Normsetzung und Entscheidungsverfahren 1

Thomas Kox | André Ullrich | Herbert Zech [Eds.]

Uncertain Journeys into Digital Futures

Inter- and Transdisciplinary Research for Mitigating Wicked Societal and Environmental Problems

Nomos

weizenbaum institut

Normsetzung und Entscheidungsverfahren

Schriftenreihe des Weizenbaum-Instituts für normative Wissenschaften

herausgegeben von | edited by

Simon Schrör

Prof. Dr. Herbert Zech

Band | Volume 1

Thomas Kox | André Ullrich | Herbert Zech [Eds.]

Uncertain Journeys into Digital Futures

Inter- and Transdisciplinary Research for Mitigating Wicked Societal and Environmental Problems

Nomos

weizenbaum institut

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de

1st Edition 2025

Published by
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Production of the printed version:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN 978-3-7560-0150-7 (Print)
ISBN 978-3-7489-4758-5 (ePDF)

DOI https://doi.org/10.5771/9783748947585

Online Version
Nomos eLibrary

Table of Contents

Shaping uncertain journeys into digital futures - perspectives on the digital and socio-ecological transformation

André Ullrich, Thomas Kox, and Herbert Zech

Our earth has already passed six of nine planetary boundaries that are critical for maintaining the stability and resilience of it (Richardson et al., 2023). Threatening environmental developments, such as overconsumption of resources (Daly, 2005; Kirsch, 2020), increasing amounts of technological waste (Notley, 2019; Saha et al., 2021), and improper handling of used products and goods (Ali et al., 2021; Zhang et al., 2022), are frighteningly compromising the health of our planet. Along with the substantial energy consumption required to train artificial intelligence (AI) tools and models (Wu et al., 2022), a major concern is that digitalisation, and especially AI, may exacerbate rather than alleviate environmental issues (van Wynsberghe, 2021). In particular, there is a need for a critical-analytical view of ecological, social, and discursive issues of sustainable AI beyond the narrative of technological solutionism, which is based on an informed and empirical foundation (Ullrich et al., 2024).

Considering the societal cohabitation of humanity, we are experiencing an erosion of democracies (Laebens & Lührmann, 2023), a new era of migration towards Europe (Osler, 2023) and North America (Massey, 2020), and increasing inequalities in various aspects, including access to education (Easterbrook & Hadden, 2021), income, and healthcare (Wang & Xu, 2023). Such disparities are accelerated by and in the digital sphere. There have been significant debates around the potentially adverse effects of increased screen time on mental and physical health, such as the excessive use of digital devices being associated with poor sleeping behaviour (Martin et al., 2021) and obesity (Fang et al., 2019) or contributing to depression and anxiety (Meier & Reinecke, 2021). Furthermore, privacy violations and cybersecurity issues compromise digital sovereignty (Momen, 2023; Timmers, 2019), affecting both individual and societal maturity. All of these are frighteningly compromising the prosperity of our futures.

In particular, sustainability and AI are two critical areas that have been gaining increasing attention in recent years. The *twin transformation* – the intertwined digital and ecological transformation of societies (Barth et al.,

2023) – holds immense potential for addressing some of the most pressing environmental challenges facing our planet today. As the global population continues to grow and natural resources become scarcer, finding innovative ways to promote sustainable developments is essential. Digitalisation has the potential to optimise resource use, reduce waste, and uncover new solutions to complex environmental problems. For example, digital platforms can be used to foster circularity and close resource loops (Körppen et al., 2024), facilitate mobility and energy transitions (Canzler & Knie, 2016; Dekeyrel & Fessler, 2024), and reduce information asymmetries of company information (Salvi et al., 2021). Additionally, they can promote the involvement and participation of citizens to influence their local spheres through the use of publicly available data (Hamm, 2020), improve climate literacy through engagement in citizen science projects on digitality and sustainability (Kox et al., 2020), and enhance relations between citizens and public organisations (Santolamazza et al., 2024).

However, the promise of participating in shaping the future seems contingent upon designated invitation, leading to civics *washing* instead of civic *involvement* (Zehner & Ullrich, 2024). Technologies and social developments thus mutually condition each other in a *co-evolutionary* (Stewart & Williams, 1998) process that precludes the identification of simple cause-and-effect relationships (Gerhold & Brandes, 2021). That emphasises the complexity of so-called *wicked problems* (Rittel & Webber, 1973) and the fundamental indeterminacy of the future. The associated uncertainty about future developments and resulting phenomena, such as mental overload, anxiety, and a weakening of social norms, can make it more difficult to deal with potential crises and their dynamics. Scientifically sound strategic foresight can be used to determine which developments are conceivable and likely to actively shape the future (Gerhold & Kox, 2024).

The Sixth Weizenbaum Conference focused on the question of how we can shape a desirable future in the face of current and future challenges and crises. This anthology comprises diverse contributions that approach this question from different perspectives against an interdisciplinary background: 1) Smart cities and urban transformation, 2) digital technologies for sustainability, 3) social justices, governance, and citizen participation, and 4) imaginaries and visions of futures.

1. Smart cities and urban transformation

In her contribution, Francesca Bria develops the building blocks for a European strategy for a digital economy based on sustainability and human rights. In view of the current polycrisis, triggered by (trade) wars, migration, and climate crises, as well as dependencies on critical raw materials and geopolitical shifts, she poses the question of a desired society. She calls for a democratic agenda to protect data, interests, critical processes, and infrastructures and an anti-monopolistic economic policy committed to climate neutrality, open-source technologies, intra-European interoperability, and competition. Mennatullah Hendawy et al. analyse the promise of smart cities and their actual (non-)implementation, rejecting it as impractical and providing avenues for possible futures of digitally enhanced cities. Bonny Brandenburger et al. investigate challenges and enabling factors for digitalisation projects in the context of smart regions from the perspective of municipality staff. They emphasise that communication within and with citizens is key and that the problems at hand must be understood as social problems so that socio-technical support can help overcome them.

2. Digital technologies for sustainability

Various aspects of the twin transformation and instruments for shaping desirable futures are presented in a number of contributions. Ina Schieferdecker looks at the uncertainties regarding climate change and the development of technologies. She emphasises the potential of sustainable, green artificial intelligence for ecological transformation and discusses accompanying myths. Lubna Rashid et al. examine the drivers of pro-environmental attitudes and behaviours among workers in high-emitting industries in Germany and highlight purpose, leadership, emotional resilience, and inclusion as decisive factors for shifting towards environmentally conscious practices. Florian Butollo et al. analyse the influence of generative AI on knowledge work and call for new requirements regarding skills, creativity, and the ability to verify the authenticity and reliability of automatically generated content for workers when engaged with generative artificial intelligence tools. Paula Scharf et al. present a mobile, energy-efficient, and low-cost sensor-based system for optimising bicycle-oriented mobility planning and safety strategies. Verena Majuntke et al. discuss the possibilities of making software more efficient and sustainable. They argue that the first step is to make these parameters measurable not just in terms of energy consumption but holistically, taking into account resource and water consumption. More collaboration between European countries is needed for international standards and effective regulation. They suggest that in addition to contributing to greater transparency, open source also enables greater choice.

3. Social justice, governance, and citizen participation

In a similar vein, several contributions elaborate on issues of social justice, governance, and citizen participation. Tina Comes et al. present a case study on the behaviour of refugees and the provision of vital infrastructure, from which concrete recommendations for improvement could be derived. Christian Herzog et al. analyse *digital sovereignty* as an *ill-structured problem* that can only be solved procedurally and call for resiliently pursuing and engaging in spaces of negotiations. Anne Mollen et al. argue along similar lines and demand a social science perspective to adequately address the question of how infrastructures of generative AI can be sustainably shaped. With this focus on sustainability, they develop a socio-technological concept of AI infrastructures that points to the socio-ecological effects of their production and discuss how AI infrastructures can be shaped sustainably. Florian Meissner et al. discuss how the Federal Office for Information Security informs the public about security problems via social media, with the sobering result that this is often inadequate, vague, and self-referential. Ame Elliott reports the results of a series of workshops and the challenges faced in cross-disciplinary research on the topic of data donation for clean energy. Most participants perceived this task as a classical engineering problem, which they felt they lacked the professional knowledge to address rather than embracing it as a wicked problem. Marc Steen's essay discusses alternative ways to design and use digital technologies to deal with societal and environmental issues by looking at non-Western indigenous cultures. Ines Weigand investigates Communities of Practice in the context of the Open Source movement regarding their specific knowledge practices and cultures. On this basis, she provides suggestions on how academia can participate in such knowledge ecosystems by focusing on the role of participatory and practice-based design research. Mennatullah Hendawy et al. identify key narratives and representations of citizens within the United Arab Emirates' national AI strategy. On this basis, they call for inclusive governance models that actively involve citizens in shaping AI policies to ensure that AI development aligns with societal needs and democratic values. Yuya Shibuya et al. investigate the potentials and limitations of civic data by contrasting it with the concepts of counter-data and open data. They reveal that there exists a limited understanding of what constitutes 'civic' data and note that issues related to power structures, diversity and inclusion, and the infrastructuring of civic data have received little attention.

4. Imaginaries and visions of futures

Several papers discuss the ideas and imaginations of the future that motivate contemporary actions. Josephine Schmitt and Samuel T. Simon analyse drawings and sketches that people had made about their visions of the future, allowing for reflections on the importance of ethical considerations in the development and deployment of digital technologies. Annemarie Witschas criticises visions of the future as powerful constructions of industrial actors working against a socio-ecological future and proposes a democratic reconquest of the future. Lorenz Erdmann presents the approach of *horizon scanning* combined with *sensemaking activities*, which could enable organisations to minimise the uncertainties of the twin transformation, eliminate systematic biases in thinking about the future, and enable meaningful action.

This anthology is intended to inform researchers, practitioners, and policymakers, encouraging them to take responsibility for our behaviour and its impact on the environment and societal life, that is, the basis of living conditions of future generations. Thus, we wish to take a step further in the right direction and help minimise the knowledge gap. But, much more importantly, we aim to provide ideas and examples that help us assess the impact of and modify our actions, as we are convinced that we currently face an action gap rather than a knowledge gap. Therefore, we call on citizens, policymakers, practitioners, and scholars: It is time to act, to individually and collectively contribute to designing convivial futures.

Acknowledgements

This work was partially funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII131 and the Open Access Publication Fund of the Weizenbaum Institute for the Networked Society, Berlin.

References

Ali, S.S., Elsamahy, T., Koutra, E., Kornaros, M., El-Sheekh, M., Abdelkarim, E.A., Zhu, D., & Sun, J., (2021). Degradation of conventional plastic wastes in the environment: A review on current status of knowledge and future perspectives of disposal. *Science of the Total Environment*, 771, 144719.

Barth, M., Gossen, M., Lang, D.J., & Santarius, T. (2023). Sustainable digitalization - fostering the twin transformation in a transdisciplinary way. *GAIA - Ecological Perspectives for Science and Society*, 32, 6–9. https://doi.org/10.14512/gaia.32.S1.3

Canzler, W., & Knie, A. (2016). Mobility in the age of digital modernity: why the private car is losing its significance, intermodal transport is winning and why digitalisation is the key. *Applied Mobilities*, 1, 56–67. https://doi.org/10.1080/23800127.2016.1147781

Daly, H.E. (2005). Economics in a Full World. *Scientific American*, 293, 100–107.

Dekeyrel, S., & Fessler, M. (2024). Digitalisation: an enabler for the clean energy transition. *Journal of Energy & Natural Resources Law*, 42, 185–209. https://doi.org/10.1080/02646811.2023.2254103

Easterbrook, M.J., & Hadden, I.R. (2021). Tackling Educational Inequalities with Social Psychology: Identities, Contexts, and Interventions. *Social Issues Policy Review,* 15, 180–236. https://doi.org/10.1111/sipr.12070

Fang, K., Mu, M., Liu, K., He, Y. (2019). Screen time and childhood overweight/obesity: A systematic review and meta-analysis. *Child*, 45, 744–753. https://doi.org/10.1111/cch.12701

Gerhold, L., & Brandes, E. (2021). Sociotechnical imaginaries of a secure future. *Eur J Futures Res*, 9, 7. https://doi.org/10.1186/s40309-021-00176-1

Gerhold, L., & Kox, T. (2024). Strategische Vorausschau im Bevölkerungsschutz, in: Bundesamt für Bevölkerungsschutz und Katastrophenhilfe (Ed.), Forschung Für Den Bevölkerungsschutz. Fachkongress Vom 12.-14. Januar 2023, Forschung Für Den Bevölkerungsschutz. Bonn, pp. 91–92.

Hamm, A. (2020). Particles Matter: A Case Study on How Civic IoT Can Contribute to Sustainable Communities, in: Proceedings of the 7th International Conference on ICT for Sustainability. Presented at the ICT4S2020: 7th International Conference on ICT for Sustainability, ACM, Bristol United Kingdom, pp. 305–313. https://doi.org/10.1145/3401335.3401815

Kirsch, S. (2020). Running out? Rethinking resource depletion. *The extractive industries and society*, 7, 838–840.

Körppen, T., Ullrich, A., Böttcher, T., & Krcmar, H. (2024). How digital platforms can foster a circular economy, in: PACIS 2024 Proceedings. Presented at the Pacific Asia Conference on Information Systems, AIS, Ho Chi Minh City, Vietnam.

Kox, T., Göber, M., Wentzel, B., Freundl, E., & Rust, H.W. (2020). Fostering weather and climate literacy among pupils by engagement in a weather citizen science project, in: Proceedings of Austrian Citizen Science Conference.

Laebens, M.G., & Lührmann, A. (2023). What halts democratic erosion? The changing role of accountability, in: *Resilience of Democracy.* Routledge, pp. 40–60.

Martin, K.B., Bednarz, J.M., & Aromataris, E.C. (2021). Interventions to control children's screen use and their effect on sleep: A systematic review and meta-analysis. *Journal of Sleep Research*, 30, e13130. https://doi.org/10.1111/jsr.13130

Massey, D.S. (2020). Immigration policy mismatches and counterproductive outcomes: unauthorized migration to the U.S. in two eras. *CMS*, 8, 21. https://doi.org/10.1186/s40878-020-00181-6

Meier, A., & Reinecke, L. (2021). Computer-Mediated Communication, Social Media, and Mental Health: A Conceptual and Empirical Meta-Review. *Communication Research*, 48, 1182–1209. https://doi.org/10.1177/0093650220958224

Momen, N. (2023). Privacy and Ethics in a Smart City: Towards Attaining Digital Sovereignty, in: Ahmed, M., Haskell-Dowland, P. (Eds.), *Cybersecurity for Smart Cities, Advanced Sciences and Technologies for Security Applications*. Springer International Publishing, Cham, pp. 47–60. https://doi.org/10.1007/978-3-031-24946-4_4

Notley, T. (2019). The environmental costs of the global digital economy in Asia and the urgent need for better policy. *Media International Australia*, 173, 125–141.

Osler, A. (2023). Education, migration and citizenship in Europe: Untangling policy initiatives for human rights and racial justice. *Global Migration and Civic Education*, 64–79.

Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S.E., Donges, J.F., Drüke, M., Fetzer, I., Bala, G., von Bloh, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hofmann, M., Huiskamp, W., Kummu, M., Mohan, C., Nogués-Bravo, D., ... Rockström, J. (2023). Earth beyond six of nine planetary boundaries. *Science Advances, 9*, eadh2458. https://doi.org/10.1126/sciadv.adh2458

Rittel, H.W., & Webber, M.M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4, 155–169.

Saha, L., Kumar, V., Tiwari, J., Rawat, S., Singh, J., & Bauddh, K. (2021). Electronic waste and their leachates impact on human health and environment: Global ecological threat and management. *Environmental Technology & Innovation,* 24, 102049.

Salvi, A., Vitolla, F., Rubino, M., Giakoumelou, A., & Raimo, N. (2021). Online information on digitalisation processes and its impact on firm value. *Journal of Business Research*, 124, 437–444.

Santolamazza, V., Mattei, G., & Grandis, F.G. (2024). Citizens' role and digitalisation in the participatory budgeting to create public value: the case of Rome. *International Journal of Public Sector Management.*

Stewart, J., & Williams, R. (1998). The Coevolution of Society and Multimedia Technology: Issues in Predicting the Future Innovation and Use of a Ubiquitous Technology. *Social Science Computer Review,* 16, 268–282. https://doi.org/10.1177/089443939801600304

Timmers, P. (2019). Ethics of AI and Cybersecurity When Sovereignty is at Stake. Minds & Machines 29, 635–645. https://doi.org/10.1007/s11023-019-09508-4

Ullrich, A., Rehak, R., Hamm, A., & Mühlhoff, R. (2024). Sustainable Artificial Intelligence Critical and Constructive Reflections on Promises and Solutions, Amplifications and Contradictions. *Weizenbaum Journal of the Digital Society*, 4. https://doi.org/10.34669/WI.WJDS/4.1.1

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1, 213–218. https://doi.org/10.1007/s43681-021-00043-6

Wang, J., & Xu, Y. (2023). Digitalization, income inequality, and public health: Evidence from developing countries. *Technology in Society*, 73, 102210.

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H., … Hazelwood, K. (2022). Sustainable AI: Environmental Implications, Challenges and Opportunities, in: Marculescu, D., Chi, Y., Wu, C. (Eds.), *Proceedings of Machine Learning and Systems*. pp. 795–813.

Zehner, N., & Ullrich, A. (2024). Dreaming of AI: environmental sustainability and the promise of participation. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02011-0

Zhang, Z., Malik, M.Z., Khan, A., Ali, N., Malik, S., & Bilal, M. (2022). Environmental impacts of hazardous waste, and management strategies to reconcile circular economy and eco-sustainability. *Science of The Total Environment*, 807, 150856.

Smart Cities and urban transformation

Europe's path to digital independence

Francesca Bria[1]

1. The age of polycrisis and a new industrial paradigm

We are currently undergoing rapid and profound changes. In Europe, the war in Ukraine continues to unfold, while in the Middle East, the ongoing conflict in Palestine threatens to spread across the region. At the same time, we are confronted with a climate emergency, a migration crisis, and the escalating US–China trade war. In these times of polycrisis, as economists and intellectuals have described, it is imperative that we seek ways to move beyond the status quo.

The technological trends we face today carry significant geopolitical implications reminiscent of the Great Power competition during the Cold War. Control over technological and scientific development – particularly in frontier technologies with disruptive societal and economic impacts – has become crucial. The ongoing race between the United States and China for technological supremacy, heavily reliant on critical raw materials, profoundly affects our ability to transition to a more sustainable and just global economic model. The vulnerabilities in supply chains and the technological decoupling of these superpowers impact all economic sectors, with substantial geopolitical and economic consequences. This situation raises the pressing question: Are we facing a new kind of Cold War, defined by technological decoupling, where the United States and China develop separate technological ecosystems?

As we navigate this era of rapid technological advancement, it is crucial to reassess the shifting geopolitical landscape and question whether we are witnessing the end of globalisation as we once knew it. The decline of hyper-globalisation challenges the neoliberal market paradigm, which advocated free trade and market efficiency as pathways to wealth creation. Instead of delivering widespread prosperity through trickle-down economics, this model has deepened inequalities, heightened wealth polarisation, accel-

1 The following text is a transcript of a keynote delivered by the author at the Weizenbaum Conference 2024.

erated climate change, and concentrated market power in the hands of a few in the Global North – often at the expense of people in the Global South and the planet.

Some experts suggest that we are entering a new phase of 'de-globalisation' or near-shoring, where nations increasingly prioritise domestic production and reduce their reliance on global supply chains. This shift is characterised by the rise of economic statecraft, where governments implement interventionist policies – such as tariffs, price controls, and restrictions on critical technologies – to protect their national interests. A notable example in the high-tech sector is the US government's decision to prohibit the Dutch company ASML from exporting its advanced lithography technology to China. This move, intended to curb China's technological advancement in key areas, underscores a growing economic protectionism trend. It marks a departure from the laissez-faire approach that dominated previous decades and prompts a reevaluation of the relationship among economic policy, state intervention, and global trade. These developments suggest that we are entering a new geopolitical landscape where traditional notions of globalisation are being reshaped by the imperatives of national security, technological sovereignty, and economic resilience.

2. The quest for Europe's technological independence

Europe views the current technological landscape as a critical political challenge that necessitates a strategic and alternative approach. In response to the Covid-19 pandemic, Europe has prioritised shared agendas and ambitious goals, including the creation of a new generation of European common debt to commit significant investments to the European Green Deal. The continuation and success of these initiatives will hinge on the effectiveness of the upcoming European Commission.

Europe faces the urgent task of reclaiming its sovereignty in the technology sector. Beyond its dependence on rare earths and raw materials – many of which are sourced from the Global South and are essential for digital and green transitions, such as cobalt and silicon, predominantly mined in Africa – Europe is also critically dependent on hardware, semiconductors, computing power, data infrastructure, AI, and digital platforms, all of which are predominantly controlled by US-based companies.

Europe's extensive reliance on external technological infrastructure has placed it in a precarious position, with its digital sovereignty increasingly dictated by foreign forces. The dominance of US-based tech giants over critical components of Europe's digital ecosystem – ranging from chips and computing power to AI and digital platforms – has led to what can be described as a new form of digital colonialism. In this scenario, Europe has, in many ways, ceded control over its digital future, with its industries and institutions deeply embedded in and dependent on technologies governed by external entities.

To address this vulnerability, European policies are increasingly focused on enhancing supply chain resilience and ensuring economic security. This approach includes advocating for decolonial and resilient supply chains that prioritise fairer and more sustainable industrial practices, as well as developing robust domestic capabilities in hardware, computing, AI, and digital platforms. By reducing these dependencies, Europe aims to reclaim its digital sovereignty and safeguard its technological future from external control.

The challenges ahead are significant. The intensifying chips war, the surging demand for computational power, and the growing concentration of AI market control underscore the urgent need for Europe to secure technological sovereignty in this rapidly evolving global landscape. Addressing these challenges is critical for Europe to maintain its strategic autonomy and ensure its digital infrastructure aligns with its long-term interests.

a) The chips war

One of the most pressing examples of technological vulnerability is the ongoing 'chips war', as highlighted in Chris Miller's book, which underscores the critical importance of microprocessors in global technology and industry. The Covid-19 pandemic starkly exposed the fragility of global supply chains, particularly in Germany's automobile sector, where a severe shortage of semiconductors led to widespread production delays and substantial economic losses. This crisis revealed the world's heavy reliance on a few key microprocessor manufacturers.

The digital revolution, far from an abstract concept, is deeply rooted in substantial physical infrastructure – submarine cables, satellites, and data centres – all powered by microprocessors. The chip supply chain, highly concentrated with significant production in Taiwan, particularly by the Taiwan Semiconductor Manufacturing Company (TSMC), and with crucial machinery provided by ASML in the Netherlands, presents significant geopolitical risks. This dependency highlights the urgent need for more secure, resilient, and diversified supply chains to mitigate the growing risks posed by geopolitical tensions and potential disruptions.

For Europe, the implications of these vulnerabilities are profound. The continent's heavy reliance on external suppliers for critical technologies places its industries, especially automotive and high-tech manufacturing, at considerable risk. As geopolitical tensions continue to escalate, particularly between the United States and China, Europe's position in the global semiconductor supply chain becomes increasingly precarious. This situation underscores the urgent need for Europe to develop its own semiconductor manufacturing capabilities and diversify its supply sources. To address these challenges, Europe must invest heavily in achieving technological sovereignty. Initiatives like the European Union's European Chips Act are essential steps in the right direction, aiming to reduce dependency on foreign suppliers and enhance resilience against supply chain disruptions.

While strengthening Europe's domestic semiconductor industry and forging strategic partnerships is essential for reducing risks associated with a concentrated global chip supply chain, this approach faces significant obstacles. The immense costs and lengthy timelines needed to build a competitive semiconductor sector, coupled with Europe's current technological lag and dependence on global supply chains, complicate the pursuit of true technological sovereignty. Furthermore, geopolitical pressures – especially from the US–China rivalry – and the inherently global nature of the semiconductor industry mean Europe may still be reliant on external expertise and partnerships for the foreseeable future. Therefore, to truly secure its tech sovereignty, Europe must implement robust regulatory frameworks and industrial policies that not only foster long-term technological independence but also ensure sustainability and resilience against geopolitical disruptions.

b) Computational power

Computing power, particularly in the realm of supercomputing, is rapidly emerging as a critical battleground in the ongoing geopolitical and technological competition. Europe is making substantial investments in both public and private computing infrastructures to secure its position in the global digital economy. These investments are essential not only for advancing scientific research but also for tackling complex societal challenges, such as climate change, and for ensuring Europe's digital sovereignty.

At the forefront of this technological race is quantum computing – a field that, although still in its early stages, has the potential to revolutionise scientific paradigms and reshape global power dynamics. Recognising the transformative potential of quantum technologies, both the public and private sectors are making heavy investments in research and development.

For Europe, the stakes are high. As quantum computing and supercomputing become integral components of the modern technological landscape, Europe's ability to compete on the global stage will increasingly depend on its capacity to lead in these areas. Initiatives like the EuroHPC Joint Undertaking and the Quantum Technologies Flagship are vital efforts to bolster Europe's capabilities, ensuring the continent remains competitive. Failure to do so could leave Europe vulnerable to technological dependencies.

Advancing computational power indeed brings significant sustainability challenges, particularly due to the resource-intensive nature of data centres. These centres, essential for storing and processing the massive amounts of data required for supercomputing and quantum computing, consume vast quantities of water, electricity, and land. According to a report published in *Nature*,[2] data centres annually consume more than 200 terawatt-hours of electricity worldwide, rivalling the energy consumption of some mid-sized countries. Moreover, Goldman Sachs' recent report highlights that AI is poised to drive a 160% increase in data centre power demand as the processing power required for training and deploying advanced AI models continues to escalate. This surge in energy demand, coupled with the significant water and land resources required for cooling and housing these facilities, amplifies the environmental challenges posed by the current extractive data centre model. As the demand for data processing grows, addressing these sustainability issues becomes increasingly critical to balance technological advancement with environmental stewardship.

Europe must, therefore, rethink its approach to data centres, focusing on more sustainable models for public computing infrastructure. Doing so could involve developing energy-efficient data centres, exploring innovative cooling methods, and ensuring that data centre locations are selected with environmental impact as a key consideration. By integrating sustainability into its strategy for enhancing computational power, Europe can build a

2 https://www.nature.com/articles/d41586-024-00478-x

digital infrastructure that not only supports technological innovation but also aligns with its broader environmental objectives.

c) Market power in artificial intelligence

The rapid rise of AI has sparked critical questions about the true scope of its advancements and the risks posed by the concentration of power in the hands of a few tech giants. Companies like Google, Amazon, Microsoft, and NVIDIA control vast reservoirs of data, computational power, and cutting-edge chips essential for training advanced AI models. This concentration has raised alarms about monopolistic practices and the potential for these companies to drive AI development in ways that may not align with the broader public interest.

Take NVIDIA, for instance – a leader in AI hardware. The company recently saw a sharp decline in its stock value after initial excitement over AI's growth potential was tempered by fears of overvaluation. This episode underscores the fragility of market confidence in AI's long-term economic impact. Some experts have even warned of a looming AI bubble, where the current hype could lead to overinvestment and eventual disillusionment, reminiscent of the dot-com bubble of the early 2000s.

Moreover, the AI boom has intensified concerns about power concentration within the tech industry. NVIDIA, which controls significant aspects of the AI stack – from chips to software – exemplifies this trend. The company's market capitalisation now surpasses that of several European countries combined, underscoring its vast influence. NVIDIA's vertical integration across the AI ecosystem allows it to dominate the market, raising serious concerns about the potential risks of an unregulated AI oligopoly for economic stability, national security, and social equity.

To address these growing challenges, a robust anti-monopoly approach to AI governance is becoming increasingly urgent. Policymakers in both Europe and the United States, including Lina Khan, head of the US Federal Trade Commission, are recognising the need to assess and mitigate the risks posed by concentrated market power in the AI sector. Khan has called for audits to identify choke points within the AI ecosystem, currently controlled by a few dominant players. This concentration of power threatens not only small and medium-sized enterprises (SMEs) but also citizens and industry leaders alike. For example, Sam Altman, CEO of OpenAI, recently claimed he aims to raise six trillion dollars to invest in chips – though this figure might be more marketing hype than economic reality.

Nonetheless, the underlying message is clear: the AI market is becoming increasingly monopolistic and concentrated. To ensure a competitive, dynamic environment that harnesses these technologies for the public good, significant industrial effort and robust regulatory frameworks are essential.

3. Surveillance capitalism and the threat to a democratic digital public sphere

Beyond its concentration of market power, the digital ecosystem poses serious threats to democratic accountability, a danger vividly outlined by Shoshana Zuboff (2019) in her concept of surveillance capitalism.[3] Zuboff describes this new phase of economic history as one driven by relentless data tracking, where every action we take on computers and mobile devices is meticulously monitored and monetised by Big Tech. This vast trove of personal data is then used to train sophisticated AI systems, which, in turn, fuel predictive models, machine learning algorithms, and deep learning technologies that increasingly shape our experiences, choices, and decisions.

In the era of surveillance capitalism, individuals are no longer just customers or products – they are the raw material. This shift brings to light significant concerns about data extraction and the growing concentration of AI capabilities within a few dominant tech companies. With near-unlimited access to personal data, these companies have unprecedented power to predict, influence, and control human behaviour. This concentration of power poses serious risks to privacy, personal autonomy, and the very foundations of democratic governance.

Recent studies reveal the disturbing extent of this issue, showing how major tech companies, in collaboration with advertisers, exploit personal data to build detailed behavioural profiles. These profiles are then used to manipulate consumer choices, political opinions, and even voting behaviour – often without users' informed consent. Such control erodes the principles of autonomy, self-determination, and accountability that are essential to a functioning democracy.

Furthermore, the unchecked manipulation and monetisation of personal data threaten to deepen existing inequalities. Those with access to vast

---

3 Zuboff, S., 2019. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power, edn. *PublicAffairs, New York.*

amounts of data can further consolidate their power and influence, raising critical concerns about who controls these technologies, how they are used, and their long-term impact on society. If left unregulated, the dominance of a few powerful entities in the digital space could lead to a future where democratic processes are undermined and individual freedoms are increasingly curtailed. To counter these threats, there is an urgent need for robust regulatory frameworks that protect fundamental rights, safeguard privacy, and hold tech companies accountable for their access to and use of data. Without such measures, the promises of the digital age could be overshadowed by a dystopian reality where our lives are governed by algorithms serving the interests of a few rather than the common good.

Reclaiming a digital public sphere is crucial to countering Big Tech's dominance. This effort includes forging new alliances with national broadcasters and public media. In some European countries, concerns are growing about state-controlled public communication turning into propaganda while digital information channels remain in the hands of media tycoons. Consider Elon Musk, who owns X (formerly Twitter) and holds vast amounts of social media data. During the German elections, Musk openly supported the far-right AfD party and now advocates for Donald Trump's re-election in the United States, underscoring the urgent need to reclaim a truly accountable digital public sphere.

It is essential to ensure that digital public spaces align with our values and protect social rights and democracy. We must confront the critical question of power in the digital age: who holds it, for what purposes, with what legitimacy, and how we can move beyond technological solutionism. As historian Evgeny Morozov (2023) argues,[4] this also involves challenging the ideologies of Silicon Valley, such as technological solutionism and the pursuit of artificial general intelligence (AGI), which often lack scientific grounding and instead reinforce the neoliberal paradigm. In a year filled with elections, the dangers of manipulation and misinformation are more pressing than ever. Fake news and targeted misinformation are exacerbated by the fact that a few media oligopolies control today's communication and information channels. To safeguard democracy, we need to rethink our approach to digital governance, prioritise the public interest, and ensure that technology serves as a tool for empowerment rather than control.

4 Morozov, E., 2023. The True Threat of Artificial Intelligence. *International New York Times.*

4. Building and governing technology in the public interest

A central component of Europe's strategy should be the deliberate shaping of its digital future. The pandemic underscored the urgency of directing the digital transition, as essential services – education, healthcare, urban management, and traffic control – have become deeply intertwined with digital technologies. These technologies have evolved into critical infrastructures, making it imperative to ensure democratic and accountable control over them. Europe has responded by crafting a regulatory framework that places rights and principles at the forefront of its proposed digital transformation. The Declaration of Digital Rights, signed by all European governments, enshrines fundamental rights such as information self-determination, autonomy, education, and equality.

Europe's new regulatory framework is designed to serve as a constitution for the digital age. Despite some scepticism regarding its potential bureaucratic hurdles, this framework is essential for fostering a different kind of innovation – one that is inclusive and rights-based. Effective enforcement is crucial to this effort. Clear regulatory goals must prioritise shifting the balance of power away from tech oligopolies and towards a more dynamic, inclusive innovation ecosystem.

Key legislative measures, including the Digital Services Act (DSA) and the Digital Markets Act (DMA), target anti-monopoly practices Big Tech companies employ. Additionally, the Artificial Intelligence Act (AIA) and the Data Governance Act (DGA) address critical aspects of AI development and data management. It is vital to ensure that these regulations are implemented in a way that does not stifle small players, academia, research institutions, or public administrations. Instead, the goal is to create public options for digital infrastructures, protect fundamental rights, and enable the development, access, and use of technologies in the public interest.

In the realm of anti-monopoly measures, Europe may need to take even more assertive steps. A major concern is the increasing vertical integration across what is known as the 'AI stack', which functions similarly to a supply chain for AI. The stack comprises four key layers: physical hardware (chips), cloud infrastructure, AI models, and applications. Vertical integration occurs when companies expand their control across multiple layers of this stack, consolidating power at each level. This trend is particularly troubling because the dominant Big Tech firms are already deeply embedded throughout the entire stack. As these companies strengthen their hold on one layer, they can leverage that position to gain an unfair advantage in other layers, potentially stifling competition and innovation. This concentration of power raises questions about whether the Digital Markets Act alone is sufficient. More drastic interventions, such as breaking up dominant companies, may be necessary to prevent monopolistic control.

As argued by different anti-monopoly scholars,[5] one approach to regulating AI is through structural separation, which would prevent businesses from operating at multiple layers of the AI stack. For example, this could mean prohibiting cloud providers from also running the applications and services that depend on their cloud infrastructure. Another related regulatory concept is nondiscrimination, which would require businesses to offer equal service and pricing to all users, preventing preferential treatment. Looking beyond these measures, the EU could consider creating a public option for cloud computing. This service would involve strengthening the network of publicly funded and managed supercomputers that serves government agencies, researchers, and others focused on solving public challenges rather than using AI to enhance the profitability of tech platforms. Such a public cloud could be trusted to prioritise societal interests and user needs over commercial gain.

a) Digital industrial policy for EU strategic autonomy

Europe's response to the current technological challenges is multi-faceted, involving robust regulation, strategic investment in green and digital industrial strategies, and a steadfast commitment to democratic control and the protection of fundamental rights.

To achieve genuine strategic autonomy, Europe must implement a comprehensive and coordinated digital industrial policy that reflects the urgency of the moment. The ongoing investments in critical areas such as 5G, cloud computing, semiconductors, quantum technologies, and batteries are steps in the right direction. However, these efforts must be part of a broader strategy that integrates both the green and digital transitions. Europe needs to harness these substantial investments to build domestic capacity, ensuring that key technologies are developed and maintained throughout the continent.

A critical aspect of this strategy is making EU infrastructures widely accessible to society. Europe's network of supercomputing centres represents

5 Sitaraman, G. and Narechania, T.N., 2023. An Antimonopoly Approach to Governing Artificial Intelligence. *Available at SSRN.*

a powerful resource that should be opened to a diverse range of stakeholders – researchers, startups, civil society organisations, NGOs, journalists, and artists. By facilitating multidisciplinary collaboration, Europe can unlock the full potential of data and technological advancements. Adopting a governance model based on public office–society cooperation could set a new standard for digital innovation, uniquely positioning Europe as a leader in inclusive and sustainable technological development.

The proposal for a €10 billion Sovereign Tech Fund for Europe is a crucial step toward this goal. In discussions with EU officials and through building strategic alliances, it has become clear that while the final figure may vary, the need for substantial resources is undeniable. The forthcoming Draghi report underscores this urgency, advocating for a €500 billion investment in European public goods. Digital public infrastructures, alongside other critical infrastructures, must be recognised as fundamental European public goods that require robust financial commitment.

Policymakers need to acknowledge that digital public infrastructure goes beyond mere physical assets – it encompasses open technology transfer, open-source standards, sovereign AI, data commons, and privacy-preserving technologies. Such infrastructure must be open, interoperable, and sovereign, managed as digital commons,[6] not controlled by the state or privatised for profit. This vision is essential for safeguarding Europe's digital future and ensuring that technological advancements serve the public interest rather than a narrow set of private interests.

In summary, Europe's strategic autonomy hinges on a well-coordinated digital industrial policy, significant investments, and the creation of open, accessible, and sovereign digital public infrastructures. These actions are not just necessary for fostering innovation and protecting privacy – they are fundamental to placing the public interest at the core of Europe's technological future.

b) A vision for an AI digital commons ecosystem

A crucial element of this vision centres on the ownership and collective control of our data. While the European Union has introduced important regulations, such as the Data Governance Act and the Data Act, it is not

6 https://www.nldigitalgovernment.nl/news/european-collaboration-for-digital-commons/

enough to simply have these rules in place. We must actively leverage them to ensure that Europe develops a distinct approach to data-driven innovation – one that promotes information self-determination, data sharing in the public interest, data portability, and data quality.

For example, private data with significant public value, such as information on pollution, mobility, climate, and housing, should not be monopolised by a few private entities. Instead, this data should be made accessible to society to help address collective challenges. One promising policy measure being advocated at the European level is the creation of a European Data Intermediary[7] in the public interest. This publicly governed data body would ensure that valuable data is not hoarded by large corporations but is available under transparent and fair rules. Currently, researchers face significant barriers to accessing essential data, even when their work is publicly funded. This situation is unacceptable, as scientific research should be a collective asset that benefits society, not just a few firms.

My work has focused on governing artificial intelligence and data as common goods that serve the public interest. As the chief technology officer (CTO) of Barcelona, I pioneered these approaches at the city level[8] and collaborated with Hamburg to develop a framework for data sharing in the public interest.[9] This model has the potential to scale and be implemented across European cities, extending beyond Hamburg. The initiative begins with urban data – data generated in public spaces – and ensures it is managed as a common good through public procurement clauses and tender requirements. Data funded by taxpayers should be managed within a data trust, benefiting society rather than being appropriated by private companies.

In today's world, data is akin to a new form of public utility that is essential for managing critical infrastructures. To fully harness its potential, we must treat data as a common good – accessible and beneficial to all – rather than allowing it to be confined to private interests.

Mobilising European cities, which are closer to citizens and equipped to manage critical digital infrastructures, provides – in my view – a unique opportunity to connect the climate agenda with data democracy and participatory governance. Citizen involvement is essential, not just to counter

7 https://publications.jrc.ec.europa.eu/repository/handle/JRC133988
8 https://www.ucl.ac.uk/bartlett/public-purpose/publications/2022/feb/new-data-deal-case-barcelona
9 https://thenew.institute/media/pages/documents/529e984d02-1698245881/the-new-hanse_blueprint_governing-urban-data-for-the-public-interest.pdf

threats to democracy but to empower individuals and harness collective intelligence for more informed decision-making. In Barcelona, where I served as CTO, 80% of policy agenda proposals originated from citizens, demonstrating the transformative power of participatory democracy. The best response to right-wing populism and democratic threats is not less democracy but more – a deeper, more inclusive participatory democracy that should be at the heart of the European project.

c) Building the Euro stack

The critical question is how to fund and develop the building blocks for a sovereign European public digital infrastructure. We must bring this term – public digital infrastructure – back into public discourse. We currently lack basic public digital infrastructures essential for building a fair, democratic, and accountable digital society.

Infrastructures like roads, electricity, and water systems have long fuelled societal progress. In the 21st century, digital infrastructures have become just as critical, underpinning the essential services upon which society depends. This new layer of infrastructure includes digital IDs, payment systems, and platforms for data management and exchange. These digital systems are now indispensable for everything from vaccine distribution to accessing social welfare, healthcare, and education. Governments, philanthropic foundations, and the private sector are investing heavily in digital public infrastructure (DPI). However, the direction of this investment is far from neutral – it is actively shaping the essential services of our society. To ensure these developments serve the public good, the EU must engage stakeholders across sectors to promote transparency, accountability, and citizen empowerment in the design and deployment of digital technologies.

Infrastructure shapes our world; thus, it is crucial that we shape our infrastructure. Failing to control and govern the digital infrastructure we rely on leaves us economically dependent and vulnerable to security risks. An ambitious and holistic approach is imperative and should involve utilising various tools and mechanisms to finance and scale existing national initiatives that align with European values and objectives. At the core of this effort should be a strong emphasis on free, open-source technology, data portability, and privacy-preserving standards. These principles of 'digital sovereignty' can be enshrined as legal obligations, embedded in procurement agreements, and made prerequisites for public financing or funding.

By mandating these standards, Europe can foster an environment that encourages innovation and collaboration while safeguarding privacy, security, digital rights, and data integrity.

Europe is working on several key initiatives, such as the digital ID project, which uses cryptography to enable people to control their data and privacy. This project would enable secure, pan-European digital transactions and application development, all while protecting privacy. Additionally, there are efforts to create public data spaces – repositories of data that can be used to build applications – along with a digital euro and digital payment system. Another focus is on open-source and fair AI ecosystems, collectively referred to as digital commons.

These initiatives form part of a broader vision for a robust digital public infrastructure in Europe. I proposed a €10 billion EU fund, called the EU Sovereign Tech Fund,[10] which aims to support this vision. Germany, for instance, already has a Sovereign Tech Fund that creates and maintains open-source software for communities. However, this effort needs to be expanded to different layers of the technology stack and connected across Europe with ambitious, dedicated resources to be truly effective.

The goal is to leverage the EU tech stack to develop EU-wide interoperable digital public services that have a tangible impact on the ground, starting from cities in areas such as mobility, climate, and housing. Local digital twins, like those in Hamburg and other German cities, are crucial for improving these services. Imagine having an interoperable European application that you can use anywhere in Europe, running on top of a European Stack – this kind of seamless integration is essential.

By integrating these elements – sovereign digital infrastructure, interoperable public services, and empowered citizen participation – Europe can build a resilient, inclusive, and democratic digital future.

5. The way forward: A new Manhattan Project for science, technology, education, and culture

Shaping the future is not just about advancing technology – it is about forging new alliances, establishing new norms, and transforming our institutions. Without reimagining and reinventing these institutions, our goals

10 https://www.euractiv.com/section/digital/opinion/towards-sovereign-ai-europes-greatest-challenge/

will remain out of reach. This vision demands multidisciplinary innovation and a holistic approach.

As a board member of the New European Bauhaus,[11] I have witnessed firsthand the transformative power of collaboration among science, technology, and the arts in advancing the European Green Deal. This movement is rethinking public spaces, pioneering bio-based materials, and designing ecological cities, fundamentally reshaping our built environment. Considering that the built environment and food production account for 70% of greenhouse gas emissions, transforming our cities and construction methods is essential for meeting climate targets and building a regenerative economy.

While there are frequent calls for massive investments in artificial intelligence, often advocating for billions in AI development, I believe what we truly need is a Manhattan Project for culture and education. Instead of focusing solely on AI, we should prioritise upgrading skills, enhancing education, fostering talent, creating quality jobs, and promoting gender equality and diversity in the era of AI and automation. By investing in culture and education, we can develop the capacity to build and govern technology and digital infrastructures in a way that serves the public interest.

The future of digital development often seems to hinge on a choice between two models: Silicon Valley's Big Tech, with its surveillance capitalism and corporate concentration of power, and China's Big State model, where the Communist Party controls data and AI systems. Neither model aligns with democratic values rooted in constitutional rights. We must move beyond Big Tech and the Big State. Instead, we need to build a 'Big Democracy' that leverages technology in the public interest through a truly democratic framework.

To implement this vision, collective action is essential. Europe's approach to innovation must be grounded in values and human rights, promoting a model of democratic digitisation with genuine democratic control. Our goals should include ensuring strategic economic competitiveness, full democratic participation, environmental protection, data privacy, and the safeguarding of fundamental rights. Addressing the challenge of governing technology in the public interest requires robust regulation, thoughtful industrial policy, anti-monopoly measures, and extensive multidisciplinary research and innovation. Investing in a €10 billion EU Sovereign Tech

11 https://new-european-bauhaus.europa.eu/

Fund dedicated to building EU digital public infrastructures is a critical step toward establishing the democratic digital society that Europe urgently needs.

By adopting these strategies, we can move beyond the current model and foster an approach to innovation that prioritises the public good, promotes equitable economic growth, and ensures a resilient and sustainable global economy. By putting people first, we can ensure that the power and potential of technology benefit everyone and support the advancement of our democratic societies.

References

Zuboff, S., (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.

https://www.nature.com/articles/d41586-024-00478-x

Morozov, E., 2023. The True Threat of Artificial Intelligence. *International New York Times*.

Sitaraman, G. and Narechania, T.N., 2023. An Anti-monopoly Approach to Governing Artificial Intelligence. *Available at SSRN*.

https://www.nldigitalgovernment.nl/news/european-collaboration-for-digital-commons/

https://publications.jrc.ec.europa.eu/repository/handle/JRC133988

https://www.ucl.ac.uk/bartlett/public-purpose/publications/2022/feb/new-data-deal-case-barcelona

https://thenew.institute/media/pages/documents/529e984d02-1698245881/the-new-hanse_blueprint_governing-urban-data-for-the-public-interest.pdf

https://www.euractiv.com/section/digital/opinion/towards-sovereign-ai-europes-greatest-challenge/

https://new-european-bauhaus.europa.eu/

Digital passivisation: The framing of citizens and their participation in the United Arab Emirates' national AI strategy

Mennatullah Hendawy, Jan Wieland, and Furkan Koc

Abstract: This article explores how the UAE's national AI strategy frames citizens and their participation in AI development and deployment. By employing both qualitative and quantitative content analysis, the study identifies key narratives and representations of citizens within the strategy. The findings reveal that the strategy predominantly positions citizens as beneficiaries of AI-driven advancements rather than active participants in policy development. This approach, termed **digital passivisation**, highlights the limited emphasis on citizen engagement and the top-down implementation of AI policies, which may hinder broader public involvement. The article concludes with a call for more inclusive governance models that actively involve citizens in shaping AI policies to ensure that AI development aligns with societal needs and democratic values.

Keywords: digital passivization, citizen participation, national AI strategy, AI governance, AI policy

1. Introduction

Citizen participation typically refers to the involvement of individuals in decision-making processes or activities related to governance, policymaking, or community development. It often implies a more active role for citizens, such as voting, attending public meetings, serving on advisory committees, or participating in grassroots initiatives (Rowe & Frewer, 2005). As we move forward into the current Digital Age, we have witnessed an increase in the relationship between digital transformation and citizen participation (Tolbert & McNeal, 2003). Kuhn (2017) asserts that technological developments have the potential to enrich both decision-making and policy processes by broadening the number of people who can participate. An example of these technologies is artificial intelligence (AI).

AI has been defined as 'the science of making machines that can perform tasks that would require intelligence if performed by humans' (Minsky, 1969, p. v). During the 2010s, progress in AI garnered significant interest from both researchers and practitioners, unlocking a wide array of advantageous prospects for utilising AI in the public sector (Pan, 2016; Wirtz et al., 2019). Several studies have demonstrated that AI can facilitate community participation and engagement in governmental policy development processes and public discussions (e.g. Battina, 2017; Havrda, 2020; Sharma et al., 2020). Additionally, AI has aided decision-making in the public sector (Reis et al., 2019; Valle-Cruz & Sandoval-Almazan, 2018). By using artificial intelligence, policymakers can adopt a more holistic and practical approach to decision-making (Milano et al., 2014). Although artificial intelligence offers tremendous potential for transformation, it has so far had only a limited positive impact on public policy (Engstrom et al., 2020; Misuraca & van Noordt, 2020; Reis et al., 2019). In fact, AI has been criticised for worsening existing disparities when not properly executed and controlled (Zajko, 2022). AI systems are often viewed as supporting 'top-down decision-making processes', such as those involved in public budgeting, where decisions are made without community input (Havrda, 2020). Furthermore, AI has been sought to exemplify how digital technologies are increasingly used for surveillance (Herasimenka, 2020).

AI, as a rapidly developing technology, holds a multitude of possible implications for economic and political life (NITI, 2018). Its significance has been recognised by various governments worldwide, leading to the publication of national AI strategies (NASs) in many countries (European Commission, 2022; Lords, 2018; National Science & Technology Council, 2016; see also Bareis & Katzenbach, 2022; Papadopoulos & Charalabidis, 2020). These NASs 'use different kinds of imaginaries – democratic, sociotechnical, and data – that help citizens envision how a future AI democracy might look'. Such efforts lead to a new type of interaction between citizens and governments (Paltieli, 2022, p. 1,613). For example, a study by Hendawy and Kumar (in press) analyses national AI strategies (NASs) in five Arab countries: UAE, KSA, Egypt, Qatar, and Morocco. The study aimed to clarify AI's portrayal within these strategies, identifying six key representations: AI as part of the national vision, a booster for economic growth, a creator of new job opportunities, a facilitator of regional leadership, potentially out of control, and a threat to employment. However, further research is still needed to investigate how citizens are affected by these narratives.

Officially published documents provide a basis for interpreting and enforcing laws, shaping public policies, and regulating professional practices (Bovens, 2009). These documents are assumed to mirror the cultural, political, economic, and ethical values and principles of a nation (Amboro et al., 2022). Given this information, the text from national AI strategies provides insights into how the strategy creators view their role in relation to AI technology and its impact on citizens. NASs have the potential to shape perceptions of the interaction between citizens and technology (for a wider discussion on the AI-citizen relation, see Hendawy & Ansari, 2023).

Given that technology, and AI in particular, has the potential to both enhance and hinder citizen participation, it becomes crucial to understand how and to what extent these effects occur through the narratives communicated by governments in their published national AI strategies. This chapter analyses the United Arab Emirates (UAE) national AI strategy, focusing on how this strategy frames citizens and their participation. Such an examination highlights the intersection of AI and citizen participation and sheds light on the overarching narratives of governmental strategies.

In October 2017, the UAE government, through the UAE AI office, unveiled the vision for the National Strategy for Artificial Intelligence 2031. The objective is to position the UAE as a global AI leader, as noted by OECD.AI (2022). The UAE was the second country worldwide and the first in the Middle East to publish a national AI strategy. Before the official announcement in 2017, the UAE had been laying the groundwork for embracing AI, recognising its transformative potential across various sectors, including governance, healthcare, education, and transportation.

This initiative is part of a broader vision to position the UAE as a knowledge-based economy and a global innovation leader. In a pioneering move, the UAE established the world's first Ministry for Artificial Intelligence in 2018 (Dutton, 2018). Subsequently, the UAE began developing a national AI strategy and creating concepts to strengthen AI implementation (United Arab Emirates Minister of State for Artificial Intelligence Office, 2018). The Ministry of AI has set eight strategic objectives to establish a comprehensive support system by 2031, transitioning from a nation that introduces AI to one that builds and mainstreams it within its institutions and society. These objectives are:

1. Build a reputation as an AI destination
2. Increase the UAE's competitive assets in priority sectors through AI deployment
3. Develop a fertile ecosystem for AI
4. Adopt AI across customer services to improve lives and governance
5. Attract and train talent for future AI-enabled jobs

6. Bring world-leading research capability to work with target industries
7. Provide the data and supporting infrastructure essential to become a test bed for AI
8. Ensure strong governance and effective regulation

According to the United Arab Emirates Minister of State for Artificial Intelligence Office (2018), the national AI strategies are structured into three distinct levels:

1. **Foundation Level:** This foundational stage includes the essential elements needed to support AI development and implementation across various sectors. It focuses on establishing a conducive ecosystem for AI innovation, attracting and training talent for AI-enabled jobs, enhancing research capabilities, and providing necessary data and infrastructure to serve as a testing ground for AI technologies.
2. **AI Activity Level:** This level aims to advance the development and deployment of AI technologies in key economic and social sectors. It includes training initiatives for students, government employees, and the broader society to equip them with AI-related skills. The goal is to standardise AI work, enhance the UAE's competitiveness through AI, and foster an environment conducive to AI integration in all areas of customer care to improve governance and quality of life.
3. **Leadership Level:** This level focuses on establishing the UAE's reputation as a forward-thinking nation and achieving the vision of becoming a global leader in AI by 2031.

The UAE's National AI Strategy is a product of expert deliberations (United Arab Emirates Minister of State for Artificial Intelligence Office, 2018). Therefore, a thorough textual examination is suitable for analysing the portrayal of citizens and their participation within the strategy as a national discourse. This chapter aims to address the research question: How does the UAE's national AI strategy frame citizens and their participation in the context of AI development and deployment? The study explores the extent to which the strategy encourages or hinders citizen participation through the narratives it communicates, analysing how citizens are portrayed within the UAE's AI strategy and examining the implications of this portrayal for public engagement and governance.

The article proceeds by presenting the method used to analyse how the UAE's national AI strategy frames citizens and citizen participation. Next, we reflect on how each aspect (citizens/participation) is mentioned in the

UAE's national AI strategy, illustrating how it frames citizens and citizen participation to answer the aforementioned research question. Afterwards, we discuss the findings and conclude that it is also important to understand the approach adopted in developing national AI strategies.

2. Method

To understand how the UAE's national AI strategy frames citizens and their participation, we employed both qualitative and quantitative content analysis. The content analysis approach systematically examines the occurrence of specific words, themes, or concepts within qualitative data sets, such as texts from policy documents or media articles (Krippendorff, 2018; Mayring, 2004; Neuendorf, 2017). Our research aimed to uncover major narratives around citizens and participation within the strategy, assessing the emphasis or lack thereof placed on citizens and their participation. Through content analysis, we analysed the frequency and context of terms related to citizen participation within the UAE's National AI Strategy document. This method facilitates a deeper understanding of how certain subjects or themes are framed or emphasised within the document (Entman, 1996; Krippendorff, 2018; Mayring, 2004). In our case, we focused on citizens and their participation, providing insights into the strategy's stance and policy direction concerning AI and its societal implications. Using content analysis as the primary methodological approach in this article is considered suitable to answer the above-mentioned research question, as it helps analyse how a text constructs a shared vision of a phenomenon through a narrative, which also has the potential to shape public perception of the phenomenon (Entman, 1996; Krippendorff, 2018; Mayring, 2004).

In particular, we incorporated both quantitative and qualitative content analysis techniques to provide a comprehensive overview of how AI and citizen participation are framed within the UAE's strategy. We focused on both the quantification and contextual use of specific terms related to citizen participation in the UAE's AI national strategy document. This combination allows for measurable data analysis and deeper insight into the underlying contexts and meanings (Elo & Kyngäs, 2008; Krippendorff, 2018; Lombard et al., 2002; Mayring, 2004; Neuendorf, 2017). Quantitative content analysis was employed to quantify the frequency of specific keywords related to citizen participation within the UAE's National AI Strategy (Riffe et al., 2019). Beyond merely counting words, we also examined the context and implications of how these terms are used within the document, incorporating qualitative content analysis techniques. Qualitative content analysis involves interpreting the meanings, themes, and patterns that emerge from the usage of these words (Elo & Kyngäs, 2008; Mayring, 2004). For instance, in this article, the application of this method revealed that the term 'citizen' is often associated with economic roles rather than political participation, indicative of the qualitative aspect of our analysis, as discussed in the following section.

The keywords used in the analysis were selected based on their relevance to the theme of citizen participation, informed by preliminary readings of the strategy and an understanding of common terms used in discussions about public participation and governance. The keywords include 'citizen(s)', 'people', 'public', 'society', 'participate/participation', and 'democratic/democracy'. An overview of the search results is presented in Table 1. The analysis in this research was initiated as part of a 2023 seminar on citizen participation in the age of digital innovation at Duisburg-Essen University, where students formed groups to analyse different national AI strategies. Collective discussions not only informed the analysis but also occasionally influenced the selection of keywords.

For the analysis, after the specific keywords related to citizen participation were identified, their context within the document was examined. We searched for these terms in the document and evaluated both the frequency and content of relevant text passages (see Bowen, 2009). Each occurrence of the keywords was analysed to capture the context and implications regarding citizen participation. This involved noting how often each keyword appears and the specific mentions that relate directly to the role of citizens in the AI strategy. Additionally, we analysed these keywords in reference to their quotes, delving deeper into the intended meaning and significance of those terms within the context of the text, taking into account any specific connotations or interpretations implied by their usage. For each keyword, the surrounding text was examined to understand the depth of citizen participation suggested – whether it indicates active participation, consultation, or mere information dissemination. Additionally, GPT (a generative pre-trained transformer) was employed after human analysis to validate the findings and assist in interpreting information about the UAE's national AI strategy in relation to citizen participation. This analysis allows us to gather insights into how citizen participation is conceptualised and represented within the UAE's National AI Strategy, providing valuable information about the strategy's approach to engaging citizens in AI-related initiatives and its implications for governance and societal development.

Table 1. Citizen participation keywords search and results in the national AI strategy of UAE (Source: Authors)

Keyword	Number of mentions	Quotations
1. 'Citizen(s)'	5	• 'The UAE has a vision to become one of the leading nations in AI by 2031 in alignment with the UAE Centennial 2071, creating new economic, educational, and social opportunities for citizens, governments and businesses and generating up to AED 335 billion in extra growth'. (p. 7) • 'Objective 4 explains how the UAE will take steps to increase the amount of government experimentation with AI to improve the lives of its citizens'. (p. 13) • 'Government can play a strong role in making sure AI delivers the greatest public value, by making citizens safer, healthier and happier'. (p. 30) • 'The X-Road platform in Estonia supports access and a combination of government and private databases, setting the stage for the application of machine learning tools. The data solution saves citizens over 800 years of working time per annum'. (p. 40) • 'Europe's General Data Protection Regulation includes new rights for consumers; it provides an opportunity to re-consider how consumer data is handled, even for customers who are not European citizens'. (p. 41)
2. 'People'	5	• 'We will transform the UAE into a world leader in AI by investing in people and industries that are key to our success'. (p. 3) • 'Therefore, the mission for this first Ministerial term is to transform the UAE into a world leader in AI by investing in the people and industries that are key to the UAE's success'. (p. 10) • '60 million people pass through Dubai Airport each year; 26 million pass through Abu Dhabi Airport'. (p. 22) • 'Jebel Ali port is the largest marine terminal in the Middle East and provides market access to over 2 billion people'. (p. 22) • '40% of the UAE workforce has good digital skills. This is less than the 56% of people with good digital skills in the UK, the top rated-nation in the AI-Readiness Index'. (p. 32)

3. 'Public'	10	• 'Public debates about AI often focus on whether or not it could take over important human decisions: from whether we go to war, to who receives medication'. (p. 13) • 'The UAE public sector is already a leader in smart public service delivery:' (p. 13) • 'Publicly accessible AI courses have already begun with large tech partners (objective 5)' (p. 18) • 'The UAI will consist of four levels of approval to include Public Sector Level, Private Sector Level, Institutional Level, and Product Level'. (p. 20) • 'These proof-of-concepts could be designed by the public sector, private sector or consortia'. (p. 25) • •'Government can play a strong role in making sure AI delivers the greatest public value, by making citizens safer, healthier and happier'. (p. 30) • •'With around 70% of Emiratis employed in the public sector, retraining of government workers is particularly critical'. (p. 32) • 'This will have a major impact on the public sector workforce and needs to be carefully managed, with a 2016 survey of Emirati workers finding the ideal future role for 54% was one in administration, i.e. a role that may not exist in the future'. (p. 32) • 'Public AI Training: Free courses are being run for UAE residents to raise awareness and understanding of AI technologies'. (p. 33) • 'Given the public sector is a major employer and potential user of AI in the UAE, The AI Office has also started specific training for government employees'. (p. 33)
4. 'Society'	7	• 'The UAE has set priority sectors – these will be the focus of initial activities. This does not mean that the UAE will stop working on AI solutions in other sectors where AI can deliver other benefits to society'. (p. 10) • 'There continues to be a range of views about the prospects of AI, and many potential future scenarios for AI in the UAE societies'. (p. 14) • 'The UAE government will play a direct role in designing and enabling AI systems that create the most value for society (objective 4)'. (p. 14) • 'For example, a new cyber research centre in Stuttgart and Tubingen, Germany (the Max Planck Society's Institute for Intelligence Systems) attracted foreign investment from Amazon, leading to an estimated 100 jobs over the next five years'. (p. 28) • 'The AI Office will also want to recognise and reward AI research with the greatest value to society'. (p. 39) • •'Innovations in AI technology often require rapid changes in regulatory settings and can create risks to society'. (p. 42) • 'A natural evolution of the Roundtable is in the formulation of an intergovernmental body, dedicated to

		providing a mechanism for governments and private companies to better understand AI and its impact on societies in order to help give a solid framework for future regulation'. (p. 43)
5. 'Participate/Participation'	2	• 'The AI Office is offering more advanced courses for government employees starting Q4 2018, focused on skills needed to work with them being the AI experts (ambassadors) in their entities. These require participants to complete a capstone project related to their current job'. (p. 33) • 'A programme to attract key AI thinkers to visit the UAE will be initiated. These key AI thinkers will participate in workshops and lectures with local universities and businesses'. (p. 39)
6. 'Democracy/ Democratic'	0	• No occurrences

3. Results

In this section, we present the findings from our exploration of how citizens were framed in the UAE's national AI strategy.

The term '**citizen**(s)' is mentioned five times in the UAE's NAS. In two of these instances, it is used in the context of discussing policies in other countries and thus does not directly refer to the citizens of the UAE (United Arab Emirates Minister of State for Artificial Intelligence Office, 2018, pp. 40–41). The other three examples illustrate the strategy's narrative regarding the role and position of citizens. For instance, the introduction of the document states the UAE's 'vision to become one of the leading nations in AI by 2031, creating new economic, educational, and social opportunities for **citizens**, governments, and businesses and generating up to AED 335 billion in extra growth' (p. 7). The quote from the UAE National AI Strategy clearly positions the deployment of artificial intelligence as a cornerstone of the country's future economic, educational, and social development. The strategy articulates a vision to transform the UAE into a leader in AI by 2031, a goal that aligns with the broader UAE Centennial 2071 ambitions. The focus on generating significant economic growth (up to AED 335 billion in extra growth) suggests a strong emphasis on AI as an engine for enhancing national prosperity. However, in terms of citizens and their participation, the quote doesn't explicitly mention how citizens will be involved in shaping or contributing to the AI initiatives. While it does indicate that the strategy intends to create opportunities for citizens, it doesn't specify mechanisms for their direct participation in decision-making or the development process. The narrative seems to lean more towards the benefits that citizens will receive from the implementation of AI rather than how they might actively participate in the strategy's execution or governance.

The strategy's emphasis on economic growth and educational opportunities without explicitly detailing ways for citizens to participate in governance or decision-making processes suggests a primary focus on the economic contributions of citizens. This perspective sees citizens mainly as beneficiaries and participants in the economic progress powered by AI rather than as active political agents shaping the direction of AI policy and its societal implications. This perspective is further supported by the other two references to 'citizens' in the strategy, which highlight that 'government experimentation with AI can improve the lives of its **citizens**' (p. 13) and 'government can play a strong role in ensuring AI delivers the greatest public value by making **citizens** safer, healthier, and happier' (p. 30). Both quotes highlight a vision where the application of AI is leveraged as a strategic tool to enhance the welfare and quality of life of citizens, reflecting a proactive approach to utilising technology for societal benefits. The first quote suggests that through experimentation with AI, the government aims to improve the lives of its citizens, indicating a focus on practical, real-world applications of AI to solve everyday problems and enhance overall living conditions. This reflects a commitment to leveraging advanced technology to address societal needs, which can be seen as a proactive approach to governance in the Digital Age. The second quote builds on this by stating the government's role in maximising AI's public value, aiming to make citizens safer, healthier, and happier. This articulation positions the government as a steward of technology, responsible for guiding AI development to ensure it benefits the public. It underscores a commitment to ensuring that AI advancements align with the broader goals of public safety, health, and happiness rather than merely economic growth or efficiency. However, these quotes also indicate a top-down approach to AI implementation, where citizens are primarily seen as recipients of AI's benefits rather than active participants in its governance or development.

The examination of the word '**people**' as a possible synonym for 'citizens' contributes little to answering this question. Although the word 'people' appears five times in the strategy document, its usage is exclusively in economic contexts. Notably, the strategy twice mentions 'investing in the **people** and industries that are key to our success' (pp. 3, 10). Policy measures related to this goal include attracting leading AI researchers, businesses, and innovators to the UAE (pp. 36–39), supporting AI-based projects in 'priority sectors' such as energy, logistics, and healthcare (pp. 22–25), and offering AI training for citizens so they can acquire the necessary skills for the future job market (pp. 32–35). While educating people on AI could be an important step toward more citizen participation, the strategy's primary motivation is to 'help them make better decisions in an economy where automation technologies enter the workplace' (p. 32). This statement suggests that citizens are primarily viewed as users/consumers of AI technologies. The strategy also refers to 'people' when discussing the country's role as a transit hub (p. 22) and the digital skills of the population (p. 32).

The term '**public**' (and variants) appears ten times in the document, six of which feature the phrase 'public sector' (plus another one referring to 'public service'). Most of these passages convey a similar message: the public sector needs to apply AI technologies to become more efficient, and its employees require training to adapt to the changing work environment (pp. 13, 30, 32, 33). Beyond public sector employees, the strategy also mentions that 'publicly accessible AI courses' open to all UAE residents have already been created with the goal to 'raise awareness and understanding of AI technologies' (pp. 18, 33). The emphasis on training and education in AI, as described in the strategy, may be more oriented towards harnessing economic growth potential rather than constituting a comprehensive approach to citizen outreach. The strategy's emphasis on enhancing public sector efficiency and providing AI education to boost skills aligns more closely with economic objectives rather than fostering broad-based citizen participation in policy development or decision-making processes. This focus reflects a drive to position the UAE as a leader in AI technology, which is expected to contribute significantly to economic development. While the provision of publicly accessible AI courses does indicate an element of outreach, the lack of mechanisms for citizens to engage meaningfully in shaping AI policies could imply that these efforts are more about preparing the workforce for an AI-driven economy rather than empowering citizens as active stakeholders in the AI discourse. Thus, it can be argued that the strategy, as it stands, leans more towards leveraging AI for economic benefits rather than cultivating a comprehensive, participatory approach to citizen engagement in AI governance.

Additionally, the word '**society**' appears seven times in the document, one instance of which can be disregarded as it refers to the German Max Planck Society (p. 28). The other six instances are similar in one important aspect: they do not portray society as an actor. Instead, there are three passages stating that AI can bring 'value to society' (pp. 14, 39) or 'benefits to society' (p. 10), one passage mentions 'risks to society' (p. 42), and another stresses the need to 'better understand AI and its impact on societies to help provide a solid framework for future regulation' (p. 43). The discussions of benefits and risks to society in the UAE's AI strategy remain abstract, with specific examples largely absent. This approach further underscores the emphasis on the economic impact of AI while potentially overlooking the broader political implications and the need for citizen involvement in shaping AI governance. It suggests a framing where society is viewed more as a recipient of AI outcomes rather than as an active participant or contributor in shaping AI policies.

While the term **'democracy'** (and its variants) is not featured at all in the strategy, the word **'participation'** (including its variants) is used only two times in the document: once in reference to training for government employees (United Arab Emirates Minister of State for Artificial Intelligence Office, 2018, p. 33) and once regarding AI research workshops and conferences (p. 39). The two quotes from the UAE National AI Strategy provide insights into the strategy's focus on developing AI expertise and cultivating a knowledge base within the country. However, they also reflect a specific framing of citizen participation in the AI landscape, which seems limited to certain segments of the population.

The first quote highlights an initiative aimed at training government employees to become AI experts within their respective entities. This is a strategic move to embed AI knowledge and capability directly into government operations, ensuring that the public sector is equipped to leverage AI technologies effectively. The requirement for a capstone project that relates to their current jobs suggests an applied learning approach, where government employees not only learn about AI but also implement AI solutions within their roles. While this enhances the capabilities of government workers and integrates AI into public administration, it does not explicitly extend AI education or participation opportunities to the general public or broader citizenry.

The second quote focuses on attracting AI thinkers and describes a programme designed to attract international AI experts to the UAE. These experts are expected to engage with local universities and businesses through workshops and lectures, which can foster a vibrant AI ecosystem and stimulate local innovation. This approach is beneficial for academic and business communities, but again, the focus is more on elite or professional circles rather than the general citizenry. The strategy, as outlined in this quote, focuses on enhancing expertise and networking at a high level without direct involvement from the broader public.

These initiatives suggest that the UAE's AI strategy is oriented towards creating a highly skilled workforce and an elite cluster of AI professionals, which can drive the country's ambitions to become a leader in AI technology. However, they also indicate a top-down approach where citizen participation in AI development is limited to specific professional groups (government employees, academics, business leaders). The broader public seems to be more of a recipient of the benefits of these initiatives rather than an active participant in shaping AI policy or contributing to the AI discourse. The strategy's focus on professional development and international collaboration is crucial for achieving technological advancement and economic growth, but there could be a greater emphasis on inclusive policies that engage all levels of society. This approach could include more public education initiatives, open forums for policy feedback, and mechanisms that encourage citizen input in shaping how AI is integrated into daily life and governance. Such inclusive approaches would not only democratise the benefits of AI but also ensure that its implementation reflects the diverse needs and values of all citizens.

The UAE National AI Strategy, while ambitious in leveraging AI to drive economic, educational, and social advancements, predominantly positions citizens as beneficiaries rather than active participants in the AI-driven transformation. The strategy details specific training for government employees and initiatives to attract global AI experts, with an emphasis on professional development and international collaboration aimed at economic growth. However, it largely overlooks broader citizen involvement in decision-making processes and public discourse on AI policy. The strategy's focus is on enhancing sectoral efficiency and individual skill sets in alignment with AI advancements, yet it lacks concrete mechanisms for inclusive citizen engagement, which could otherwise enrich the policymaking process and ensure that AI development aligns with the wider societal needs and democratic values. The limited portrayal of citizens as active stakeholders in shaping AI policy underscores a need for a more participatory approach that embraces all segments of society, ensuring that AI benefits are not only received but also shaped by citizens.

In sum, we can draw two main conclusions from our analysis. First, the primary focus of the UAE's AI strategy appears to be boosting economic growth, with citizen engagement and participation largely side-lined. Secondly, when citizens or society are mentioned in the document, they are depicted as workers, users, or consumers rather than as political actors capable of voicing their opinions and holding the government accountable.

4. Discussion: The passivisation of citizens through the use of technology

The above analysis uncovers insights about how citizens are envisioned within the strategy as passive recipients rather than active participants. In essence, the strategy acknowledges the profound implications of AI for society but does not detail mechanisms through which society could actively participate in the AI discourse, decision-making process, or the AI life cycle. The strategy highlights the positive outcomes of AI for citizens but does not explicitly address how citizens can influence or participate in decision-making processes regarding AI use. This situation might suggest that the strategy envisions a model where the government dictates the trajectory of AI development and deployment, with society and citizens playing a more passive role. For example, the strategy's references to AI bringing 'value' and 'benefits' to society, while positive, depict society as a passive entity merely receiving the advantages generated by AI technologies. Similarly, mentions of 'risks to society' position society as a potential victim of AI's negative impacts rather than as an engaged stakeholder in mitigating these risks. Additionally, the call to 'better understand AI and its impact on societies' (p. 34) in the strategy to inform future regulation acknowledges the need for informed governance structures. However, this does not explicitly suggest a role for society in the regulatory process. Simultaneously, while the strategy positions citizens as central to its success through their roles in the AI-enhanced economy, it appears to limit their participation to the roles of users and beneficiaries of AI technology rather than as co-creators or governors of AI policy. Therefore, the framing of citizens and their participation in the UAE strategy suggests a top-down approach where citizens are positioned as beneficiaries rather than active participants. This approach could impact the public's perception and reception of AI initiatives, influencing the degree to which AI technologies are accepted and integrated into society.

Therefore, the strategy's language suggests a view of society as an external entity to be protected and benefitted rather than as a collaborative partner in the AI development process. This interpretation could imply a need for the strategy to evolve towards more inclusive governance models that actively involve various societal stakeholders, not only as beneficiaries but also as participants in shaping AI's integration and regulatory frameworks. This inclusion could help ensure that AI developments align more closely with societal needs, values, and aspirations, creating a more dynamic and participatory approach to AI policy and implementation.

Reflecting on this, future iterations of the strategy should include clearer roles for citizen engagement and participation to ensure that AI development aligns with the public's needs and values, thus fostering a more inclusive AI-driven future. Ultimately, while the government plays a central role in driving the UAE National AI Strategy, achieving its objectives requires collaboration and participation from all segments of society. By fostering a culture of inclusivity, transparency, and accountability, the UAE can harness the full potential of AI technologies to promote sustainable development, economic prosperity, and societal well-being for all its citizens. For a more holistic and participatory approach, future iterations of the strategy could benefit from including mechanisms for citizen engagement in AI governance, ensuring that AI development not only serves but also reflects the public's needs and values.

The view of citizens as mere recipients suggests passivity, focusing on the idea of receiving outcomes or benefits without participating in their creation. In fact, achieving successful AI implementation in the public sector requires a collaborative effort between public authorities and civilians (European Commission, 2018; AGID, 2018). This emphasises that citizens are at the end of decisions, merely benefiting from or suffering the consequences of policies they had no hand in shaping. The current language underscores the unilateral nature of decision-making and implies that actions are done for or to the citizens, who have little to no role in influencing those actions.

Given the two main conclusions outlined previously – the UAE's AI strategy's primary focus on economic growth with limited emphasis on broader citizen engagement and participation and the frequent depiction of citizens more as workers, users, or consumers rather than active participants in decision-making – it appears that the strategy contributes to a passive role for citizens regarding their engagement with technology. The strategy's emphasis on economic growth and the utilisation of AI technologies primarily position citizens as recipients or end-users rather than active contributors or decision-makers. This approach to technology use effectively passivises citizens, limiting their role to that of beneficiaries who do not actively participate in shaping the technological landscape or policies. This finding highlights how AI and technology generally have the potential to be used to reduce the active role of individuals in shaping policies or technological deployments, relegating them to more passive roles as recipients or end-users without significant influence or control. This suggests a critical view of how technology, while potentially empowering, can also be structured to limit public engagement and maintain control within a select group of decision-makers.

Hence, we posit that this process can be seen as '**digital passivisation**', a concept that can describe a scenario where citizens are increasingly placed in passive roles within digital strategies, initiatives, and solutions. It refers to situations where technology is implemented in ways that do not actively involve citizens in decision-making processes, treating them not as active agents but as recipients or end-users of the technology without significant input or control. This concept highlights a shift from active participation to passive reception in the context of digital transformation, particularly in how citizens interact with and influence public policies or technological developments. In the context of a national AI strategy, 'digital passivisation' would emphasise how the deployment of AI technologies and related policies might be structured to keep citizens at the margins of influence, focusing instead on economic growth or other objectives where citizen input is minimal or largely ceremonial.

As AI becomes increasingly pervasive in various aspects of daily life, efforts to engage citizens in discussions about AI governance, ethics, and societal impacts become imperative. Empowering individuals with knowledge about AI technologies and involving them in dialogues about its implications can help cultivate a more informed and inclusive approach to AI governance as well as reduce the intensified digital divide. This strategy may involve initiatives such as public consultations, citizen forums, and educational campaigns aimed at raising awareness and fostering dialogue around AI-related issues throughout the AI life cycle (see De Silva & Alahakoon, 2022). Future research could utilise critical discourse analysis to explicate the power dynamics in the development of the AI strategy and, more importantly, the elements of power (e.g. between the government and citizens) embodied in the language of the strategy. Highlighting 'power' as the focal point in the analysis would aid in reflecting on the political consequences of the strategy.

5. Conclusion

The analysis of the UAE's national AI strategy reveals a predominant focus on economic growth, positioning citizens primarily as beneficiaries rather than active participants in the AI-driven transformation. The strategy emphasises enhancing sectoral efficiency and individual skill sets in alignment with AI advancements but lacks concrete mechanisms for inclusive citizen engagement. This limited portrayal of citizens as active stakeholders in shaping AI policy underscores a need for a more participatory approach that embraces all segments of society.

Two main conclusions can be drawn from the analysis. (1) Economic Focus: The primary focus of the UAE's AI strategy is on boosting economic growth, with citizen engagement and participation largely sidelined. The strategy highlights the benefits AI can bring to society, but these are often framed in terms of economic contributions rather than political or participatory roles for citizens. (2) Citizens as Beneficiaries: When citizens or society are mentioned, they are depicted more as workers, users, or consumers rather than political actors capable of influencing decisions and holding the government accountable. This perspective suggests a top-down approach where citizens are positioned as recipients of AI benefits rather than active participants in its governance or development.

The concept of digital passivisation is introduced to describe this phenomenon, where technology is implemented in ways that do not actively involve citizens in decision-making processes, treating them as passive recipients of the technology. This approach can limit public engagement and maintain control within a select group of decision-makers, emphasising economic growth over democratic participation.

To foster a more inclusive AI-driven future, the strategy should evolve to include clearer roles for citizen engagement and participation, which could involve public consultations, citizen forums, and educational campaigns aimed at raising awareness and fostering dialogue around AI-related issues. By empowering individuals with knowledge about AI technologies and involving them in discussions about its implications, the UAE can ensure that AI development not only serves but also reflects the public's needs and values. This approach would help bridge the digital divide and create a more dynamic and participatory AI policy framework.

Acknowledgements

We extend our gratitude to Zeinab Al Masry for her invaluable research assistance during her internship at Impact Circles e.V., Berlin. Additionally, we thank Ola Khaled for her diligent research support. Our appreciation also goes to the reviewers of the Weizenbaum Journal and other venues where this research was submitted, including the reviewers of the Participatory Design Conference 2024 and others. Special thanks to all the students of the 'Participation in the Age of Digital Innovation' seminar at Duisburg-Essen University in WS 22/23 for their insightful discussions.

References

Amboro, K., Kuswono, K., & Setiawan, A. (2022). Identification of cultural values and nation's character. *Dalung bojong inscription of the Sultanate of Banten in Lampung. JUSPI (Jurnal Sejarah Peradaban Islam*), *5*(1), 1–12. http://dx.doi.org/10.30829/juspi.v5i1.9471

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values, 47*(5), 855–881. https://doi.org/10.1177/01622439211030007

Battina, D. S. (2017). Research on artificial intelligence for citizen services and government. *International Journal of Creative Research Thoughts*, *5*(2), 769–773. http://www.ijcrt.org/papers/IJCRT1134263.pdf

Bovens, L. (2009). The ethics of nudge. In *Preference change: Approaches from philosophy, economics and psychology* (pp. 207–219). Springer Netherlands. https://doi.org/10.1007/978-90-481-2593-7_10

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

De Silva, D. & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, *3*(6).

Dutton, T. (2018). An overview of national AI strategies. *Politics+ AI*, *28*. Retrieved from https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd (Accessed 7 May 2024).

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Engstrom, D., Ho, D., Sharkey, C., & Cuéllar, M. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3551505

Entman, R. M. (1993). Framing: Toward clarification of A fractured paradigm. *Journal Of Communication*, *43*(4), 51–58.

European Commission. (2018). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: *Artificial Intelligence for Europe*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN (Accessed: 6 May 2024).

European Commission. (2022). Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions. *European Commission: Brussels, Belgium*. https://www.reteambiente.it/repository/normativa/48230_comunicazione18_maggio_2022repowereu.pdf (Accessed 4 May 2024).

Havrda, M. (2020). artificial intelligence's role in community engagement within the democratic process. *International Journal of Community Well-Being*, *3*(4), 437–441. https://doi.org/10.1007/s42413-020-00100-8

Hendawy, M. (in press), The intensified digital divide: Comprehending genAI. *Internet Policy Review*.

Hendawy, M., & Kumar, N. (in press). How AI is portrayed in national AI strategies of the Arab region. *Arab Reform Initiative*.

Hendawy, M., & Ansari, Z. (2023). AI-citizens' relationship: Analyzing Egypt and Mauritius national AI strategies. *Working Paper Series*, Witz University.

Herasimenka, A. (2020). *Adjusting democracy assistance to the age of digital dissidents*. German Marshall Fund of the United States.

Krippendorff, K. (2018). Content analysis: An introduction to its methodology. Sage Publications.

Kuhn, T. (2017). Digitaler Zwilling. *Informatik Spektrum* 40, 440–444. https://doi.org/10.1007/s00287-017-1061-2

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, *28*(4), 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x

Lords, H. O. (2018). AI in the UK: Ready, willing and able? https://www.politico.eu/wp-content/uploads/2018/04/AI-in-the-UK-ReadyWillingAndAble-April-2018.pdf (Accessed 4 May 2024).

Mayring, P. (2004). Qualitative content analysis. *A Companion To Qualitative Research*, *1*(2), 159–176.

Milano, M., O'Sullivan, B., & Gavanelli, M. (2014). Sustainable policy making: A strategic challenge for artificial intelligence. *Ai Magazine*, *35*(3), 22–35. https://doi.org/10.1609/aimag.v35i3.2534

Minsky, M. L. (1969). *Semantic information processing*. The MIT Press.

Misuraca, G., & van Noordt, C. (2020). Overview of the use and impact of AI in public services in the EU. *Publications Office of the European Union: Luxembourg*. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC120399/jrc120399_misuraca-ai-watch_public-services_30062020_def.pdf (7 May 2024).

National Science & Technology Council. (2016). The National Artificial Intelligence Research And Development Strategic Plan. *National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee.* http://large.stanford.edu/courses/2018/ph241/cheng1/docs/ai-eop-oct16.pdf (Accessed 4 May 2024).

Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage Publications.

Niti, A. (2018). National Strategy For Artificial Intelligence. *Government of India/NITI Aayog, New Delhi, India.*

OECD.AI. (2022). UAE National Strategy For AI. Retrieved from https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-25387 (Accessed 7 May 2024).

Paltieli, G. (2022). The political imaginary of national AI strategies. *AI & SOCIETY*, *37*(4), 1613–1624. https://doi.org/10.1007/s00146-021-01258-1

Pan, Y. (2016). Heading toward artificial intelligence 2.0. *Engineering*, *2*(4), 409–413. https://doi.org/10.1016/J.ENG.2016.04.018

Papadopoulos, T., & Charalabidis, Y. (2020). What do governments plan in the field of artificial intelligence? Analysing national AI strategies using NLP. In *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance* (pp. 100–111). https://doi.org/10.1145/3428502.3428514

Reis, J., Santo, P. E., & Melão, N. (2019). Impacts of artificial intelligence on public administration: A systematic literature review. In *2019 14th Iberian Conference On Information Systems And Technologies (CISTI)* (pp. 1–7). IEEE. https://doi.org/10.23919/CISTI.2019.8760893

Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019). *Analyzing Media Messages: Using Quantitative Content Analysis In Research*. Routledge. https://doi.org/10.4324/9780429464287

Rowe, G., & Frewer, L. J. (2005). A typology of public engagement mechanisms. *Science, Technology, & Human Values*, *30*(2), 251–290. https://doi.org/10.1177/0162243904271724

Sharma, G. D., Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures*, *2*, 100004. https://doi.org/10.1016/j.sftr.2019.100004

Tolbert, C. J., & McNeal, R. S. (2003). Unraveling the effects of the Internet on political participation? *Political Research Quarterly*, *56*(2), 175–185. https://doi.org/10.1177/106591290305600206

United Arab Emirates Minister of State for Artificial Intelligence Office. (2018). *UAE National Strategy for Artificial Intelligence 2031*. Retrieved from https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf (Accessed 7 May 2024).

Valle-Cruz, D., & Sandoval-Almazan, R. (2018). Towards an understanding of artificial intelligence in government. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (pp. 1–2). https://doi.org/10.1145/3209281.3209397

Wilson, C. (2022). Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly*, *39*(1), 101652. https://doi.org/10.1016/j.giq.2021.101652

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector – Applications and challenges. *International Journal of Public Administration*, *42*(7), 596–615. https://doi.org/10.1080/01900692.2018.1498103

Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, *16*(3), e12962. https://doi.org/10.1111/soc4.12962

Empowering smart regions: addressing challenges and leveraging enabling factors in municipal digital transformation

Bonny Brandenburger, Andrea Hamm, Caroline Krohn, Felix Sühlmann-Faul, Manuel Atug, Nicole Döpp, and André Ullrich

Abstract: The advancing digital transformation of society creates a wide range of opportunities for improved access to information and resources that contribute to ensuring the availability of public services and the development of sustainable living spaces. This underlying potential does not only apply to urban areas; digitalisation projects are also being implemented in rural municipalities in order to exploit the potential of digital transformation. Nevertheless, the field of so-called smart regions has yet to receive substantial focus in research. To understand the specific challenges and enabling factors of digital transformation activities in urban-rural areas, a workshop was organised with the cooperation of municipal representatives of a model region in Schleswig-Holstein, Germany. Therein, specific technical, economic and social challenges as well as enabling factors of municipal digitalisation projects aimed at developing smart regions were identified. The results show that the success of digital transformation measures in urban-rural areas is not only determined by the expansion of a corresponding technical infrastructure but in particular by the acceptance of citizens and municipality employees, as well as economic viability. This research further informs municipalities and future researchers on the critical factors required to effectively conduct digitalisation projects in the smart region context.

Keywords: digital transformation, smart region, challenges and enabling factors, sustainable regional development

1. Introduction

When it comes to fully exploiting the potential of digital transformation (DT), municipalities are faced with numerous challenges, mainly due to limited resources (Fila et al., 2023). DT creates opportunities to improve and guarantee access to public services (van der Hoogen, 2024) and expand smart living spaces while maintaining a focus on ensuring the greatest possible added value for citizens (Heuser et al., 2020, p. 15; Sahu et al., 2020). In this respect, we understand DT as a holistic organisational approach to using and providing digital technologies within and for municipal administrations, emphasising the imperative for ongoing adaptations of processes, services and products to meet external demands (Mergel et al., 2019, p. 10). Thus, DT shall improve relationships between public administrations and their stakeholders, increase citizen satisfaction and foster a shift in organisational culture (Mergel et al., 2019, p. 10). To understand and cope with the challenges of DT, Lafioune et al. (2023) developed a framework based on a systematic literature review that included an investigation of multiple concepts concerning DT and the implications for municipalities. Applying this framework enables the identification of predominant issues hindering municipalities' DT.

While the relationship between the challenges and enabling factors of smart city projects has already been researched thoroughly (Andejany et al., 2023; De Guimarães et al., 2020; Kogan & Lee, 2014), research focusing specifically on **smart regions** and their prerequisites for a successful DT is still in its infancy (Geisberger, 2023, p. 31; Matern et al., 2020). Municipalities with lower-density populations and less infrastructure have even fewer human and financial resources (Fila et al., 2023), which makes DT measures more difficult to implement (Jakob & Krcmar, 2018). Usually, a critical mass of users is required in order to establish and maintain existing and new technical and social infrastructure (Ruscheinski, 2023, p. 19). Against this background, this paper is guided by the following research question:

How can the challenges for municipalities on their way to smart regions be tackled and suitable support measures be designed?

Lafioune et al.'s (2023) framework was developed by focusing on larger urban contexts, so we sought to determine to what extent the framework model is applicable for planned and promoted DT in urban-rural areas, where initial conditions are different. We accessed the field by approaching municipal administrators currently realising DT in their day-to-day practices. We conceptualised and held a workshop with municipal representatives from a model region and discussed their most recently implemented DT projects. The framework developed by Lafioune et al. (2023) served as the theoretical and conceptual basis used to identify the challenges. The workshop results were then analysed in the context of the existing academic discourse on DT. This article will provide two main contributions: first, we will identify the challenges of DT in urban-rural areas based on the conducted fieldwork and examine how these challenges can be addressed by identifying enabling factors and suitable support measures. Second, the applied theoretical framework will be tested and discussed in terms of its possible utilisation in further studies on urban-rural contexts. DT is not just a matter of solving urban problems through technology, and because cities cannot be pictured as isolated entities (Amin & Thrift, 2002), we believe that the identified factors and presented measures can guide future **smart region** projects.

2. Towards a deeper understanding of smart regions

For more than twenty years, concepts centred on digital technologies for the design of urban spaces have been discussed under the umbrella of **smart cities** (Geisberger, 2023, p. 27). This term not only refers to the possibilities of information and communication technologies (ICT) but also increasingly focuses on the economic, ecological and social aspects present in the course of the DT. With the beginning of the 21st century, cities are no longer understood as isolated entities but rather as regionalised spaces (Amin & Thrift, 2002; Brenner and Schmid, 2015; Soja, 2000); urban and rural areas do not constitute two distinct territories that can be considered separate from one another (Delgado-Viñas & Gómez-Moreno, 2022). On the contrary, the classifications of **urban or city** and **rural** are to be replaced with the term **smart region**, and their existing interdependencies are to be examined. In recent years, there has been an increasing amount of literature on DT strategies within the political and administrative boundaries of cities (Ben Letaifa, 2015; Masik et al., 2021; Shamsuzzoha et al., 2021; Slavova & Okwechime, 2016;), while in contrast, concepts for **smart regions** are still in their infancy (Geisberger, 2023, p. 31). Matern et al. (2020, p. 2064) define **smart regions** as:

> diverse **urban-rural areas** that are spatially reframed by digital technologies and the respective social practices in a variety of fields (citizenship, governance, economy, environment, mobility, infrastructure) on a discursive, implemental and regulative level. The concept of **smart regions** follows a relational and social constructivist understanding of spaces and

> emphasi[z]es an integrated approach towards the social (re-)construction of **smart regions** by actors and their networks.

Thus, the concept of a **smart region** bridges urban and rural environments and has a wide conceptual overlap with **urban-rural municipalities**. A fundamental aspect lies in fostering a collaborative approach so individuals impacted by DT are empowered to actively shape their communities through ICT. Central to the notion of a **smart region** is the deliberate and thorough integration of various sectors, institutions, themes, administrative bodies and diverse social groups through ICT infrastructure (Will, 2021, p. 459). Here, the essence of **smartness** transcends mere technological solutions, encompassing a broader capacity for integrating diverse stakeholders into a region's innovation ecosystem instead of thinking only within urban boundaries (Markkula & Kune, 2015). This holistic perspective not only leverages the potential of ICT but also increasingly emphasises the sustainable economic and social dimensions within the framework of DT.

a) Challenges for digitalisation projects in smart regions

The increasing dynamics of peripheralisation in rural areas and centralisation in metropolitan areas are reducing the economic performance of rural regions (Ruscheinski, 2023, p. 12), resulting in fewer companies choosing to settle there. This, in turn, can lead to structural economic weakness; furthermore, an ageing society due to declining birth rates is also characteristic (Statistisches Bundesamt, 2023) that results, among other things, in a declining capacity of technical and social infrastructure. Krishnan et al. (2020) identify the following aspects as specific challenges in the DT of municipalities: a lack of control over sustainability activities, a lack of methods and processes to promote the participation of citizens and insufficient knowledge about structures and value creation through ecosystems. In municipalities with a low population density, the question is not only how systematic and highly integrated networking using ICT can contribute to ensuring public services, but also to what extent the associated expansion costs are profitable (Geisberger, 2023, p. 37). The functional, financial and personnel requirements in rural areas therefore confront decision-makers with challenges different than those they would face in a solely urban context (Ruscheinski, 2023, p. 14). Nevertheless, rural regions in particular represent a special opportunity, as they can function not only as places to work and live but also as recreational areas and cost-effective business locations (Geisberger, 2023, p. 31). Against this background, research into **smart regions** is particularly complex. A large number of heterogeneous initial conditions (such as population growth versus migration, public transport infrastructure versus individual mobility) must be taken into account when implementing DT projects to ensure the provision of general-interest services in urban-rural municipalities.

b) Enabling factors for municipal digitalisation projects

In the context of municipal digitalisation projects, citizen participation, regulatory framework conditions, funding and stable telecommunications infrastructure are grouped under the category of environmental enabling factors (Jonathan, 2020); this approach can also be found in more recent works (cf. Gudmundsdottir et al., 2024; Lafioune et al., 2023). In the information technology category, interoperability, data security, IT architecture and data-driven agility are emphasised (Jonathan, 2020). Organisational and management factors include a sufficient degree of business IT alignment, management commitment, further training opportunities for employees or an organisational culture geared towards change. Gudmundsdottir et al. (2024) emphasise the significance of managers and middle managers in actively shaping and taking responsibility for municipal DT processes. Furthermore, the critical nature of sustainability in resource management and the innovation in service provision associated with DT were identified as important determinants (Gudmundsdottir et al., 2024). Access to the internet independent of time and place is also emphasised as an essential prerequisite for success, as it creates individual, social and economic development opportunities, as well as possibilities for participation (Geisberger, 2023, p. 37). A tangible added value for the local population is also required for the successful implementation of digital processes or solutions (Heuser et al., 2020, p. 15). In this regard, identifying relevant social groups is necessary for the social construction of technology. Different social groups attribute different meanings to technology depending on its perceived usefulness or ease of use (Davis, 1985; Pinch & Bijker, 1984). Thus, during a participatory negotiation process between these groups, the design of the technology continues until a consensus is reached, which occurs when the negotiation process has ensured that the technology works for all groups involved (Bijker, 1995; Pinch, 1996; Pinch & Bijker, 1984, p. 411ff.;). Based on this understanding, it is clear that the success of

DT in terms of digital artefacts stands and falls with the participation of all affected groups in the process of creating the technologies. This means that the perspectives of providers and users are incorporated into the creation process, which significantly increases the subsequent acceptance of the technology.

3. The Süderbrarup case study and methodological approach

To identify challenges and enabling factors in digitalisation projects, group discussions were held in December 2023 as part of a half-day workshop. In addition to the two researchers who acted as moderators, a total of 12 participants took part. The participants were representatives from the rural municipality of Süderbrarup (individuals from the Süderbrarup Smart City office and two mayors) and representatives from the areas of the smart border region, regional development and digitalisation offices (Chief Digital Officer, IT and Digitalisation Department) in the district and from an independent city in the region. This group of participants specifically invites a view of DT processes beyond the urban context. The Süderbrarup model region has already published a Smart City strategy[1,2], launched a wide range of digitalisation projects and thus taken the first steps towards becoming a digital pioneer (Office Süderbrarup, 2021). Both the rural structure and the transformation impulses make the Süderbrarup model region a particular object of investigation in relation to the realisation of digitalisation projects towards a **smart region**.

After collecting information on the participants' expectations at the beginning of the workshop, their views on the specific challenges of digitalisation projects were gathered. Current digitalisation projects were discussed first. The key questions were 'Which project are you currently working on?' and 'Which digitalisation projects are taking place in your area of work?'. Challenges were then determined by asking, 'What are the specific challenges in the respective projects?' In Block 2 of the workshop, enabling factors for digitalisation projects were identified. Based on the key question, 'What makes a digitalisation project actually successful?', the participants had the opportunity to make various suggestions using moderation cards.

1 https://smartcityamtsuederbrarup.de/wp-content/uploads/2021/05/Finale_Strategie.pdf
2 The use of the term Smart City was prescribed by the funding guidelines of the state of Schleswig-Holstein, with no distinction made between city and region.

Individual suggestions were illustrated using examples from the aforementioned projects and discussed during the presentation so that a common understanding and agreement could be reached. The participants then evaluated the identified enabling factors. Each participant had the opportunity to award three points so that the factors could be ranked according to their perceived relevance.

During the event, the two moderators recorded their general observations as well as the content developed during the discussions; the data were then collected, modified, synthesised and discussed thematically using whiteboards. Following the workshop, the participants' responses were consolidated and deductively categorised using Lafioune et al.'s (2023) framework for the DT of municipalities. The framework consists of 22 elements that are clustered into six overarching categories: strategy, process management, organisational culture, policy, network and digital information ecosystem. This categorisation of challenges is also in line with recent research on challenges at the municipal level (van der Hoogen et al., 2024). For example, municipal representatives have reported that they lack an overview of the individual digitalisation measures in the administration. These challenges were assigned to the third category – *Organisational culture* and *3.2 Lack of communication* – in the DT framework.

4. Results

The interviewees identified a total of 11 different *inter*- and *intra-organisational digitalisation projects*. The term *inter-organisational projects* describes projects that aim to create a network of relationships among municipalities, civil society and the economy that extends beyond any one organisation (Zundel, 1999, p. 19). In contrast, *intra-organisational projects* encompass network-like patterns of relationships within an organisation or community. This section describes the challenges and the enabling factors identified in the course of the discussions on digitalisation projects and then explores the connections between them.

a) Challenges of municipal digitalisation projects

The participants identified the processes for administrative digitalisation and the establishment of a central database as well as an electronic document management (EDM) system for the municipality of Süderbrarup as challenges in the context of *intra-organisational digitalisation projects*. These projects aim projects to convert workflow and communication channels into digital technologies within the administration, thereby improving access to information and increasing the effectiveness of decision-making processes. For example, in August 2020, a project implementing the EDM system in the Süderbrarup region was completed. The primary goal was to store all documents digitally within one system. The system includes forgery-proof file storage and the application of an automated stamp on documents. In addition, a dynamic filing system for managing the lifecycle of documents before electronic archiving is integrated. Thus, every change to a document can be retraced; this is intended to prevent forgeries. Furthermore, minutes of meetings and dates are made available or are announced to citizens via the council information system. An additional feature of the EDM system is the creation of electronic files that consolidate information gathered from diverse sources, meaning that various documents relating to a citizen can be collected in one folder, even if the responsibilities within the municipal administration (e.g. the tax office or civil engineering office) are different.

Acceptance of change by administrative staff turned out to be a decisive factor when facing challenges while implementing the three *intra-organisational digitalisation projects*. Technical change requires a cultural flux in terms of individual willingness to change. The simultaneous nature of these two transformation endeavours was described as particularly challenging. This aspect can be found in Lafioune et al. (2023) under 3.1 *Lack of organisational change management*. One participant reported that some old administrative processes continued to exist alongside new ones, leading to unnecessary additional work. In the specific example with the EDM system, participants reported that there were administrative employees who still tended to use the analogue document filing system while others used the new system and thus experienced the associated simplification of administrative processes. In addition, interviewees mentioned that administrative staff lacked an overview of the particular digitalisation measures in the municipality in order to be able to assess what was already running successfully (3.2 *Lack of communication*). The lack of personnel (5.3 *Lack of human resources*) in municipal administrations was also emphasised as another major challenge in the course of DT. In the eyes of the interviewees, the implementation of transformation projects often depends on a few people who frequently have to perform other tasks that are not a part of the implementation of **smart region** projects.

In the area of *inter-organisational digitalisation projects*, software-based applications for communication between the administration and citizens were mentioned most frequently by the participants. These include the *MensaMax*[3] (a digital application for organising school lunches), a booking platform, a data platform for exchanging and sharing data and information with citizens and a mobility app called *smartes DorfSHUTTLE*[4]. Local public transport in Süderbrarup is currently designed primarily for the transportation of schoolchildren. Outside of school hours, there are very few connections between municipalities. The mobility app seeks to mitigate this problem by supporting demand-oriented transportation in the form of a ride-pooling-on-demand service. In addition, general administrative services, such as applying for an ID card, are to be digitalised as part of the Online Access Act (OZG), thereby simplifying the process for citizens.

The workshop participants identified challenges in both social and technical aspects of DT projects; these areas are often interlinked. For example, the issue of technical interfaces was raised (6.2 *Incompatibility and/or inadequacy of existing information systems*) in the context of the need for cross-service software solutions that make it easier for citizens to use different areas of general-interest services (e.g. the financial or citizens' office). There are often different applications and standards that, from the perspective of municipal representatives, lead to conflicts between administrative staff and citizens. The usability of these digital applications is often unfriendly to consumers (6.1 *Inadequate technologies and tools*), resulting in acceptance problems among the population (5.2 *Resistance to change* and 5.4 *Perception by the citizens*). This social challenge is exacerbated by the issue of data protection and security (2.2 *Absence of data management*). Citizens do not recognise the exact purpose for which data are collected and what added value and danger this poses. There is no explanation of the benefits of data acquisition and analysis for local civil society (3.2 *Lack of communication*). The municipal representatives emphasised the need for the participation and early involvement of citizens in the development of digital applications, such as data or booking platforms. Thus, both internal and external communication were mentioned as major challenges in the realisation of software-based digitalisation projects.

3 https://mensamax.de/schule/
4 https://smartes-dorfshuttle.de/

Another major digitalisation project in the municipality of Süderbrarup is the *makerspace*[5], a smart city lab that opened in 2022. It serves as a physical meeting place and a space for development and experimentation for citizens and businesses, where digital tools for prototyping, crafts and IT are available and can be tested. Additionally, the lab is used for training courses to promote digital skills (Amt Süderbrarup, 2021, p. 49). In order to be able to present *makerspace* in a decentralised manner and to bring its offerings to surrounding areas, the project aims to create digital hubs, e.g. through a temporary version of the entire offering. The extracurricular activities available through *makerspace* and the youth centre are communicated via social media channels to reach young people.

It can be emphasised that these physical places – *makerspace* and the youth centre – do not sufficiently address the needs of the local population; for example, the added value of *makerspace* is hardly appreciable for older people. In the local population, a lack of communication gives rise to fears. Another central challenge is the funding logic, which is associated with a limited project duration and dependence on third-party funding. The lack of financial sustainability can also be emphasised in the context of the other digitalisation projects (1.6 *Funding problems* and 4.1 *Insufficient policy and support from elected officials*). Moreover, the participants reported that digitalisation projects require time beyond the funding period for people to recognise the added value for themselves (5.4 *Perception by the citizens*).

b) Enabling factors of municipal digitalisation projects

The participants repeatedly brought up social and organisational issues when discussing enabling factors. With a score of five, the two highest-ranking enabling factors were (1) 'making life easier by creating added value for citizens' and (2) 'economic viability'. These two factors coincide with the challenges described above, in which the perceptions of citizens and sustainable funding problems were also strongly emphasised. In addition, 'pride in the project' on the part of those involved was underlined, receiving four points. This goes hand in hand with the importance of an organisational culture in which all participants see themselves as part of a successful project. With three points each, respondents mentioned 'high usage figures', 'the fulfilment of project objectives' and 'the subsequent use

5 https://www.diz.digital/ueber-das-diz

of infrastructure' even after the end of the project as important for determining its success. These factors were applied both to internal digitalisation projects within the administration and those with external interest groups. In addition, 'user satisfaction', 'quick decisions' and 'good design and planning of the project' were emphasised as key factors for success, receiving one point each. All in all, the participants described the importance of bundling needs in order to demonstrate added value, ensure acceptance among the citizenry and guarantee that technical measures, such as the MensaApp, are also used.

c) Connecting challenges and enabling factors

In order to demonstrate the connections between the *intra-* and *inter-organisational projects*, the identified challenges and enabling factors were transferred into a superordinate systematic structure (Figure 1) utilising Lafioune et al.'s (2023) six categories. For example, the participants discussed the lack of sufficient clarity necessary for citizens to understand what happens to their data in the newly introduced data platform; this challenge was categorised under 2.2 *Absence of data management* and 3.2 *Lack of communication*. The enabling factors are highlighted in grey at the bottom of the figure. In order to meet the challenges mentioned regarding the data platform, for example, improvement at the planning stage of the digitalisation project by involving various user groups ('good design and planning of the project') was suggested. The superordinate systematic structure therefore illustrates the relationships between the challenges and enabling factors, which can help enable decision-makers to identify the most frequently mentioned challenges within various digitalisation projects by means of a systematic presentation and thus develop initial solutions using the identified enabling factors. The example of the Süderbrarup municipality shows that a lack of communication, especially with citizens, is a major hurdle in the implementation of digitalisation projects.

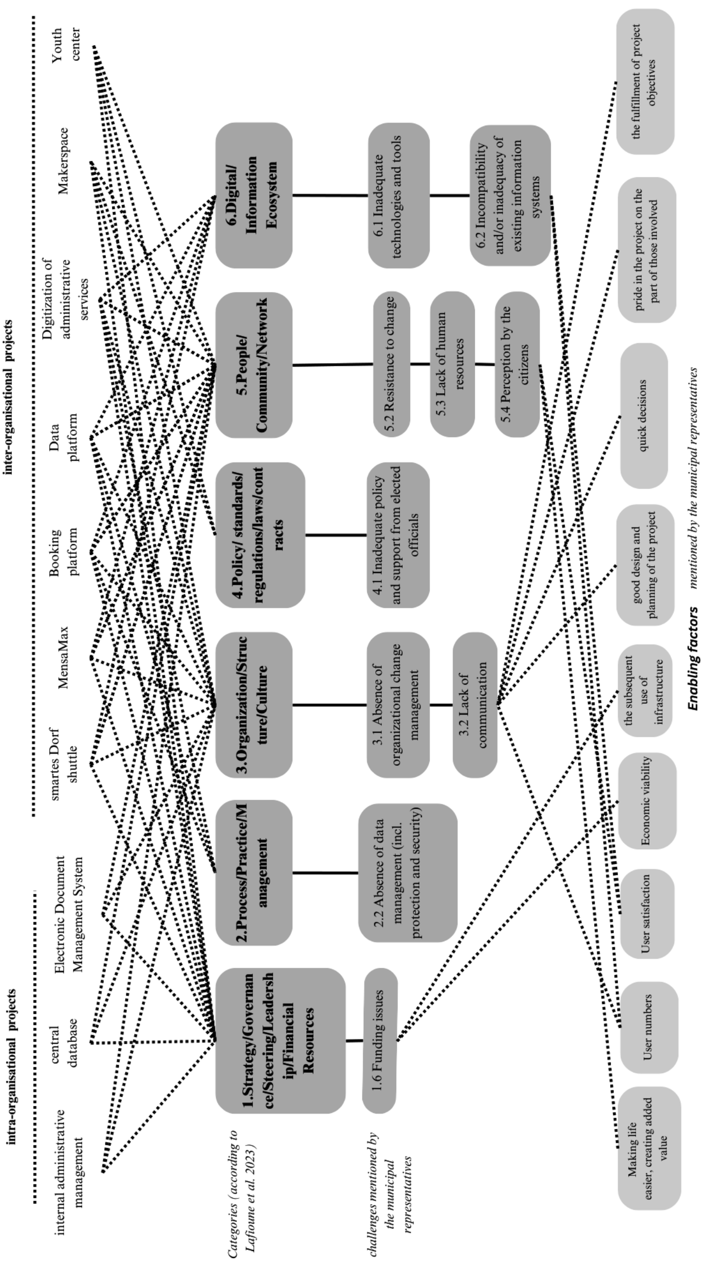

Figure 1: The relationships between digitalisation projects, challenges and enabling factors in Süderbrarup.

5. Discussion

The participants provided many responses in the subcategories of financial issues, the lack of communication and resistance to change. The Süderbrarup municipality should therefore pay particular attention to these challenges when implementing digitalisation projects. However, these aspects relate more to structural and inter-organisational problems that can only be solved to a limited extent at the project level. The challenges highlighted by the representatives of the municipality should thus be discussed not in isolation among project members but from a systemic perspective involving various interest groups. Linking the challenges with the enabling factors represents a first step in exploring how challenges can be met in the course of implementing digitalisation projects. However, these factors remain very general in their wording, as seen in the example 'making life easier, creating added value'; it remains unclear what 'added value' means for each individual. Moreover, some of the enabling factors mentioned contradict one another; for example, the participants mentioned 'quick decisions' as an enabling factor while also claiming that projects should be well-planned and involve a diverse user group at an early stage. In this regard, it remains questionable whether the involvement of citizens in the design and decision-making processes of digitalisation projects can lead to quick decisions. Thus, the proposed presentation not only provides a systematic overview of DT measures but also reveals the contradictions in the statements made by the municipal representatives.

The example of Süderbrarup demonstrates the applicability of Lafioune et al.'s (2023) framework for urban-rural regions. The proposed categories and subcategories were suitable for systematising the challenges described by the participants for each digitalisation project. The framework thus allows for the systematic analysis, framing and comparison of the challenges posed by DT in both urban and rural municipalities. For instance, differences can be found when comparing this research with the results from a survey of Canadian communities (cf. Lafioune et al., 2023). Respondents in the urban Canadian context referred to 2.2 *Absence of data management*, 2.3 *Misalignment of current practices and management* and 6.2 *Incompatibility and/or inadequacy of existing information systems* as particularly major challenges in DT. Comparing the results of the challenges of DT in the urban and rural contexts can assist in the development of **smart region** strategies. The DT framework is also valuable because it incorporates elements for sustainable digitalisation under the second category, *Processes,* *practices and management*, such as data protection (Herlo et al., 2023; Sühlmann-Faul, 2024, p. 141). Furthermore, the perception of the citizenry is emphasised under the fifth category, *People, community and network*. From the perspective of municipal representatives, citizens should be involved in the design process of municipal digitalisation projects from the beginning in order to promote the acceptance and long-term use of these applications (Hamm et al., 2023). **Smartness** is therefore understood not only in terms of what is technically possible but also what is useful for citizens in the long term. The applied framework has so far only enabled a systematic presentation of the challenges faced in the course of the DT of municipalities. This article proposes an extension by adding enabling factors for digitalisation projects to the framework.

However, the identified enabling factors remain very general and, to a certain extent, even contradictory. Among other issues, it remains unclear how the economic viability of projects can be measured in concrete terms. It would therefore be useful for future researchers to add additional items to the list of enabling factors. Furthermore, connections between projects and enabling factors could be investigated to map the relationships between digitalisation projects, challenges and enabling factors more clearly. This means that each person responsible for a respective project can develop specific strategies to overcome challenges such as the *incompatibility of existing information systems* (6.2). However, some of the challenges mentioned cannot be solved at the project implementation level. Factors such as a lack of sustainable financial and personnel resources must be addressed at the organisational level. In addition, further case study investigations in the context of **smart regions** with both municipal representatives and citizens are necessary in order to increase the generalisability.

These results can be used as a first step towards developing a holistic model that allows for the understanding of the necessary prerequisites for the successful implementation of digitalisation projects at the regional level. Other municipalities can thus acquire a concrete idea of which projects are being implemented in other areas on the way to becoming a **smart region** and reflect on the applicability and usefulness for their own case. The findings also provide insight into the challenges of the still-young research field of *smart regions*, offering a basis for future researchers to compare the challenges and enabling factors in the course of the DT in different regions.

6. Conclusions

By applying the framework developed by Lafioune et al. (2023), the prevailing challenges that hinder the DT of municipalities in urban-rural areas could be identified. The example of the Süderbrarup municipality shows that a lack of communication, especially with citizens, is one of the major hurdles in the implementation of digitalisation projects. In order better bring about the enabling factors mentioned, such as higher usage numbers, user satisfaction and subsequent use of (digital) infrastructure like the *smartes DorfSHUTTLE* app, the people who are affected must be involved in the planning of the projects at an early stage. The acceptance hurdles mentioned by the workshop participants – problems at the organisational level and social change processes as well as user problems in dealing with software products – clearly point to a necessity that coincides with theories concerning the social construction of technology: users need to be involved in the design process from the outset. For users of internal software, this will address the enabling factors ('pride in being part of the change' and 'having been involved in a successful project'), and for end users outside of the organisational levels, a product will be created that reflects the needs and requirements of future users. Acceptance can also be increased by increasing clarity and transparency concerning, for example, how software handles user data. All in all, only when the problems described are understood as social problems can socio-technical support help to overcome them.

Acknowledgements

This work was partly funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII131 ("Deutsches Internet-Institut").

References

Amin, A., & Thrift, N. (2002). *Cities: Reimagining the Urban*. Wiley.

Amt Süderbrarup (2021). *SMART CITY STRATEGIE*. https://smartcityamtsuederbrarup.de/wp-content/uploads/2021/05/Finale_Strategie.pdf

Andejany, M., Malik, A., Ahmad, W., Alhabri, A. M., & Umar, S. (2023). Transformation of Urban Cities to Sustainable Smart Cities – Challenges and Opportunities faced by Saudi Arabia. *Journal of Theoretical and Applied Information Technology*. Vol.(No.), Seiten. evtl. doi fehlt

Ben Letaifa, S. (2015). How to strategize smart cities: Revealing the SMART model. *Journal of Business Research*, *68*(7), 1414–1419. https://doi.org/10.1016/j.jbusres.2015.01.024

Bijker, W. E. (1997). Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change. *Science, Technology, & Human Values*, *22*, 127–130. http://www.jstor.org/stable/689970

Bradley, S., & Mahmoud, I. H. (2024). Strategies for Co-Creation and Co-Governance in Urban Contexts: Building Trust in Local Communities with Limited Social Structures. *Urban Science*, *8*(1), 9. https://doi.org/10.3390/urbansci8010009

Brenner, N., & Schmid, C. (2015). Towards a new epistemology of the urban? *City*, *19*(2–3), 151–182. https://doi.org/10.1080/13604813.2015.1014712

Davis, F. D. (1985). *A Technology Acceptance Model for Empirically Testing New End-user Information Systems: Theory and Results* [Massachusetts Institute of Technology, Sloan School of Management]. https://dspace.mit.edu/handle/1721.1/15192

De Guimarães, J. C. F., Severo, E. A., Felix Júnior, L. A., Da Costa, W. P. L. B., & Salmoria, F. T. (2020). Governance and quality of life in smart cities: Towards sustainable development goals. *Journal of Cleaner Production*, *253*, 119926. https://doi.org/10.1016/j.jclepro.2019.119926

Delgado-Viñas, C., & Gómez-Moreno, M.-L. (2022). The Interaction between Urban and Rural Areas: An Updated Paradigmatic, Methodological and Bibliographic Review. *Land*, *11*(8), 1298. https://doi.org/10.3390/land11081298

Fila, D., Fünfgeld, H., & Dahlmann, H. (2023). Climate change adaptation with limited resources: adaptive capacity and action in small- and medium-sized municipalities. *Environment, Development and Sustainability*, *26*(3), 5607–5627. https://doi.org/10.1007/s10668-023-02999-3

Geisberger, M. (2023). Smarte Regionen – Ländlicher Raum als Chancenraum? In D. Ahrens (Ed.), *Smart Region: Angewandte digitale Lösungen für den ländlichen Raum* (pp. 25–41). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38236-0_3

Gudmundsdottir, S., Sigurjonsson, T. O., Hlynsdottir, E. M., Fridriksdottir, S. D., & Ingibjargardottir, I. S. (2024). Sustainable Digital Change: The Case of a Municipality. *Sustainability*, *16*(3), 1319. https://doi.org/10.3390/su16031319

Hamm, A., Ullrich, A., Rehak, R., Becker, K., & Nölting, B. (2023). Municipalities' digitalization and sustainability transformations: A network approach for designing a transdisciplinary knowledge community. *The 11th International Conference on Communities and Technologies (C&T)*, 174–179. https://doi.org/10.1145/3593743.3593780

Herlo, B., Ullrich, A., & Vladova, G. (2023). Sustainable Digital Sovereignty - Interdependencies Between Sustainable Digitalization and Digital Sovereignty. *WEIZENBAUM SERIES (32) working paper \ June 2023*. https://doi.org/10.34669/WI.WS/32

Heuser, L., Illigmann, K., Kassyda, C., & Kliesch, M. (2020). *DIN SPEC 91387:2020-08, Kommunen und digitale Transformation_- Übersicht der Handlungsfelder*. Beuth Verlag GmbH. https://doi.org/10.31030/3181316

Jakob, M., & Krcmar, H. (2018). Which barriers hinder a successful digital transformation in small and medium-sized municipalities in a federal system? *Central and Eastern European EDem and EGov Days*, *331*, 141–150. https://doi.org/10.24989/ocg.v331.12

Jonathan, G. M. (2020). Digital Transformation in the Public Sector: Identifying Critical Success Factors. In M. Themistocleous & M. Papadaki (Eds.), *Information Systems,* (*381*), 223–235. Springer International Publishing. https://doi.org/10.1007/978-3-030-44322-1_17

Kogan, N., & Lee, K. J. (2014). Exploratory Research on the Success Factors and Challenges of Smart City Projects. *Asia Pacific Journal of Information Systems*, *24*(2), 141–189. https://doi.org/10.14329/apjis.2014.24.2.141

Krishnan, B., Arumugam, S., & Maddulety, K. (2020). Critical success factors for the digitalization of smart cities. *International Journal of Technology Management & Sustainable Development*, *19*(1), 69–86. https://doi.org/10.1386/tmsd_00016_1

Lafioune, N., Poirier, E. A., & St-Jacques, M. (2023). Managing urban infrastructure assets in the digital era: challenges of municipal digital transformation. *Digital Transformation and Society*, Vol (No), pages. https://doi.org/10.1108/DTS-02-2023-0010

Markkula, M., & Kune, H. (2015). Making Smart Regions Smarter: Smart Specialization and the Role of Universities in Regional Innovation Ecosystems. *Technology Innovation Management Review*, *5*(10), 7–15. https://doi.org/10.22215/timreview/932

Masik, G., Sagan, I., & Scott, J. W. (2021). Smart City strategies and new urban development policies in the Polish context. *Cities*, *108*, 102970. https://doi.org/10.1016/j.cities.2020.102970

Matern, A., Binder, J., & Noack, A. (2020). Smart regions: insights from hybridization and peripheralization research. *European Planning Studies*, *28*(10), 2060–2077. https://doi.org/10.1080/09654313.2019.1703910

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, *36*(4), 101385. https://doi.org/10.1016/j.giq.2019.06.002

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Ruscheinski, T. (2023). Herausforderungen ländlicher Räume – das Ziel gleichwertiger Lebensverhältnisse. In D. Ahrens (Ed.), *Smart Region: Angewandte digitale Lösungen für den ländlichen Raum* (pp. 9–23). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38236-0_2

Sahu, A. K., Padhy, R. K., & Dhir, A. (2020). Envisioning the Future of Behavioral Decision-Making: A Systematic Literature Review of Behavioral Reasoning Theory. *Australasian Marketing Journal*, *28*(4), 145–159. https://doi.org/10.1016/j.ausmj.2020.05.001

Shamsuzzoha, A., Nieminen, J., Piya, S., & Rutledge, K. (2021). Smart city for sustainable environment: A comparison of participatory strategies from Helsinki, Singapore and London. *Cities*, *114*, 103194. https://doi.org/10.1016/j.cities.2021.103194

Slavova, M., & Okwechime, E. (2016). African Smart Cities Strategies for Agenda 2063. *Africa Journal of Management*, *2*(2), 210–229. https://doi.org/10.1080/23322373.2016.1175266

Soja, E. W. (2000). *Postmetropolis: critical studies of cities and regions*. Blackwell Pub.

Statistisches Bundesamt (2023). *Bevölkerung nach Altersgruppen*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/bevoelkerung-altersgruppen-deutschland.html

Sühlmann-Faul, F. (2024). *Der goldene Käfig des Digitalkapitalismus: nichts kostet mehr als kostenlos*. oekom.

Van Der Hoogen, A., Fashoro, I., Calitz, A. P., & Luke, L. (2024). A Digital Transformation Framework for Smart Municipalities. *Sustainability*, *16*(3), 1320. https://doi.org/10.3390/su16031320

Will, O. C. (2021). Smart Regions: Elemente einer Transferagenda. In A. Mertens, K.-M. Ahrend, A. Kopsch, & W. Stork (Eds.), *Smart Region* (pp. 457–474). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29726-8_25

Zundel, P. (1999). Inter- und intra-organisationale Netzwerke. In P. Zundel, *Management von Produktions-Netzwerken* (pp. 19–74). Deutscher Universitätsverlag. https://doi.org/10.1007/978-3-663-08712-0_2

Digital Technologies for Sustainability

Climate change and AI: A research agenda for sustainable intelligence

Ina K. Schieferdecker

Abstract: Artificial intelligence (AI) represents one of the most powerful technologies currently available to humanity, offering immense potential for transformative change. It has the potential to assist us in addressing our most pressing climate change concerns. It can also enhance our ability to promote and lead the transition to sustainability. However, it is also important to acknowledge that AI is a technology that requires significant resources. It is estimated that by 2025–2026, the daily training costs for the large AI models will exceed the global computing capacity. This chapter presents the latest developments in AI for sustainability transformation and sustainable AI, with a particular focus on reducing resource consumption. It introduces the concept of sustainable intelligence and discusses a research agenda featuring a multidisciplinary, transformative approach to the design and development of AI technologies and AI-based systems, with the objective of ensuring their development in an environmentally friendly and socially responsible manner, and with the aim of investigating their impact on the natural, technical and societal environment.

Keywords: digital transformation, information and communication technologies, artificial intelligence, sustainable AI, AI for sustainability, sustainable intelligence

1. Technologies and societal progress – demystifying technical progress

In the context of climate change and the potential of artificial intelligence (AI) to develop solutions to mitigate its effects, this section begins with a discussion of key concepts of technology, technological progress, and technical infrastructures in societies. Building on this, Section 2 discusses information and communication technologies and the dual transition for digital transformation and sustainable transformation in line with the United Nations Sustainable Development Goals. Finally, Section 3 focuses on AI as an emerging powerful technology and its relationship to the transformation towards sustainability. Based on this discussion and a review of related research, the concept of sustainable intelligence and a research agenda for sustainable intelligence are presented and discussed.

This section introduces key concepts of engineering in the construction of the technical infrastructures on which modern societies depend, such as energy supply, transport and mobility services, or information distribution. A basic understanding of engineering is important to place the social science debate on contextualised dichotomies in digital transformation; on the relationship between society, resources, and pollution; or on power relations into a broader, multidisciplinary perspective that includes technical sciences and engineering.

In his book on the 'technological society', Ellul (2021, p. XXV) states:

> The term technique, as I use it, does not mean machines, technology, or this or that procedure for attaining an end. In our technological society, technique is the totality of methods rationally arrived at and having absolute efficiency (for a given stage of development) in every field of human activity. Its characteristics are new; the technique of the present has no common measure with that of the past.

And Ellul (2021, p. 427):

> The human race is confusedly beginning to understand at last that it is living in a new and unfamiliar universe. The new order was meant to be a buffer between man and nature. Unfortunately, it has evolved autonomously in such a way that man has lost all contact with his natural framework … Enclosed within his artificial creation, man finds that there is no 'exit' … The new milieu has its own specific laws which are not the laws of organic or inorganic matter. Man is still ignorant of these laws.

Sixty years later, we are still struggling to realise that in the Anthropocene the human-centred view must be replaced by or, at least, combined with a planet-centred view, where humans are just one of the species on this planet. If the well-being of the majority of humanity is to prevail, then the well-being of the whole planet must be the top priority towards which all available resources and efforts must be directed.

Another obstacle is that this artificial technological universe, as the source of climate change, is also the foundation for the solution to the transformation towards sustainability and climate resilience (see section 2). However, the initial theory of technocracy has proved inadequate in addressing the challenges that emerged in the context of political and industrial crises. Technocracy gave rise to expectations that were not fully aligned with the capabilities of the technocratic approach, which was also perceived as a means of declining ideologies and even politics in a knowledgeable society. Boorstin (1978, frontmatter) posits that the converging powers of technology will ultimately prevail, overcoming the barriers posed by '[i]deology, tribalism, nationalism, the crusading spirit in religion, bigotry, censorship, racism, persecution, immigration and emigration restriction, tariffs, and chauvinism'.

In a subsequent statement, Gunnell (1982, p. 9) asserts that:

> The precise nature of the impact of technology on politics is sometimes ambiguous, but it seems to involve three distinct – though not mutually exclusive – theories, dimensions, or levels of analysis:
> 1. In circumstances in which political decisions necessarily involve speciali[s]ed knowledge and the exercise of technical skills, political power tends to gravitate toward technological elites.
> 2. Technology has become autonomous, hence politics has become a function of systemic structural determinants over which it has little or no control.
> 3. Technology (and science) constitute a new legitimating ideology that subtly masks certain forms of social domination.
> In each level of analysis, there is a concern about the depreciation of the political realm, the subversion of traditional bases of authority, and the ascendancy of instrumental over political rationality.

Gunnell's evaluation provides a more accurate depiction of the interrelationship between technique (and technologies) and politics (and society). In the context of the technocratic paradigm, however, the contributions of technologies and engineers are frequently excluded from social and political discourse on the grounds that they would be unable to offer meaningful solutions to the challenges societies face. This attitude is particularly noteworthy in light of the fact that, as a consequence of technological advancement, global well-being, satisfaction, and happiness are on a steady upward trajectory (Azuh et al., 2020; Hausken & Moxnes, 2019; Kowal & Paliwoda-Pekosz, 2017).

Myths about technical progress include:

- We would have been overrun by technology.
- Technology would be harmful.
- Technology would be accountable.

Notably, technology is a science and the study of engineering. It is concerned with the ways in which (raw) materials and components can be transformed into technical systems, including digital products, services, and information/data. These transformation processes use scientific knowledge and tools. The results of engineering are technical systems that are deployed, operated, maintained, and used, and the sum of these technical systems constitutes the technical infrastructure of a society. Its modernity and availability are directly related to a society's ability to be innovative and resilient: 'in the longer view a secure society involves innovation in strong infrastructure and social systems' (Allenby & Fink, 2005, p. 1034).

Moreover, because technologies, technical systems, and infrastructures have no agency, they cannot be held accountable, but stakeholders such as clients, contractors, developers, operators, consumers, regulators, researchers, entrepreneurs, or firms can. Technologies, actors, and infrastructures (technical, institutional, and operational) constitute technological systems:

> A technological system may be defined as a network of agents interacting in a specific economic/industrial area under a particular institutional infrastructure or set of infrastructures and involved in the generation, diffusion, and utili[s]ation of technology. Technological systems are defined in terms of knowledge/competence flows rather than flows of ordinary goods and services. They consist of dynamic knowledge and competence networks. In the presence of an entrepreneur and sufficient critical mass, such networks can be transformed into development blocks, i.e. synergistic clusters of firms and technologies within an industry or a group of industries. (Carlsson & Stankiewicz, 1991, p. 111)

Such technological systems are sources of innovation (Blatter 2004). They can be regional, such as the microelectronics core in Saxony, or cross-border, such as Silicon Valley and the Baja California region between the United States and Mexico, or the Øresund region between Denmark and Sweden.

In addition, technologies, technical systems, infrastructures, technological systems, and innovations are human-made. The actors involved in their development decide on the objectives, the values pursued, and the final impact on society. The closer developers get to the final artefacts, the smaller the set of choices available. Along the development processes – whether it is a concrete product, a technical infrastructure, or an innovation region – more and more values are hard-coded. Therefore, it is essential to address value sets from the outset of analysis and design. The speed and scale of today's information and communication technologies (ICTs) used in digital transformation make it particularly important to ensure respect for democratic values. Important fields include privacy, data security, and decision-making powers (Subirats, 2002), sustainability goals like the United Nations' Sustainability Development Goals (UN SDGs; see Pedersen, 2018), and business values that include corporate social responsibility (CSR; Ali et al., 2017) and environmental, social, and governance performance (ESG; Huang, 2021).

Finally, Esmark (2020, p.79) 'clears up the considerable confusion surrounding the relationship between technocracy, bureaucracy and democracy, which provides the foundation for the empirical analysis of the anti-bureaucratic and pro-democratic nature of contemporary technocracy', leading me to conclude that there is no dichotomy between technology and sustainability and no dichotomy between technology and prosperity, but that there may be an imbalance between the technical, environmental, and political environment, as will be discussed in the following sections.

Before discussing this further, the paper presents a brief overview of the components of the ICTs that provide the fundamental capabilities for digital transformation (Figure 1). Hardware and software are used to build the devices, which are interconnected by networks for telecommunications and mobile communications. The devices essentially store, compute, interpret, and present data that can be stored in databases. (Cyber) security is important to ensure the integrity of systems and services, to protect sensitive data, to prevent cybercrime, to protect privacy, and to build and maintain trust in digital solutions. More and more IT services are being virtualised and delivered via cloud platforms as infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS), and data as a service (DaaS) – in short, everything as a service (XaaS). Emerging technologies for ICT include:

- Internet of Things (IoT) for connecting the physical and cyber worlds, where IoT refers to a network of interconnected physical devices that can collect, exchange, and act on data over the Internet

- Cloud computing for the delivery of computing services such as storage, processing, and software over the Internet, allowing users to access and use these resources on demand from anywhere
- Edge computing focuses on processing data closer to where it is generated or needed, reducing latency and bandwidth consumption by processing data locally on devices or close to the data source
- Distributed ledger technologies to improve security, immutability, and/or transparency through decentralised systems where multiple participants maintain and validate a synchronised, immutable record of transactions across a network
- Computer vision to analyse, understand, and interpret visual information from digital images, video, or 3D models to make decisions or to perform specific tasks to interpret and make decisions based on visual data from the world
- Virtual reality (VR) immerses the user in a fully artificial digital environment, and augmented reality (AR) overlays digital information onto the real world to enhance the user's perception of their surroundings
- AI essentially simulates human intelligence processes by computer systems, enabling them to perform tasks such as learning, reasoning, problem-solving, perception, and language understanding
- Next generation software engineering improves the efficiency, scalability, and quality of software development and maintenance processes by applying new methods such as AI, DevOps, or microservices

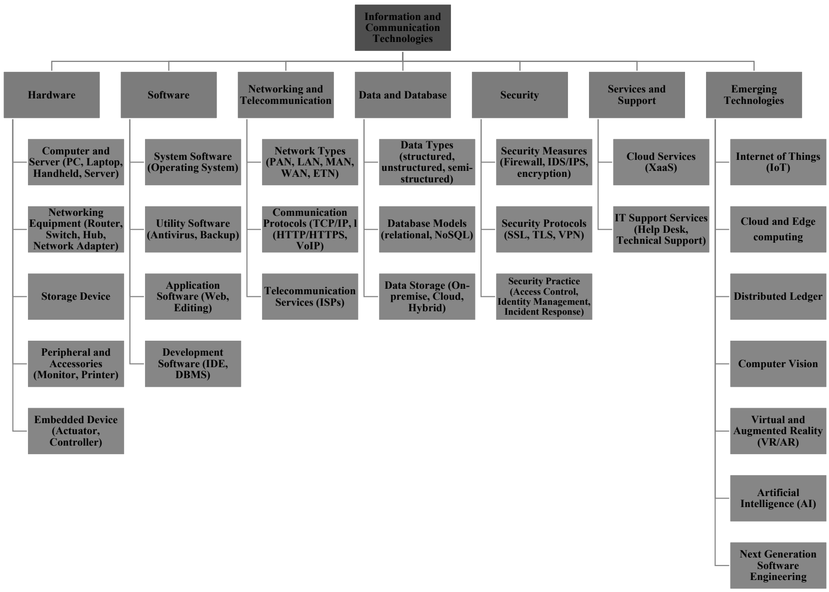

Figure 1. Overview of information and communication technologies (ICTs)

ICTs provide the basis for digital socio-technical systems, e.g. in Industry 4.0 (Aceto et al., 2019). The term socio-technical system was coined in the 1960s and refers to the

> joint optimization of the social and technical systems. The technical and social systems are independent of each other in the sense that the former follows the law of natural science while the latter follow the laws of the human sciences and is a purposeful system. Yet they are correlative in that one requires the other for the transformation of an input to an output, which comprise the functional task of a work system. Their relationship represents a coupling of dissimilars which can only be jointly optimized. (Trist, 1981, p. 24)

More recent publications define digital transformation (DT) as 'the process of organizational or societal changes driven by innovations and developments of ICT. DT includes the ability to adopt technologies rapidly and affects social as well as technical elements of business models, processes, products and the organizational structure' (Bockshecker et al., 2018, p. 9).

ICT has led to the development and evolution of the Internet 2 (Leiner et al., 2009; Ryan, 2010), which began as an Internet of Information for the exchange of data and documents through hypertexts. With the commercialisation of the Internet after the 1990s, it has evolved into the Internet of Services, where different services are seamlessly connected and delivered. The Internet of Services focuses on the provision of a wide range of services offered, such as infrastructure, platform, and business process services, enabling greater accessibility, efficiency, and automation. It aims to create a service-oriented ecosystem where services can be dynamically discovered, provisioned, managed and consumed. Since the 2000s, the Internet of Things has also added connections to the physical world through smart sensors and actuators. Currently, the Internet is evolving into the Internet of Collaboration by providing unified platforms, enhanced connectivity and real-time collaboration for instant interaction and synchronised workflows. It enables remote working, education, and participation at a new scale, as well as inclusive participation.

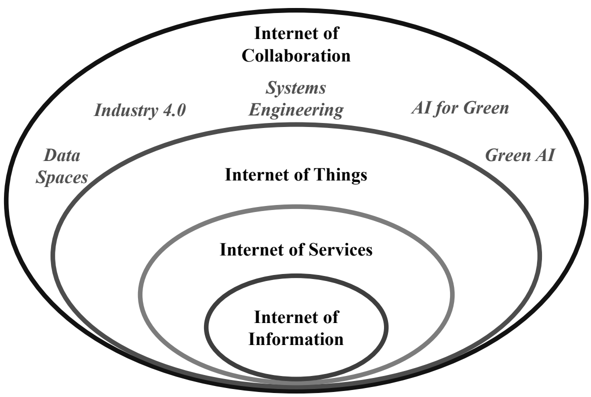

Figure 2. The evolution of the Internet

The Internet has become a technique that has transformed the way individuals, businesses, and societies interact. It has become a central part of the technical infrastructures of societies to the extent that the ability to use the Internet as a means of consuming and providing information is being discussed as a fundamental human right – for the freedom of expression through the free access to the Internet (Reglitz, 2020). Its emergence also led to the term Digital Age (Messner et al., 2019), originally called the

Information Age (Toffler, 1982). While the Internet fundamentally changed the exchange of information and services in societies, AI will fundamentally change their production and consumption, including their exchange (see also section 3).

2. Sustainability and digitalisation

Before discussing the implications of AI further, this section briefly reviews aspects of sustainability and digitalisation in general: At the UN Conference on Sustainable Development in Rio de Janeiro in 2012, the "The Future We Want" document on sustainable development and a green economy was adopted. This document set the stage for the development of the UN Sustainable Development Goals (UN SDGs). The UN SDGs (Saxena et al., 2021) are a globally adopted canon of values to combat climate change with balanced means for environmental, social, and economic progress. They consist of 17 goals and 169 targets to advance societies economically and politically, reducing poverty and increasing prosperity in a world with mitigated climate change. The SDGs address environmental, social, and economic sustainability for a coordinated approach to climate change (see Figure 3).

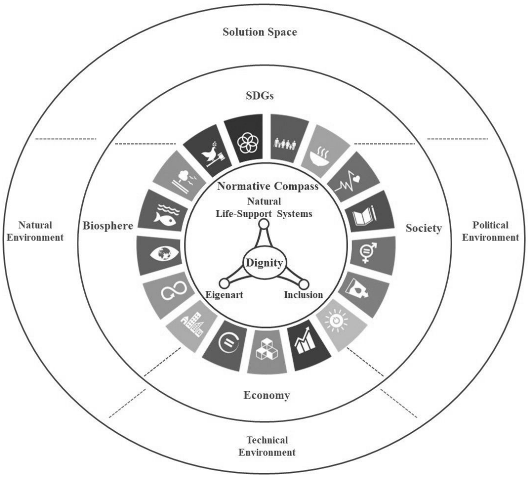

Figure 3. Solution space, SDGs and canon of values for sustainability

The transformation towards sustainability in the Digital Age is based on a normative compass. Its starting point is human dignity, i.e. its inviolability, respect, and protection. In addition, the normative compass has three dimensions (Messner et al., 2019, p. 3; for further reading, see Chowdhary, 2020; Enholm et al., 2022; Zhang & Lu, 2021):

- 'Sustaining the natural life-support systems': Comply with planetary guardrails and avoid or solve local environmental problems.
- 'Inclusion': Ensure universal minimum standards for substantive, political and economic inclusion.
- 'Eigenart': Recognise the value of diversity as a resource for successful transformation and as a condition for well-being and quality of life.

As an aside, it is important to note that researchers question (Stober, 2019; Tarcan et al., 2022) whether the normative compass for the transformation towards sustainability should remain human-centred, as it is today, or shift to a more-than-human-centred or nature-centred leadership, which has been debated often and remains an ongoing discussion. However, this topic does not directly contribute to the essence of this chapter and, therefore, will not be discussed further here.

The solution space for sustainability transformation in the Digital Age consists of solutions that target the natural, technical, or political environment. All three environments are interconnected, and a solution in one environment rarely works without taking into account the others, also referred to as dimensions. Most solutions for the transformation towards sustainability will have to be socio-technical-ecological solutions, going far beyond the conventional concept of purely socio-technical systems (Ahlborg et al., 2019; Smith & Stirling, 2010).

Due to the complexity of the SDGs, this chapter focuses on the environmental aspects of sustainability in view of digitalisation in general; later sections centre on AI. Because ICTs and digitalisation, like the SDGs, are very complex (see section 1), the chapter below focuses on the specific implications of AI for environmental sustainability in view of the natural, technical, and political environment.

There are a number of surveys on environmental sustainability and digitalisation:

- Chen et al. (2020) investigate whether digital applications in manufacturing positively or negatively impact the environment by looking at improved resource efficiency in product design, production, transport, and customer service. They also examine the environmental impacts of resource and energy use, including waste and emissions from the manufacture, use, and disposal of digital systems. The authors propose a lifecycle perspective on the environmental impacts of a product and the technology lifecycle. Such a lifecycle perspective is supported by the ecological footprint, which should also include the footprint of the digital services used.

- Broccardo et al. (2023, p.15) review how companies have built sustainable business models through the use of digitalisation. They identify the benefits of digitalisation in helping to 'i) share resources, ii) improve relationships and collaboration, iii) reuse, and iv) recycle. All these actions have a positive impact on the cost structure, efficiency, and the creation of new revenue flows'.
- Lange et al. (2020, frontmatter) investigate the relationship between digitalisation, ICT usage, and energy consumption. They conclude that 'ICT decreases energy demand via energy efficiency and sectoral change. ICT increases energy demand via a growing ICT sector, rebounds and economic growth'. Since digitalisation cannot decouple economic growth from energy consumption, it finally leads to a rise in energy consumption.
- Guandalini (2022, p. 466) analyses the relationship between digital transformation and sustainability improvements, calling for 'the development of a new stream of literature' dubbed 'digital sustainability'. The author claims that the 'identified research gaps are expected to foster future investigations with more focused outputs from management scholars to the practical community' and highlights selected research questions.

All this suggests that if ICT-based solutions for sustainability are resource-efficient and socially acceptable, they can make a real contribution to tackling climate change and improving climate resilience (Santarius et al., 2023; Rome, 2019). This view is also shared by the European Green Deal's policy objective of a digital and environmental 'twin transition', in which the two dynamics of digital and sustainable transformation reinforce each other (Salvi et al., 2022). For further reading on the digital revolution for sustainable development, see Fouquet and Hippe (2022), Mäkitie et al. (2023), and Sachs et al. (2019).

3. The impact of AI on environmental sustainability

Before discussing the implications of AI on the natural environment, this section provides a brief overview of AI as a key technology (see Figure 4) applied in almost every application domain of digitalisation. While there is no widely accepted definition of AI (Monett & Lewis, 2018), according to (Union, 2024, Article 3), an

> 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Thus, AI systems are generally defined as cyber-physical systems that exhibit adaptability, however defined. Cyber-physical systems are

> [o]pen, networked systems ... that use sensors to capture data about what is going on in the physical world, interpret these data and make them available to network-based services, whilst also using actuators to directly affect processes in the physical world and control the behaviour of devices, objects and services. (Geisberger & Broy, 2015, p. 13)

Wang (2019, p. 1) defines (artificial) intelligence as capabilities for 'adaptation with insufficient knowledge and resources'. These capabilities include knowledge acquisition, language processing, search and pattern recognition, reasoning and learning, decision-making, and problem-solving (see Figure 4). On several benchmarks, AI-based systems have surpassed some human capabilities, such as image classification, visual reasoning, and language understanding, but not yet the capabilities needed for more complex problem-solving in mathematics, visual reasoning, or planning, to name a few (Perrault & Clark, 2024).

Nevertheless, AI capabilities are used in a wide variety of applications, including expert systems, natural language processing, computer vision, multi-agent systems, autonomous systems, and robotics and motion control. A wide variety of methods can be used to build AI-based systems. These can be divided into symbolic, statistical, and sub-symbolic methods, each of which has a number of subcategories (e.g. Chowdhary, 2020; Enholm et al., 2022; Zhang & Lu, 2021).

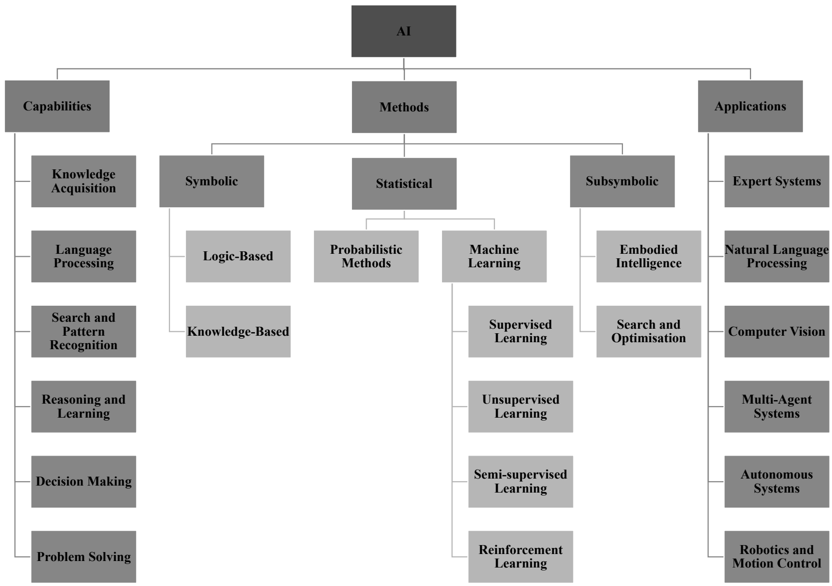

Figure 4. Overview of artificial intelligence

Digital systems (Belkhir & Elmeligi, 2018) and AI-based systems (Luccioni et al., 2024) have several orders of magnitude of impact on the environment and society (Bohnsack et al., 2022):

- 1st order: The use of resources and energy in the development, production, construction, operation and maintenance of AI-based technical infrastructures (Luccioni et al., 2024).
- 2nd order: Consumption effects through differently organised AI-based processes and through new processes such as in industry, administration, or smart cities (Bibri et al., 2024).
- 3rd order: Consumption and lifestyle effects through AI-based digital systems like online marketplaces (Bai, 2022).

The first, second, and third-order impacts of AI add up to the total impact, which can be positive or negative in terms of CO_2 emissions or resource use, impacting environmental sustainability. Currently, the first-order impacts are already very large in terms of energy consumption and carbon release, for example, for AI training and the use of AI systems such as ChatGPT or Llama (Luccioni et al., 2024):

> For instance, Meta's Llama2 70B model released approximately 291.2 tonnes of carbon, which is nearly 291 times more than the emissions released by one traveler on a round-trip flight from New York to San Francisco, and roughly 16 times the amount of annual carbon emitted by an average American in one year. (Perrault & Clark, 2024, p. 154)

The calculation of second and third-order effects is challenging due to the rapid pace of change in AI. No concrete figures are available. Nevertheless, research (Vinuesa et al., 2020) suggests that sustainable AI-based systems will make a significant contribution to tackling climate change. Some evidence indicates that AI can enable 134 targets across all SDGs, i.e. it can have a positive impact on 79% of the SDG targets, while it can inhibit 59 targets, i.e. it can have a negative impact on 34%. Further, the positive impact can only be achieved if there is 'regulatory insight and oversight for AI-based technologies to enable sustainable development' (Vinuesa et al., 2020, p. 1).

The potential second and third-order effects of AI have been investigated in numerous comprehensive and supplementary studies:

- Nishant et al. (2020, p. 1) argue that 'AI can support the derivation of culturally appropriate organizational processes and individual practices to reduce the natural resource and energy intensity of human activities. The true value of AI will not be in how it enables society to reduce its energy, water, and land use intensities, but rather, at a higher level, how it facilitates and fosters environmental governance'.
- Uriarte-Gallastegi et al. (2023, p. 662) demonstrate that 'Artificial Intelligence can significantly impact resource efficiency and provide a competitive edge to organizations, primarily by reducing energy and material consumption'.
- Regarding the potential impact of AI on the SDGs, Gupta et al. (2021, p. 2) state that 'the Environment category entails the highest potential with 93% of the targets being positively affected' and that 'when taking into account the type of evidence indicating the connection with AI, we observed that the positive effects on the Environment … were quite robust'.
- Bibri et al. (2024) present the concept of smarter eco-cities as the convergence of AI and the Internet of Things (AIoT), which has significant potential to address complex environmental challenges by improving the performance and efficiency of smart cities.

Despite the existence of a multitude of research agendas pertaining to AI as a general-purpose technology, there is a notable absence of an integrated research agenda that encompasses both sustainable AI and AI for sustainability. To date, publications have focused on either sustainable AI or AI for sustainability. However, this separation of concerns carries the risk of overlooking important aspects, framework conditions, or side effects of AI. To illustrate, Vinuesa et al. (2020), Nishant et al. (2020), and Dwivedi et al. (2021) provide comprehensive accounts of AI for sustainability, but they do not address the specific issue of sustainable AI. More specifically, they fail to address the necessity of making AI more resource-efficient, including the issue of reducing its energy and data consumption. Another example is Mumtaz et al. (2022), who address potential unintended consequences, threats, and hazards of AI, including, for example, the performance threat. However, the article fails to consider the carbon footprint of AI. This omission is somewhat surprising given that Hilty and Aebischer (2015) had already identified the necessity for a new research field in the field of ICT for sustainability, highlighting the importance of reducing ICT-induced energy and material flows. However, the article by Hilty and Aebischer (2015) was somewhat lacking in its consideration of the societal need for trustworthy AI, although it also discussed ethical aspects of ICT. Verdecchia et al. (2021) present a technological landscape for the development of energy-efficient digital infrastructure. The paper's focus on the technical environment for sustainable transformation precludes any consideration of the natural and societal environments. While the authors identify the need for paradigm shifts and environmental and social solutions within the technical context, these are viewed exclusively from a technical perspective.

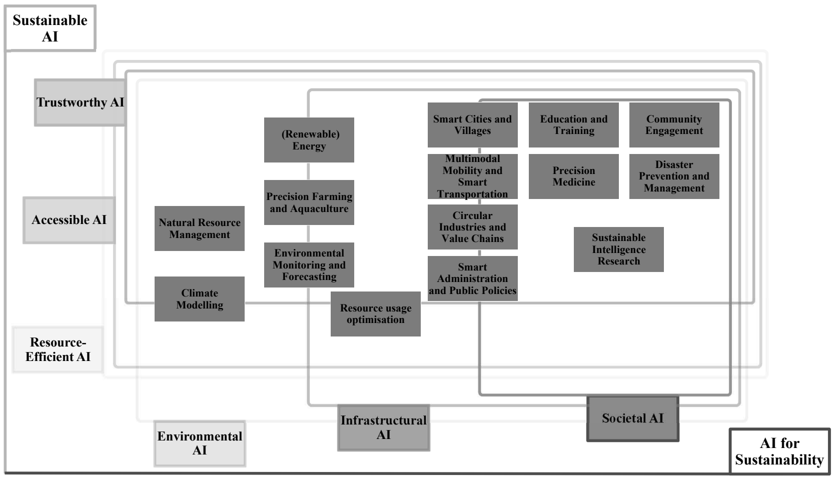

Figure 5. Research agenda for sustainable intelligence

It is time to coin a new concept: **sustainable intelligence** refers to the comprehensive approach of integrating sustainability principles into the development, deployment, and application of AI technologies. This concept has two primary dimensions:

1. Sustainable AI will address the sustainable development of AI by considering the environmental, technical, and societal impacts of AI in its design and implementation.

- Sustainable AI minimises carbon footprints, energy consumption, and resource use by optimising AI methods for data and energy efficiency, adopting the reduce, reuse, repair, and recycle principles for AI-based solutions, and using renewable energy to power them.
- Sustainable AI is to be made accessible through appropriate AI infrastructures, tools, and licences, including open access to AI and open-source AI methods, algorithms, models, and training and validation data for the public.
- Sustainable AI needs to be made trustworthy by making AI-based systems reliable, robust, and performant; more transparent through explanation, documentation, and disclosure; fairer through inclusivity and bias mitigation; accountable through explicit responsibilities and recourse mechanisms; secure by security design and secure operation

and through data, privacy, and security protection; and ethically aligned with the normative compass of society through value alignment.

By addressing these issues, sustainable AI can build and maintain trust among users and stakeholders, fostering greater acceptance and the responsible adoption of AI.

2. AI for sustainability to accelerate sustainability in society through AI-based solutions for a society's natural, infrastructural, and political environment.

- AI-based solutions for the natural environment involve the application of artificial intelligence technologies to monitor, manage, and protect natural ecosystems through remote sensing and (extra-)terrestrial and satellite imagery; image, sound, and pattern recognition; monitoring, modelling, simulation, and forecasting; and optimisation of resource use.
- AI-based solutions for the technical environment include the intelligent digitalisation of society's infrastructure, making it more robust, reliable, resilient, and secure through AI-based methods. These solutions also encompass the development and operation of AI infrastructures for industry, government, and the public, including data, computing, and storage centres.
- AI-based solutions at the societal level include enhancing public welfare and improving the quality of life in communities, using machine learning, data analytics, natural language processing, and other AI methods to address issues related to education, healthcare, public safety, governance, and more.

By applying AI technologies in these areas, we can develop innovative and effective solutions to further develop our democratic societies while protecting nature and biodiversity for future generations by reducing the impact of human activities on nature. Sustainable intelligence aims to create a symbiotic relationship between AI and sustainability, ensuring that technological advances contribute positively to the environmental, social, and economic dimensions of sustainability, while making AI development processes themselves more sustainable.

The research agenda for sustainable intelligence is shown in Figure 5, which highlights the highest-priority applications of AI and positions them in relation to the three-by-three dimensions of sustainable intelligence. It also highlights the fact that any AI-based solution in the public interest should be resource-efficient and accessible to relevant stakeholders. The majority of AI-based solutions must also be trustworthy, especially those that are societal AI solutions, as they interact within society with individuals, stakeholders, and interest groups. For simplicity's sake, the various AI-based solutions will not be detailed here. Instead, reference is made to the research mentioned above for further reading. The added value of the Research Agenda for Sustainable Intelligence lies in its multidimensional, multidisciplinary, and transformative approach, which is important to avoid siloed approaches to AI.

4. Conclusion

AI has and will continue to surpass human capabilities, enables novel solutions, and can be a socio-technical game changer in the transition towards sustainability. As climate change is the most pressing grand challenge facing humanity, and as digital solutions, especially AI-based systems, are at the same time the fastest-growing large energy consumers worldwide, it is important to engage in multidisciplinary, transformative research into the wide range of concepts, approaches, and empirical evidence of sustainable AI and AI for sustainability.

The concepts of sustainable AI and AI for sustainability collectively constitute sustainable intelligence. This term reflects the necessity to develop AI in a manner that is both resource-efficient and accessible while also ensuring that it is trustworthy. Furthermore, it encompasses the objective of utilising AI to address environmental, infrastructural and societal sustainability challenges.

In order to facilitate the successful development of sustainable artificial intelligence, it is essential that focused research adopts a multidisciplinary and transformative approach, investigating the enablers, inhibitors, power structures, and socio-technical basis for successful change towards sustainability. It is imperative that research be conducted to develop a comprehensive and well-founded knowledge base regarding the most appropriate technologies, designs, processes, and empirical models for facilitating a sustainable transformation. Furthermore, research should be conducted to gain a deeper understanding of the societal dynamics involved, including an analysis of the potential barriers and conditions for success.

Acknowledgements

This article has been made possible by the work and discussions with my colleagues and friends at the Weizenbaum Institute for the Networked Society, the Weizenbaum Conference on Uncertain Journeys into Digital Futures, and the German Advisory Council on Global Change. Thanks to all of you.

Disclosure of Interests

Ina Schieferdecker was a co-founding director of the Weizenbaum Institute. She is now an independent researcher with no connection to the Weizenbaum Institute other than as an alumna.

References

Aceto, G., Persico, V., & Pescapé, A. (2019). A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges. *IEEE Communications Surveys & Tutorials*, *21*(4), 3467–3501.

Ahlborg, H., Ruiz-Mercado, I., Molander, S., & Masera, O. (2019). Bringing technology into social-ecological systems research – motivations for a socio-technical-ecological systems approach. *Sustainability*, *11*(7), 1-23.

Ali, W., Frynas, J. G., & Mahmood, Z. (2017). Determinants of corporate social responsibility (CSR) disclosure in developed and developing countries: A literature review. *Corporate Social Responsibility and Environmental Management*, *24*(4), 273–294.

Allenby, B., & Fink, J. (2005). Toward inherently secure and resilient societies. *Science*, *309*(5737), 1034–1036.

Azuh, D. E., Ejemeyovwi, J. O., Adiat, Q., & Ayanda, B. A. (2020). Innovation and human development perspectives in West Africa. *Sage Open*, *10*(4), 1-10.

Bai, L. (2022). Analysis of the change of artificial intelligence to online consumption patterns and consumption concepts. *Soft Computing*, *26*(16), 7559–7569.

Belkhir, L., & Elmeligi, A. (2018). Assessing ICT global emissions footprint: Trends to 2040 & recommendations. *Journal of Cleaner Production*, *177*, 448–463.

Bibri, S. E., Krogstie, J., Kaboli, A., & Alahi, A. (2024). Smarter eco-cities and their leading-edge artificial intelligence of things solutions for environmental sustainability: A comprehensive systematic review. *Environmental Science and Ecotechnology*, *19*, 1-31.

Blatter, J. (2004). From 'spaces of place' to 'spaces of flows'? Territorial and functional governance in cross-border regions in Europe and North America. *International Journal of Urban and Regional Research*, *28*(3), 530–548.

Bockshecker, A., Hackstein, S., & Baumö U. (2018). Systematization of the term digital transformation and its phenomena from a socio-technical perspective: A literature review. In P. M. Bednar, U. Frank, & K. Kautz (Eds.), *26th European Conference on Information Systems: Beyond Digitization – Facets of Socio-Technical Change, ECIS 2018, Portsmouth, UK, June 23–28, 2018*, 1-17.

Bohnsack, R., Bidmon, C. M., Pinkse, J., eds. (2022). *Sustainability in the Digital Age: Intended and unintended consequences of digital technologies for sustainable development*. Wiley.

Boorstin, D. J. (1978). *Republic of technology: Reflections on our future community*. Harper and Row Publishers, Inc.

Broccardo, L., Zicari, A., Jabeen, F., & Bhatti, Z. A. (2023). How digitalization supports a sustainable business model: A literature review. *Technological Forecasting and Social Change*, *187*, 1-16.

Carlsson, B., & Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of Evolutionary Economics*, *1*, 93–118.

Chen, X., Despeisse, M., & Johansson, B. (2020). Environmental sustainability of digitalization in manufacturing: A review. *Sustainability*, *12*(24), 1-31.

Chowdhary, K. (2020). *Fundamentals of artificial intelligence*. Springer.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., et al. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 1-47.

Ellul, J. (2021). *The technological society*. Vintage. First published 1964.

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). Artificial intelligence and business value: A literature review. *Information Systems Frontiers*, *24*(5), 1709–1734.

Esmark, A. (2020). *The technocratic regime: Technocracy, bureaucracy and democracy*. Bristol, UK: Bristol University Press.

European Union. (2024). The Artificial Intelligence Act – Regulation (EU) 2024/1689.

Fouquet, R., & Hippe, R. (2022). Twin transitions of decarbonisation and digitalisation: A historical perspective on energy and information in European economies. *Energy Research & Social Science*, *91*, 102736.

Geisberger, E., & Broy, M. (2015). *Living in a networked world: Integrated research agenda Cyber-Physical Systems (agendaCPS)*. Herbert Utz Verlag.

Guandalini, I. (2022). Sustainability through digital transformation: A systematic literature review for research guidance. *Journal of Business Research*, *148*, 456–471.

Gunnell, J. G. (1982). The technocratic image and the theory of technocracy. *Technology and Culture*, *23*(3), 392–416.

Gupta, S., Langhans, S. D., Domisch, S., Fuso Nerini, F., Felländer, A., Battaglini, M., Tegmark, M. & Vinuesa, R. (2021). Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level. *Transportation Engineering*, *4*, 1-7.

Hausken, K., & Moxnes, J. F. (2019). Innovation, development and national indices. *Social Indicators Research*, *141*, 1165–1188.

Hilty, L. M., & Aebischer, B. (2015). ICT for sustainability: An emerging research field. In L. M. Hilty and B. Aebischer (Eds.), *ICT Innovations for Sustainability* (pp. 3–36). Springer International Publishing.

Huang, D. Z. (2021). Environmental, social and governance (ESG) activity and firm performance: A review and consolidation. *Accounting & Finance*, *61*(1), 335–360.

Kowal, J., & Paliwoda-Pekosz, G. (2017). ICT for global competitiveness and economic growth in emerging economies: Economic, cultural, and social innovations for human capital in transition economies. *Information Systems Management*, *34*(4), 304–307.

Lange, S., Pohl, J., & Santarius, T. (2020). Digitalization and energy consumption. Does ICT reduce energy demand? *Ecological Economics*, *176*, 1-14.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., & Wolff, S. (2009). A brief history of the Internet. *ACM SIGCOMM Computer Communication Review*, *39*(5), 22–31.

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, & Transparency* (pp. 85–99).

Mäkitie, T., Hanson, J., Damman, S., & Wardeberg, M. (2023). Digital innovation's contribution to sustainability transitions. *Technology in Society*, *73*, 102255.

Messner, D., Schlacke, S., Fromhold-Eisebith, M., Grote, U., Matthies, E., Pittel, K., Schellnhuber, H. J., Schieferdecker, I., Schneidewind, U., Augenstein, K., et al. (2019). Towards our common digital future. WBGU flagship report.

Monett, D., & Lewis, C. W. (2018). Getting clarity by defining artificial intelligence – a survey. In Vincent C. Müller (ed.), *Philosophy and theory of artificial intelligence* (pp. 212–214). Springer.

Mumtaz, N., Izhar, T., Pandey, G., & Labhasetwar, P. K. (2022). Utilizing artificial intelligence for environmental sustainability. In *Artificial intelligence for renewable energy systems* (pp. 259–279). Elsevier.

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, & a research agenda. *International Journal of Information Management*, *53*, 1-13.

Rome, Club of. (2019). Open letter in response to the European Green Deal, https://www.clubofrome.eu/eu-green-deal-the-reaction-from, Accessed Aug 22, 2024.

Pedersen, C. S. (2018). The UN sustainable development goals (SDGs) are a great gift to business! *Procedia Cirp*, *69*, 21–24.

Perrault, R., & Clark, J. (2024). Artificial intelligence index report. Stanford University.

Reglitz, M. (2020). The human right to free internet access. *Journal of Applied Philosophy*, *37*(2), 314–331.

Ryan, J. (2010). *A history of the Internet and the digital future*. Reaktion Books.

Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., & Rockström, J. (2019). Six transformations to achieve the sustainable development goals. *Nature Sustainability*, *2*(9), 805–814.

Salvi, M., Jensen, K., Stoermer, E., Scapolo, F., Asikainen, T., & Muench, S. (2022). *Towards a green & digital future*. Publications Office of the European Union.

Santarius, T., Dencik, L., Diez, T., Ferreboeuf, H., Jankowski, P., Hankey, S., Hilbeck, A., Hilty, L. M., Höjer, M., Kleine, D., et al. (2023). Digitalization and sustainability: A call for a digital green deal. *Environmental Science & Policy*, *147*, 11–14.

Saxena, A., Ramaswamy, M., Beale, J., Marciniuk, D., & Smith, P. (2021). Striving for the United Nations (un)sustainable development goals (SDGs): What will it take? *Discover Sustainability*, *2*, 1–14.

Smith, A., & Stirling, A. (2010). The politics of social-ecological resilience and sustainable socio-technical transitions. *Ecology and Society*, *15*(1), 13.

Stober, S. S. (2019). Nature-centered leadership: Challenging the 'rules of the game'. *International Journal of Social Sustainability in Economic, Social & Cultural Context*, *15*(2), 1–13.

Subirats, J. (2002). *The dilemmas of an inevitable relationship: Democratic innovation and the information and communication technology*. Edward Elgar Publishing.

Tarcan, B., Pettersen, I. N., & Edwards, F. (2022). Making-with the environment through more-than-human design. In D. Lockton, S. Lenzi, P. Hekkert, A. Oak, J. Sádaba, & P. Lloyd (Eds.), *DRS2022: Bilbao*. Elsevier, 1-19.

Toffler, A. (1982). *Alvin Toffler's The Third Wave*. Triwave Productions Incorporated.

Trist, E. L. (1981). *The evolution of socio-technical systems*, Volume 2. Ontario Quality of Working Life Centre Toronto.

Uriarte-Gallastegi, N., Landeta-Manzano, B., Arana-Landin, G., & Laskurain-Iturbe, I. (2023). Influence of artificial intelligence on resource consumption. In *IFIP International Conference on Advances in Production Management Systems* (pp. 662–673). Springer.

Verdecchia, R., Lago, P., Ebert, C., & De Vries, C. (2021). Green IT and green software. *IEEE Software*, *38*(6), 7–15.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., S. Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso-Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, *11*(1), 1–10.

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1–37.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, *23*, 100224.

Determinants of pro-environmental attitudes and behaviours in the digital era: The case of workers in high-emitting industries in Germany

Lubna Rashid, Jordana Composto, and Elke Weber

Abstract: Addressing the interconnected challenges of digital futures and environmental pressures demands a deep understanding of human decision-making and the factors that influence sustainable behavioural change and adaptation. This study examines the drivers of pro-environmental attitudes (PEAs) and behaviours (PEBs) among workers in Germany in sectors with substantial environmental impacts, such as digital technology, mobility, and manufacturing. We focus on this worker population due to their critical role in shaping workplace values, visions, and actions and the importance of establishing inclusive and thoughtful workplace environments and structures that enhance their engagement and wellbeing, given the tensions and expectations associated with their line of work. Analyzing survey responses from 297 workers using linear regression modelling, we find varying and nuanced impacts of personal convictions, experiences, wellbeing, entrepreneurialism, and perceptions of social norms and organizational support on PEAs and various PEBs. Our results highlight the importance of purpose, leadership, emotional resilience, and inclusion in fostering a shift towards environmentally conscious practices. This research aims to guide workers, managers, and policymakers in the design of workplaces that promote, rather than hinder, ecological sustainability.

Keywords: ecological sustainability, environmental psychology, emotional and normative decision-making, behavioural change, workplace transformation, technology industry, purpose

1. Introduction

Amidst the rapid pace of technological advancements and the evolving nature of work, the urgency of safeguarding the environment has never been greater, particularly in wealthy, industrialized nations with high levels of resource consumption and a large share of global greenhouse gas emissions. While top-down, system-level approaches to environmental protection, such as new policies, technologies, and business process designs, are critical, they must be complemented by individual-level, bottom-up transformations in behaviour. This is particularly pertinent with respect to individuals working in key industries that significantly contribute to climate change emissions yet are also responsible for producing the innovations and technologies that could help mitigate long-term environmental harm.

Workers are key drivers of change as initiators of intrapreneurship and shapers of endogenous values at the organizational level (Lamm et al., 2015). Furthermore, understanding workers' decision-making and behaviour helps design support structures, workplace cultures, and leadership approaches that foster inclusion, engagement, and wellbeing, especially in high-pressure, rapidly evolving environments with high expectations for performance and impact. Although awareness of environmental issues is growing and workers are frequently exposed to climate change narratives, this increased knowledge does not always translate into effective pro-environmental decision-making or prioritization of ecological goals in future-relevant industries (Grothmann et al., 2023; Lammers et al., 2022). Evidence shows that, even in the presence of sufficient environmental knowledge and belief in climate change, predictors of pro-environmental attitudes (PEAs) and pro-environmental behaviours (PEBs) may differ (Munro et al., 2023; Park & Lin, 2020; Rashid, 2022). This contributes to the well-documented attitude-behaviour gap wherein PEAs do not fully translate to PEBs (Fahy, 2005; Farjam et al., 2019; Kennedy et al., 2009). Additionally, the blurring of boundaries between work and home life due to the rise of remote and digital-based work designs necessitates an understanding of PEA and PEB determinants in and across the private and professional spheres.

Accordingly, this study utilizes a quantitative empirical design and an online survey of workers in Germany to identify factors that shape their PEAs and PEBs, inspired by a similar recent study in the US context (Composto et al., 2023). The present survey targets employees in industries that are critical to the current global sustainability transition, namely, digital technology, transportation and mobility, extraction, and manufacturing, as well as engineering and environmental services. Understanding the unique contextual factors that impact PEAs and PEBs in this particular setting is crucial given the situational nature of decision-making predictors. This research also distinguishes between various types of PEBs – such as personal habits and engagement with others on environmental issues – both at home and in the workplace. Ultimately, we aim to specifically address the following research question: *What shapes the pro-environmental attitudes and behaviours of workers in climate-relevant industries in Germany?*

2. Research design and hypotheses

This study views worker decision-making processes and behavioural consequences from the perspective of Kahneman's (2003) expanded Dual Process Theory model and Weber and Lindemann's (2007) Functional Taxonomy of Decision Modes. Accordingly, human decision-making processes fall under two systems, one that is calculative, reflective, and effortful, requiring logical reasoning (System 2) and another that is intuitive, automatic, and effortless, highly driven by emotional experiences and social norms (System 1). This means decisions are made using a variety of mental processes, or "decision modes": on the one hand, rationally and deliberately ("with the head"); on the other hand, instinctively and emotions-driven ("with the gut") or by following rules ("by the book") (Kahneman, 2013; Weber & Lindemann, 2007). Emotions- and rule-driven decision-making is less cognitively taxing and happens automatically and in parallel to rational deliberation. When encouraged wisely, these two modes can encourage pro-environmental decisions (Reeck et al., 2022) and potentially promote positive spillover across decision contexts (Truelove et al., 2014).

With respect to environmental decisions, an understanding of personal convictions and their interplay with experiences, states, traits, and perceptions that impact emotion-based and norm-guided decision-making is central to deciphering behavioural outcomes (see Figure 1). Convictions encompass an array of ideological, political, and spiritual beliefs, as well as one's sense of the purpose of work and life and moral standing. Beliefs serve as foundational pillars in shaping an individual's self-concept, aspirations, and corresponding role in society (Blaine et al., 1998; Chen & Urminsky, 2019), including their position on environmental issues and consequent attitudes and behaviours (Maheshwari et al., 2024). Spiritual beliefs and moral standards may shape environmental values as part of larger philosophical views and existential significance. Political beliefs also reflect an individual's ideological orientation with respect to personal freedom and collective responsibility, extending to environmental issues and influencing the degree to which environmental behavioural adaptation is seen as a burden versus an obligation. Having a sense of purpose may motivate an individual to behave in ways that fulfil that purpose. As such, convictions such as spiritual beliefs, political activity and party identification, sense of purpose, and moral engagement in environmental topics are expected to directly impact PEAs and PEBs.

Hypothesis 1: Personal convictions predict pro-environmental attitudes and behaviours.

Nevertheless, acting on those convictions involves complex decision-making pathways in which emotions and social norms play a significant role. Therefore, even when an individual holds strong beliefs and values related to environmental protection, their emotional state can significantly impact their ability to act. Emotional wellbeing and personal experiences, particularly those involving extreme events, can influence the relationship between personal convictions and attitudes and actions. Additionally, certain traits may enhance emotional resilience in risky situations and further impact this relationship. In the environmental context, personal experiences with environmental calamities, such as extreme weather events, may strengthen emotions-driven decision-making via affective and associative learning pathways (Dai et al., 2015; Weber, 2010). Such experiences may also invoke empathy and subsequent pro-social behaviours (Singer & Lamm, 2009). Furthermore, heightened psychological wellbeing may enhance the resilience and self-efficacy needed to engage in new behaviours and cope with environmental stressors (Rashid, 2022), while poor mental health may undermine an individual's capacity to prioritize environmental concerns and reduce engagement with ethical behaviour (Kouchaki & Desai, 2015). Additionally, individuals with a strong entrepreneurial spirit may be more likely to demonstrate increased emotional resilience and enhanced ability to take risks, innovate, and instigate changes in attitude and behaviour against a backdrop of ambiguity and uncertainty (Conz et al., 2023). Therefore, emotional decision-making facilitators such as experience with extreme weather events (climate change experience), psychological wellbeing, and an entrepreneurial spirit are expected to activate System 1 of decision-making and thereby influence the conviction-PEA and conviction-PEB relationship.

Hypothesis 2: Emotional decision-making impacts the relationship between personal convictions and pro-environmental attitudes and behaviours.

Furthermore, a person's perception of their social milieu highly influences their own environmental attitudes and behaviours (Caggiano et al., 2023; Constantino et al., 2022). Social norms dictate which attitudes and behaviours are more appropriate and acceptable within a particular context, and individuals frequently and unknowingly engage in behavioural mimicry, unconsciously adopting the actions of those around them. When individuals observe that their colleagues or peers are engaging in specific PEBs, they are more likely to perceive these behaviours as socially desirable, creating social pressure to conform to the behaviours to avoid disapproval and maintain social harmony (Chung & Rimal, 2016). This may reinforce and normalize the adoption of PEBs within a social group, producing a cascading effect where more individuals are motivated to align their own actions with perceived social norms. Similarly, when employees perceive their organization as authentically contributing to environmental sustainability, they may be more likely to internalize these values as a form of normative alignment and organizational identification (Servaes et al., 2022). Furthermore, organizational guidance and inspiration through committed leadership and the recognition of environmentally responsible behaviours may motivate employees to take ownership of their role in environmental stewardship. Hence, the observed social norms of peers (friends and family/colleagues) and perceived organizational purpose are expected to activate System 1 of decision-making and thereby influence the conviction-PEA and conviction-PEB relationship.

Hypothesis 3: Normative decision-making impacts the relationship between personal convictions and pro-environmental attitudes and behaviours.

While our study measures PEA using a single index, we measure PEBs using three indices: PEB-Personal Habits is concerned with daily activities, such as eating habits and mobility choices, that primarily concern the individual; PEB-Engagement at Home relates to the individual's engagement and exchange with others in their private lives; PEB-Engagement at Work concerns the individual's engagement and exchange with others in their professional setting. All variables are detailed in Appendix 1.

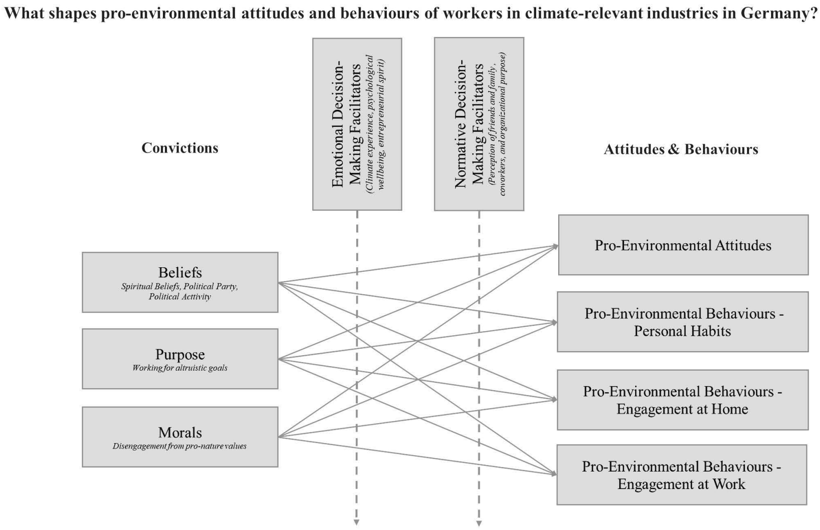

Figure 1: Research question and conceptual model.

3. Methods

A total of 297 workers in Germany participated in this study. Data collection took place in January 2024 using an approximately 20-minute survey. Although most survey items were based on existing validated measures, some measures were created for the purpose of this study (see Appendix 1). The survey was made available in both the English and German languages, and both versions were tested by several individuals for length, language, and understanding by native English and German speakers prior to formal data collection. The survey was programmed using Qualtrics software. Participants were recruited via the online panel provider Prolific, ensuring diversity across industries that are critical to the global (and local) climate transition. This includes technology (e.g. software development and hardware), transportation and mobility (e.g. automotive, aviation, and logistics), extraction and manufacturing (e.g. construction and energy), engineering (e.g. electrical installations), and environmental services. Participants' demographic information, including age, gender, ethnicity, educational level, and job role were collected to ensure sample representativeness and use as control variables. The survey and recruitment materials were reviewed and approved by the ethics review boards of both Princeton University and the Technische Universität Berlin.

Pearson correlation tests were performed to delineate the relationships between all tested variables except the control variables (see Appendix 2). Analysis of variance (independent sample t-test) was performed to test whether PEAs and PEBs significantly differ between workers operating in different industries and work modes, and linear regression modelling was then employed using SPSS software to test the conceptual model.

4. Results

The sample is predominantly male (78.1%), the average age is 32 years *(Standard Deviation = 8.5)*, 34 respondents (11.4%) identify as black, indigenous or people of colour (BIPoC), and 203 respondents are university-educated (68.4%). Among the participants, 232 reported working either completely remotely or in a hybrid format (78.1%). Industries represented include technology (55.6% of the sample), transportation, logistics, and mobility (24.2%), extractive industries such as energy, utilities, construction, manufacturing, and agriculture (16.8%), engineering, carpentry, and electrical installations (14.8%), and chemicals, plastics, and environmental services (5.4%).

The results of the ANOVA tests identify some differences between workers in the technology industry compared to other workers, as well as between workers who at least sometimes work from the office versus those who work completely remotely. Comparing the technology industry to non-technology industries reveals no significant differences in PEA and PEBs between workers except with respect to PEB-Personal Habits. In this respect, workers in the technology industry appear to demonstrate significantly higher PEB-Personal Habits compared to those who do not work in technology *(one-sided $p=0.016$)*. Comparing remote workers to others, we find that remote workers have significantly less PEB-Engagement at Work *(one-sided $p=0.030$)*, although no significant differences are observed regarding the remaining dependent variables.

Regression modelling was performed on the entire sample. No multicollinearity is detected in the regression models, and the Cronbach alpha measures of scale validity ranged between 0.6 and 0.9 for all variables. The linear regression modelling results indicate clear differences between predictors of PEA and various types of PEB; these are detailed in Tables 1–4. Significant predictors of higher PEA include older age, BIPoC identification, affiliation with left-leaning political parties, a strong sense of purpose, and higher moral engagement with environmental issues. Positive perception of environmental behaviour among friends, family members, and coworkers has a moderating effect on political activity, whereas being less politically active significantly predicts PEA only in the presence of the positive perception of those norms. A negative perception of organizational authenticity, a dimension of organizational purpose, appears to significantly predict PEA.

The results look quite different with respect to predicting PEB-Personal Habits. Greater PEB-Personal Habits are significantly predicted by not working in a leadership position. All independent variables exhibit significance, with spirituality showing a negative association with personal habits and political activity exhibiting a positive association, contrary to PEA results. Notably, the negative perception of organizational authenticity is a significant predictor, similarly to the PEA case. Climate change experience, which does not significantly predict PEA, has a significant moderator effect on both political activity and beliefs while perceptions of the behaviour of friends and family moderates the relationship between spiritual as well as political beliefs and PEB-Personal Habits.

Whereas having no leadership role at work significantly relates to PEB-Personal Habits, the opposite is true for PEB-Engagement at Home. Meanwhile, demographic variables generally seem to have no significant predictive power in the context of PEB-Engagement at Work. Higher education levels positively relate to both PEB-Engagement at Home and PEB-Engagement at Work, while identifying with the female gender only predicts PEB-Engagement at Home. Both at home and at work, purpose is the only significant predictor from the tested independent variables across all models. All other variables related to conviction appear insignificant in predicting PEB.

Zooming in on the moderator variables, greater PEB-Engagement at Home and PEB-Engagement at Work are both significantly predicted by experience with extreme weather events, psychological health, and an entrepreneurial spirit. There is no significant moderation effect on the independent variables in the case of PEB-Engagement at Work. The positive impact of belonging to a left-leaning political party on PEB-Engagement at Home is only apparent when accounting for entrepreneurial spirit or psychological health, although these effects are not seen in the workplace. The positive perception of organizational inspiration, which is also a dimension of organizational purpose, predicts PEB-Engagement at Work. Neither norm perception variable significantly predicts work or home PEB.

Table 1: Linear regression models predicting Pro-Environmental Attitudes

Linear Regression Models Predicting **PEAs**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Control Variables																
Age	**0.012****	0.004	**0.009***	0.004	**0.010***	0.004	**0.009***	0.004	**0.009***	0.004	0.006	0.004	**0.009***	0.004	0.006	0.004
Education	0.081	0.078	-0.003	0.075	-0.007	0.075	0.004	0.076	0.002	0.076	-0.019	0.069	-0.017	0.076	-0.021	0.070
Gender	0.059	0.089	0.132	0.091	0.144	0.091	0.135	0.092	0.137	0.092	0.096	0.083	0.138	0.092	0.128	0.084
Ethnic Minority	**0.418*****	0.114	**0.304****	0.111	**0.301****	0.11	**0.296****	0.111	**0.302****	0.111	**0.206***	0.101	**0.333****	0.111	**0.208***	0.102
Leadership Role	-0.068	0.076	-0.057	0.073	-0.082	0.073	-0.047	0.074	-0.051	0.074	-0.057	0.066	-0.064	0.073	-0.074	0.068
Tech Industry	0.087	0.076	-0.017	0.074	-0.013	0.073	-0.009	0.074	-0.013	0.074	-0.027	0.067	-0.036	0.075	-0.020	0.069
Remote Work	-0.141	0.097	-0.049	0.095	-0.045	0.095	-0.059	0.096	-0.049	0.095	-0.054	0.086	-0.033	0.097	-0.045	0.089
Independent Variables																
Spirituality			-0.113	0.077	-0.115	0.076	-0.114	0.077	-0.113	0.077	-0.104	0.070	-0.126	0.077	-0.106	0.070
Political Beliefs			**0.082****	0.026	**0.076****	0.026	**0.081****	0.026	**0.082****	0.026	**0.062***	0.024	**0.080****	0.027	**0.048***	0.025
Political Activity			-0.125	0.079	-0.146	0.08	-0.124	0.079	-0.125	0.079	**-0.152***	0.072	-0.131	0.080	**-0.182***	0.073
Purpose			**0.368*****	0.049	**0.348*****	0.05	**0.386*****	0.054	**0.381*****	0.057	**0.229*****	0.048	**0.349*****	0.053	**0.248*****	0.057
Moral Disengagement			**-0.099****	0.031	**-0.109*****	0.032	**-0.104*****	0.032	**-0.101****	0.032	**-0.103*****	0.028	**-0.106*****	0.031	**-0.122*****	0.029

Linear Regression Models Predicting **PEAs**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Moderator Variables																
Climate Change Experience					0.065	0.037									0.056	0.033
Psychological Wellbeing							-0.041	0.052							-0.042	0.059
Entrepreneurial Spirit									-0.026	0.058					-0.029	0.066
Friends & Family Perception											**0.345*****	0.065			**0.360*****	0.065
Coworker Perception											**0.129***	0.061			**0.125***	0.064
Organizational Contribution													0.017	0.047	0.012	0.043
Organizational Authenticity													-0.119	0.067	**-0.122***	0.062
Organizational Guidance													**0.153***	0.064	0.108	0.058
Organizational Inspiration													-0.031	0.053	-0.022	0.049
Adjusted R^2	0.049		0.324		0.326		0.318		0.317		0.444		0.327		0.455	

*p <= .05; **p <= .01; ***p <= .001

Table 2: Linear regression models predicting Pro-Environmental Behaviours – Personal Habits

Linear Regression Models Predicting **PEB – Personal Habits**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Control Variables																
Age	-0.004	0.007	-0.009	0.007	-0.007	0.007	-0.008	0.007	-0.008	0.007	-0.009	0.007	-0.010	0.007	-0.010	0.007
Education	0.178	0.126	0.041	0.124	0.031	0.123	0.030	0.125	0.034	0.125	0.080	0.125	-0.002	0.125	0.029	0.126
Gender	-0.260	0.143	-0.213	0.150	-0.184	0.149	-0.219	0.151	-0.222	0.152	-0.192	0.150	-0.221	0.151	-0.155	0.152
Ethnic Minority	0.078	0.183	-0.058	0.182	-0.065	0.180	-0.045	0.183	-0.055	0.183	-0.081	0.183	-0.037	0.183	-0.049	0.183
Leadership Role	**-0.258***	0.122	**-0.314****	0.119	**-0.369****	0.120	**-0.330****	0.121	**-0.324****	0.122	**-0.328****	0.119	**-0.312****	0.120	**-0.390****	0.123
Tech Industry	0.233	0.122	0.030	0.121	0.038	0.120	0.019	0.122	0.025	0.122	0.054	0.121	-0.009	0.124	0.018	0.123
Remote Work	-0.043	0.156	-0.066	0.157	-0.057	0.155	-0.051	0.158	-0.067	0.157	-0.052	0.156	-0.010	0.159	0.046	0.159
Independent Variables																
Spirituality			**-0.258***	0.126	**-0.261***	0.125	**-0.256***	0.126	**-0.258***	0.126	-0.231	0.126	**-0.275***	0.128	-0.239	0.126
Political Beliefs			**0.093***	0.043	0.079	0.043	**0.094***	0.043	**0.094***	0.044	0.077	0.044	0.086	0.044	0.056	0.044
Political Activity			**0.285***	0.130	0.237	0.130	**0.283***	0.130	**0.284***	0.130	**0.275***	0.129	0.247	0.131	0.185	0.131
Purpose			**0.420*****	0.080	**0.375*****	0.081	**0.390*****	0.089	**0.398*****	0.094	**0.385*****	0.087	**0.402*****	0.088	**0.318****	0.103
Moral Disengagement			**-0.129***	0.052	**-0.150****	0.052	**-0.121***	0.052	**-0.125***	0.052	**-0.121***	0.051	**-0.135***	0.052	**-0.137****	0.053
Moderator Variables																

Linear Regression Models Predicting **PEB – Personal Habits**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Climate Change Experience					**0.147***	0.060									**0.137***	0.060
Psychological Wellbeing							0.065	0.085							0.133	0.105
Entrepreneurial Spirit									0.042	0.096					-0.090	0.118
Friends & Family Perception											**0.274***	0.117			**0.309****	0.117
Coworker Perception											-0.157	0.111			-0.199	0.115
Organizational Contribution													0.091	0.078	0.111	0.077
Organizational Authenticity													**-0.247***	0.111	**-0.263***	0.111
Organizational Guidance													0.062	0.105	0.060	0.105
Organizational Inspiration													0.074	0.088	0.061	0.088
Adjusted R^2	0.033		0.216		0.233		0.214		0.213		0.229		0.221		0.250	

*p <= .05; **p <= .01; ***p <= .001

Table 3: Linear regression models predicting Pro-Environmental Behaviour – Engagement at Home

Linear Regression Models Predicting **PEB-Engagement at Home**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Control Variables																
Age	-0.006	0.010	-0.014	0.010	-0.011	0.009	-0.013	0.010	-0.013	0.010	-0.015	0.010	-0.013	0.010	-0.010	0.009
Education	**0.435***	0.184	0.138	0.174	0.112	0.166	0.093	0.174	0.075	0.173	0.137	0.177	0.087	0.176	0.032	0.171
Gender	-0.300	0.208	-0.213	0.211	-0.136	0.202	-0.237	0.209	-0.295	0.210	-0.225	0.213	-0.264	0.212	-0.228	0.206
Ethnic Minority	**0.606***	0.267	0.082	0.256	0.062	0.244	0.134	0.254	0.108	0.252	0.034	0.258	0.134	0.257	0.102	0.248
Leadership Role	**0.591*****	0.177	**0.377***	0.167	0.231	0.162	0.312	0.168	0.289	0.168	**0.375***	0.168	**0.419***	0.169	0.199	0.166
Tech Industry	0.159	0.178	-0.095	0.17	-0.076	0.162	-0.142	0.170	-0.143	0.168	-0.096	0.171	-0.188	0.175	-0.179	0.167
Remote Work	-0.273	0.228	-0.022	0.22	0.004	0.209	0.039	0.219	-0.026	0.216	-0.022	0.220	0.047	0.224	0.066	0.216
Independent Variables																
Spirituality			0.226	0.177	0.217	0.169	0.232	0.175	0.220	0.174	0.235	0.178	0.157	0.179	0.183	0.171
Political Beliefs			0.116	0.061	0.079	0.059	**0.121***	0.060	**0.124***	0.060	0.104	0.062	**0.131***	0.062	0.094	0.060
Political Activity			0.341	0.183	0.215	0.176	0.335	0.181	0.339	0.180	0.327	0.183	0.35	0.184	0.208	0.177
Purpose			**0.972*****	0.112	**0.853*****	0.110	**0.851*****	0.123	**0.777*****	0.130	**0.904*****	0.123	**0.857*****	0.124	**0.567*****	0.139
Moral Disengagement			0.005	0.072	-0.052	0.070	0.035	0.073	0.035	0.072	0.005	0.073	-0.008	0.073	-0.023	0.071

Linear Regression Models Predicting **PEB-Engagement at Home**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Moderator Variables																
Climate Change Experience					**0.393*****	0.081									**0.369*****	0.082
Psychological Wellbeing							**0.268***	0.118							0.111	0.143
Entrepreneurial Spirit									**0.376****	0.132					0.217	0.161
Friends & Family Perception											0.199	0.166			0.168	0.159
Coworker Perception											0.032	0.157			-0.034	0.155
Organizational Contribution													0.061	0.109	0.111	0.105
Organizational Authenticity													-0.119	0.155	-0.083	0.150
Organizational Guidance													0.101	0.148	0.056	0.142
Organizational Inspiration													0.138	0.124	0.059	0.119
Adjusted R^2	0.068		0.339		0.400		0.351		0.359		0.339		0.343		0.410	

*p <= .05; **p <= .01; ***p <= .001

Table 4: Linear regression models predicting Pro-Environmental Behaviour – Engagement at Work

Linear Regression Models Predicting **PEB-Engagement at Work**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Control Variables																
Age	0.005	0.011	-0.004	0.010	-0.001	0.010	-0.003	0.010	-0.003	0.010	-0.007	0.010	-0.002	0.010	-0.001	0.009
Education	**0.543****	0.191	0.259	0.181	0.230	0.172	0.202	0.180	0.196	0.180	0.202	0.182	0.157	0.178	0.059	0.171
Gender	-0.273	0.217	-0.172	0.220	-0.088	0.209	-0.203	0.217	-0.254	0.219	-0.236	0.219	-0.261	0.216	-0.250	0.206
Ethnic Minority	**0.559***	0.278	0.111	0.267	0.089	0.253	0.176	0.264	0.137	0.263	0.009	0.266	0.202	0.261	0.121	0.248
Leadership Role	**0.462***	0.185	0.299	0.174	0.139	0.168	0.217	0.174	0.211	0.175	0.311	0.173	**0.373***	0.172	0.143	0.167
Tech Industry	0.174	0.185	0.041	0.177	0.062	0.168	-0.018	0.176	-0.007	0.176	0.006	0.176	-0.129	0.177	-0.137	0.167
Remote Work	**-0.605***	0.237	-0.271	0.229	-0.242	0.217	-0.193	0.227	-0.275	0.226	-0.29	0.226	-0.137	0.227	-0.119	0.216
Independent Variables																
Spirituality			0.065	0.185	0.055	0.175	0.073	0.182	0.059	0.182	0.053	0.183	-0.055	0.182	-0.033	0.171
Political Beliefs			0.059	0.064	0.019	0.061	0.066	0.063	0.067	0.063	0.049	0.063	0.084	0.063	0.039	0.060
Political Activity			0.258	0.190	0.120	0.182	0.251	0.188	0.256	0.188	0.234	0.188	0.258	0.187	0.091	0.177
Purpose			**1.059*****	0.117	**0.929*****	0.114	**0.907*****	0.128	**0.864*****	0.136	**0.917*****	0.126	**0.856*****	0.126	**0.518*****	0.140
Moral Disengagement			0.090	0.075	0.027	0.072	0.128	0.076	0.120	0.075	0.078	0.075	0.066	0.074	0.043	0.071
Moderator Variables																

Linear Regression Models Predicting **PEB-Engagement at Work**

	Model 1 *Demographics*		**Model 2** *Demographics, Convictions*		**Model 3** *Demographics, Convictions, Emotional Decision-Making Facilitator #1 (Climate Experience)*		**Model 4** *Demographics, Convictions, Emotional Decision-Making Facilitator #2 (Psychological Wellbeing)*		**Model 5** *Demographics, Convictions, Emotional Decision-Making Facilitator #3 (Entrepreneurial Spirit)*		**Model 6** *Demographics, Convictions, Normative Decision-Making Facilitator #1 (Friends & Family/ Colleagues Perception)*		**Model 7** *Demographics, Convictions, Normative Decision-Making Facilitator #2 (Organizational Purpose)*		**Model 8** *All*	
Variable	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*	*B*	*SE B*
Climate Change Experience					**0.431*****	0.084									**0.395*****	0.082
Psychological Wellbeing							**0.338****	0.122							0.176	0.143
Entrepreneurial Spirit									**0.375****	0.137					0.162	0.161
Friends & Family Perception											0.179	0.171			0.169	0.159
Coworker Perception											0.305	0.162			0.184	0.156
Organizational Contribution													0.152	0.111	0.187	0.105
Organizational Authenticity													-0.286	0.158	-0.249	0.150
Organizational Guidance													0.182	0.150	0.102	0.142
Organizational Inspiration													**0.253***	0.126	0.177	0.119
Adjusted R^2	0.068		0.334		0.402		0.353		0.352		0.351		0.371		0.451	

*p <= .05; **p <= .01; ***p <= .001

5. Discussion

This study distinguishes the demographic, individual, and social predictors of PEAs and PEBs in a sample of workers in traditionally high-emitting, climate-key industries in Germany, most of whom work in the technology sector. The study's findings reveal interesting patterns regarding convictions and their relationship with environmental attitudes and behaviours while providing evidence for the vital role of emotions and social norms in environmental decision-making.

Workers identifying with left-leaning political parties and exhibiting higher moral engagement with environmental issues demonstrated stronger environmental attitudes, which aligns with the observations of previous studies (Currie & Choma, 2018). However, these convictions did not necessarily translate into PEBs, potentially attributable to a reluctance to discuss these topics at work, perhaps due to pluralistic ignorance (Geiger & Swim, 2016; Sparkman et al., 2022), where individuals underestimate the extent to which others share their views. A related finding is that having spiritual or religious beliefs relates negatively to environmental habits, contradicting findings in other contexts such as China (Li et al., 2023) and Malaysia (Abdullah & Keshminder, 2020). This might be due to a preference for adhering to familiar, norm-conforming habits rather than adopting changes that might challenge (conservative) values.

Notably, having a higher sense of purpose, which involves embracing pro-social, community-oriented, and altruistic values, emerged as a significant predictor for all dependent variables. This suggests that – irrespective of political and spiritual beliefs – an intrinsic altruistic drive and passion for making a positive impact on the world may be a most critical determinant of sustainable behaviours with regards to personal convictions. This underscores the importance of fostering a sense of purpose and altruism to promote pro-environmental actions across the entire spectrum of ideological perspectives and might also suggest that having a sense of purpose does not necessarily intertwine with religious or spiritual beliefs, as previously expected (Oishi & Diener, 2013). Given the young sample in this study (average age = 32), this potentially provides evidence for the contextual and generational views of purpose and spirituality; however, further research is needed in this regard.

The study found that personal experience with climate change was a strong predictor of the PEBs examined. This suggests that direct exposure to the impacts of climate change may play a crucial role in motivating individuals to adopt more sustainable behaviours, confirming findings from previous research (Broomell et al., 2015; Constantino et al., 2022). This implies that simulating future impacts of environmental issues in a personal and tangible manner, such as through engaging storytelling about a beloved city's future under climate change (e.g. Staud & Reimer, 2021) – could bolster PEBs by activating emotions-driven decision-making processes. Also in-line with our expectations, higher levels of psychological wellbeing and entrepreneurial spirit were found to promote engagement with others on environmental topics, both at home and in the workplace, indicating that emotional resilience and an innovative mindset might be essential for fostering proactive environmental collaboration.

Interestingly, the relationship between political ideology and environmental engagement at home was moderated by psychological wellbeing and entrepreneurial spirit. This finding suggests that these personal attributes may influence how political beliefs translate into action in the domestic sphere. It is possible that those individuals are better equipped to navigate potential ideological differences and identify creative and collaborative solutions to environmental challenges within their households. Further research could help clarify the specific mechanisms underlying these relationships and their implications.

The study's findings also provide intriguing insights into the role of the perception of social norms in shaping PEAs and PEBs. Perception of friends and family as well as colleagues' environmental behaviour significantly relates to an individual's PEA. Also, the perception of the behaviour of friends and family members was found to moderate the relationship between spiritual and political beliefs and PEB-Personal Habits. This suggests that individuals may prioritize the perceived expectations of their close social circles over their personal convictions when adopting particular lifestyle practices. Surprisingly, the perception of the behaviour of friends, family, and colleagues did not significantly relate to PEB engagement neither at home nor at work. This finding indicates that social norms might not always play a significant role in motivating environmental action, warranting context-specific behavioural norm interventions (Constantino et al., 2021).

Unexpectedly, negative perceptions of organizational authenticity were associated with higher PEAs and PEB-Habits. This could imply that individuals attempt to compensate for their organization's perceived shortcomings in their private lives without feeling secure enough to engage with others on these matters. The findings also signal the ambiguous contribution of organizational authenticity to PEB (Lopes et al., 2023). On the other hand, an inspiring work environment embodying a higher cause seems to motivate workers to engage with environmental initiatives in the workplace, perhaps due to the expectation of being rewarded for such actions. However, this positive perception of organizational inspiration did not relate to PEA, PEB-Personal Habits, or PEB-Engagement at Home, suggesting the distinct role of organizational culture in shaping environmentally focused behaviour within professional settings (Robertson & Barling, 2013).

Our results also indicate that being a BIPoC individual significantly predicts PEA when accounting for all variables in the study (Models 1–8 in Table 1), while the impact of being a BIPoC on PEB at home and at work is less prominent. This may suggest that BIPoC individuals feel hesitant to express or act upon their pro-environmental views. Explanations could be feelings of exclusion or "invisibility" in the workplace (see Buchanan & Settles, 2019), potentially fuelled by the rise in cases of racism and islamophobia in Germany particularly since October 2023 (Deutsche Welle, 2024a, 2024b). This provides evidence for the importance of creating safe spaces and authentic inclusion and empowerment of workers across ethnic backgrounds to truly enhance sustainability. However, further research is needed to explain this finding. Additionally, older age was associated with stronger PEAs, but given the relatively young average sample age, this finding should be interpreted with caution.

Higher education levels were linked to increased PEB both at home and at work but not necessarily with stronger PEA or PEB-Personal Habits. This could be attributable to the confidence that higher education provides, empowering individuals to motivate others to adopt environmentally friendly practices. Furthermore, those in professional leadership positions were found to engage with others both at home and work, but having a leadership position was negatively associated with practising personal environmental habits. This may suggest a tendency towards intellectual virtue signalling (Levy, 2023), particularly among leaders and also those with a higher education level, but further research is needed to explain this finding. It may also be that leaders perceive engagement with sustainability and environmental issues to be a popular trend or a social expectation, prompting them to advocate for it without necessarily internalizing the importance of practising sustainable behaviours in their own lives.

Finally, we find that working in the technology sector might be linked to more environmentally friendly personal habits than employment in non-tech industries. This could be due to the workplace location flexibility that tech jobs often offer. Working from home, which may be more common in such fields, can reduce the need for daily commuting or enable workers to have more control over their dietary choices. It may also be due to selection bias regarding the type of workers that join the tech industry. This possibility needs further explanation, appearing to conflict with recent research suggesting that technology company leaders in Germany might be less environmentally inclined than their counterparts in other industries (Lammers et al., 2022). Meanwhile, our results also show that working solely remotely appears to have a significant and negative impact on PEB with respect to workplace engagement. This could be due to the reduced workplace embeddedness and sense of organizational belonging experienced by online workers, which diminishes the motivation or ability to engage in shaping organizational values or investing in workplace awareness activities and community initiatives.

6. Conclusion

Purpose, understood as striving to make the world a better place through one's work, is the unifying predictor of PEAs and PEBs across home and work domains. Workers increasingly seek mission-driven organizations; combined with the rapid digital transition across all work domains, this will inevitably impact technology-based work. Our study offers novel insights into the environmentally impactful attitudes and behaviours of workers in Germany mainly operating in the technology sector. Addressing climate change undoubtedly requires that organizations make immense and fast changes. Workers are a crucial part of achieving that change and operationalizing new plans, products, and services. The unifying role of purpose suggests that organizations should adopt an environmental mission and that workers are, already, finding outlets for PEB at work and at home, with tech workers appearing more willing or able than others to embrace pro-environmental habits.

Furthermore, cultivating a work environment that supports emotional resilience, personal growth, and work-life balance can positively influence both employee wellbeing and environmental sustainability, while empowering underrepresented minorities to share their voice may foster inclusive change. Leaders should not only actively foster a purposeful, wellbeing-promoting, and inclusive workplace but be open to learning from team members about environmental habits. In uncertain, increasingly digital futures, the importance of these initiatives is amplified as technology reshapes the nature of work and introduces new challenges. Embracing digital tools and practices while maintaining a focus on purpose, wellbeing, and inclusivity will be crucial to thriving in this evolving landscape, ensuring that workers remain resilient and adaptive in the face of uncertainty while being part of the solution rather than the problem with respect to environmental sustainability.

Our study features some noteworthy limitations. First, because data collection was based on self-reported survey responses, subjectivity in responses and social desirability bias may be inevitable. Second, the data and analyses are correlational and cannot provide evidence about causal relationships between measured variables. Furthermore, the analyses are based on a relatively small sample of workers who were all recruited using a single platform. This may have consequences for the representativeness of the sample and, therefore, the generalizability of our results.

References

Abdullah, M. S., & Keshminder, J. S. (2020). Religion and pro-environmental behaviour: A comparative analysis towards environmental protection. *International Journal of Environment and Sustainable Development*, *19*(2), 174–194.

Blaine, B. E., Trivedi, P., & Eshleman, A. (1998). Religious belief and the self-concept: Evaluating the implications for psychological adjustment. *Personality and Social Psychology Bulletin*, *24*(10), 1040–1052. https://doi.org/10.1177/01461672982410002

Broomell, S. B., Budescu, D. V., & Por, H.-H. (2015). Personal experience with climate change predicts intentions to act. *Global Environmental Change*, *32*, 67–73. https://doi.org/10.1016/j.gloenvcha.2015.03.001

Buchanan, N. T., & Settles, I. H. (2019). Managing (in)visibility and hypervisibility in the workplace. *Journal of Vocational Behavior*, *113*, 1–5. https://doi.org/10.1016/j.jvb.2018.11.001

Caggiano, H., Constantino, S. M., Lees, J., Majumdar, R., & Weber, E. U. (2023). Community-engaged research is best positioned to catalyze systemic change. *The Behavioral and Brain Sciences*, *46*, e152. https://doi.org/10.1017/S0140525X23001024

Chen, S. Y., & Urminsky, O. (2019). The role of causal beliefs in political identity and voting. *Cognition*, *188*, 27–38. https://doi.org/10.1016/j.cognition.2019.01.003

Chung, A., & Rimal, R. (2016). Social norms: A review. *Review of Communication Research*, *4*, 1–28. https://doi.org/10.12840/issn.2255-4165.2016.04.01.008

Composto, J. W., Constantino, S. M., & Weber, E. U. (2023). Predictors and consequences of pro-environmental behavior at work. *Current Research in Ecological and Social Psychology*, *4*, 100107. https://doi.org/10.1016/j.cresp.2023.100107

Constantino, S. M., Pianta, S., Rinscheid, A., Frey, R., & Weber, E. U. (2021). The source is the message: The impact of institutional signals on climate change-related norm perceptions and behaviors. *Climatic Change*, *166*(3–4), 35. https://doi.org/10.1007/s10584-021-03095-z

Constantino, S. M., Sparkman, G., Kraft-Todd, G. T., Bicchieri, C., Centola, D., Shell-Duncan, B., Vogt, S., & Weber, E. U. (2022). Scaling up change: A critical review and practical guide to harnessing social norms for climate action. *Psychological Science in the Public Interest*, *23*(2), 50–97. https://doi.org/10.1177/15291006221105279

Conz, E., Magnani, G., Zucchella, A., & De Massis, A. (2023). Responding to unexpected crises: The roles of slack resources and entrepreneurial attitude to build resilience. *Small Business Economics*, *61*(3), 957–981. https://doi.org/10.1007/s11187-022-00718-2

Currie, S., & Choma, B. (2018). Sociopolitical ideology and the morality of green behaviour. *Environmental Politics*, *27*(2), 247–266. https://doi.org/10.1080/09644016.2017.1413727

Dai, J., Kesternich, M., Löschel, A., & Ziegler, A. (2015). Extreme weather experiences and climate change beliefs in China: An econometric analysis. *Ecological Economics*, *116*, 310–321. https://doi.org/10.1016/j.ecolecon.2015.05.001

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research*, *97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Deutsche Welle. (2024a, January 12). *Muslims in Germany: Life post-Hamas-attack "like after 9/11."* Deutsche Welle. https://www.dw.com/en/muslims-in-germany-life-post-hamas-attack-like-after-9-11/a-67959092

Deutsche Welle. (2024b, June 25). *Germany: Annual discrimination report logs 22% rise in cases*. Deutsche Welle. https://www.dw.com/en/germany-annual-discrimination-report-logs-22-rise-in-cases/a-69463601

Fahy, F. (2005). The right to refuse: Public attitudes and behaviour towards waste in the west of Ireland. *Local Environment*, *10*(6), 551–569. https://doi.org/10.1080/13549830500321618

Farjam, M., Nikolaychuk, O., & Bravo, G. (2019). Experimental evidence of an environmental attitude-behavior gap in high-cost situations. *Ecological Economics*, *166*, 106434. https://doi.org/10.1016/j.ecolecon.2019.106434

Geiger, N., & Swim, J. K. (2016). Climate of silence: Pluralistic ignorance as a barrier to climate change discussion. *Journal of Environmental Psychology*, *47*, 79–90. https://doi.org/10.1016/j.jenvp.2016.05.002

Grothmann, T., Frick, V., Harnisch, R., Münsch, M., Kettner, S. E., & Thorun, C. (2023). *Umweltbewusstsein in Deutschland 2022*. Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV). https://www.umweltbundesamt.de/publikationen/umweltbewusstsein-in-deutschland-2022

Jasinenko, A., & Steuber, J. (2023). Perceived organizational purpose: Systematic literature review, construct definition, measurement and potential employee outcomes. *Journal of Management Studies*, *60*(6), 1415–1447. https://doi.org/10.1111/joms.12852

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D. (2013). *Thinking, fast and slow* (Reprint edition). Farrar, Straus and Giroux.

Kennedy, E., Beckley, T., Mcfarlane, B., & Nadeau, S. (2009). Why we don't "walk the talk": Understanding the environmental values/behaviour gap in Canada. *Human Ecology Review*, *16*(2), 151–160.

Kouchaki, M., & Desai, S. D. (2015). Anxious, threatened, and also unethical: How anxiety makes individuals feel threatened and commit unethical acts. *Journal of Applied Psychology*, *100*(2), 360–375. https://doi.org/10.1037/a0037796

Lamm, E., Tosti-Kharas, J., & King, C. E. (2015). Empowering employee sustainability: Perceived organizational support toward the environment. *Journal of Business Ethics*, *128*(1), 207–220. https://doi.org/10.1007/s10551-014-2093-z

Lammers, T., Rashid, L., Kratzer, J., & Voinov, A. (2022). An analysis of the sustainability goals of digital technology start-ups in Berlin. *Technological Forecasting and Social Change*, *185*, 122096. https://doi.org/10.1016/j.techfore.2022.122096

Levy, N. (2023). Intellectual virtue signaling. *American Philosophical Quarterly*, *60*(3), 309–322.

Li, J., Zhai, S., & Li, X. (2023). Religion and enterprise pollution behavior: Evidence from China. *Journal of Cleaner Production*, *384*, 135454. https://doi.org/10.1016/j.jclepro.2022.135454

Lopes, J. M., Gomes, S., & Trancoso, T. (2023). The dark side of green marketing: How greenwashing affects circular consumption? *Sustainability*, *15*(15), Art. 15. https://doi.org/10.3390/su151511649

Maheshwari, A., Bhagwat, S., & Karpate, Y. (2024). How can faith-based approaches protect biodiversity? *Journal for Nature Conservation*, 126578. https://doi.org/10.1016/j.jnc.2024.126578

Meynhardt, T., Steuber, J., & Feser, M. (2023). The Leipzig Leadership Model: Measuring leadership orientations. *Current Psychology*. https://doi.org/10.1007/s12144-023-04873-x

Munro, P., Kapitan, S., & Wooliscroft, B. (2023). The sustainable attitude-behavior gap dynamic when shopping at the supermarket: A systematic literature review and framework for future research. *Journal of Cleaner Production*, *426*, 138740. https://doi.org/10.1016/j.jclepro.2023.138740

Oishi, S., & Diener, E. (2013). Residents of poor nations have a greater sense of meaning in life than residents of wealthy nations. *Psychological Science*, *25*(2). https://doi.org/10.1177/0956797613507286

Park, H. J., & Lin, L. M. (2020). Exploring attitude–behavior gap in sustainable consumption: Comparison of recycled and upcycled fashion products. *Journal of Business Research*, *117*, 623–628. https://doi.org/10.1016/j.jbusres.2018.08.025

Qin, X., Shepherd, D. A., Lin, D., Xie, S., Liang, X., & Lin, S. (2020). The dark side of entrepreneurs' creativity: Investigating how and when entrepreneurs' creativity increases the favorability of potential opportunities that harm nature. *Entrepreneurship Theory & Practice*. https://doi.org/10.1177/1042258720915582

Rashid, L. (2022). Bursting the bubble: Why sustainability initiatives often lack adequate intention to action translation. *Small Business Economics*, 1–9. https://doi.org/10.1007/s11187-022-00599-5

Reeck, C., Gamma, K., & Weber, E. U. (2022). How we decide shapes what we choose: Decision modes track consumer decisions that help decarbonize electricity generation. *Theory and Decision*, *92*(3), 731–758. https://doi.org/10.1007/s11238-022-09874-z

Robertson, J. L., & Barling, J. (2013). Greening organizations through leaders' influence on employees' pro-environmental behaviors. *JOURNAL OF ORGANIZATIONAL BEHAVIOR*, *34*(2), 176–194. https://doi.org/10.1002/job.1820

Servaes, M., Nguyen Thi Thu, H., Kluijtmans, T., & Crucke, S. (2022). Don't talk the talk, but walk the walk: The role of authentic CSR in fostering beneficial employee outcomes. *Management Decision*, *61*(3), 569–588. https://doi.org/10.1108/MD-11-2021-1515

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*(1), 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Sparkman, G., Geiger, N., & Weber, E. U. (2022). Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nature Communications*, *13*(1), 4779. https://doi.org/10.1038/s41467-022-32412-y

Staud, T., & Reimer, N. (2021). *Deutschland 2050: Wie der Klimawandel unser Leben verändern wird* (8. edition). KiWi-Paperback.

Truelove, H. B., Carrico, A. R., Weber, E. U., Raimi, K. T., & Vandenbergh, M. P. (2014). Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Global Environmental Change*, *29*, 127–138. https://doi.org/10.1016/j.gloenvcha.2014.09.004

Weber, E. U. (2010). What shapes perceptions of climate change? *WIREs Climate Change*, *1*(3), 332–342. https://doi.org/10.1002/wcc.41

Weber, E. U., & Lindemann, P. G. (2007). From intuition to analysis: Making decisions with our head, our heart, or by the book. In H. Plessner, C. Betsch, & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 191–208). Lawrence Erlbaum.

Appendix 1: Survey Items and Measures

Variable Type	Variable Name	Item	Response Type	Source
Scoping	Industry Branch	Which of the following best describes the industry that you, personally, work in (regardless of your job title)?	Multiple Choice	(Composto et al., 2023)
Control	Age	How old are you in years?	Open Text	
Control	Education	What is your highest obtained academic qualification?	Single Choice	
Control	Gender	What is your gender?		
Control	Ethnic Minority	Do you consider yourself a person of colour/BIPoC?		
Control	Leadership Role	Which of the following best describes your role at work?		
Independent	Spirituality	Which of the following best describes your religious and spiritual beliefs?		-
Independent	Political Beliefs	Which political party do you most associate with?		(Composto, Constantino, and Weber 2023)
Independent	Political Activity	In the last 5 years, have you participated in any political action or protest?		
Independent	Purpose	I want my work to benefit society	Likert Scale (1–6)	(Meynhardt et al., 2023)
		I want my work to contribute to changing the world for the better		
		I focus what I do on the common good		
		I check whether my decisions are in line with social expectations		
		I forgot personal gain if it harms the common good		
		I take the common good into account in conflicts		
		I am passionate about work that has social relevance		
		I am proud if I can make a contribution to society with my work		
		I motivate employees to see their work in a larger context than our organization		
		I think that organizations should be guided mainly by their social function		
Independent	Moral Disengagement	It is okay to sacrifice the natural environment for economic development	Likert Scale (1–7)	(Qin et al., 2020)
		Polluting the natural environment is okay as long as you can restore it later		

Variable Type	Variable Name	Item	Response Type	Source
		Considering that other people are destroying the natural environment, it's hardly a big deal to pollute it casually		
		People shouldn't be held accountable for doing things that harm the natural environment if an authority figure encouraged them to do so		
		People can't be blamed for doing things that are bad for the natural environment when all their friends are doing it too		
		Doing things that temporarily harm the natural environment is no big deal		
		The natural environment can be harmed because it can restore itself		
		Nature was holding back human survival and advancement		
Moderator	Climate Change Experience	I have been experiencing extreme weather events more frequently over the past three years	Likert Scale (1–5)	-
		At least one extreme weather event in the past three years has caused me or someone close to me physical harm (e.g. bodily injury)		-
		At least one extreme weather event in the past three years has caused me or someone close to me material harm (e.g. property damage)		-
		At least one extreme weather event in the past three years has caused me or someone close to me financial harm (e.g. job or income loss)		-
		At least one extreme weather event in the past three years has caused me or someone close to me mental harm (e.g. emotional or psychological hardship)		-
Moderator	Psychological Well-being	I lead a purposeful and meaningful life	Likert Scale (1–7)	(Diener et al., 2010)
		My social relationships are supportive and rewarding		
		I am engaged and interested in my daily activities		
		I actively contribute to the happiness and well-being of others		
		I am competent and capable in the activities that are important to me		
		I am a good person and live a good life		
		I am optimistic about my future		
		People respect me		
Moderator	Entrepreneurial Spirit	I want to help new things get off the ground	Likert Scale (1–6)	(Meynhardt et al., 2023)

Variable Type	Variable Name	Item	Response Type	Source
		I take risks in order to try out new things		
		I am enthusiastic about new ideas		
		I solve problems creatively		
		I encourage people around me to try new things		
		I see change as positive		
		I am a fountain of ideas		
		I find it easy to let go of routines		
		I want to develop as a person		
		I look for challenges		
Moderator	Friends & Family Perception	"Companies should prioritize, equally, their workers, customers, suppliers, community, and the environment." Most of my friends would…	Likert Scale (1–5)	(Composto, Constantino, and Weber 2023)
		"The private sector has caused environmental degradation." Most of my friends would…		
		"Companies that do not adapt and address climate change now will become less successful over time." Most of my friends would…		
		"Policies and regulations are the best way to address climate change." Most of my friends would…		
		"Innovation and technology are the best way to address climate change." Most of my friends would…		
		"Individuals have a responsibility to protect the environment." Most of my friends would…		
Moderator	Coworker Perception	"Companies should prioritize, equally, their workers, customers, suppliers, community, and the environment." Most of my coworkers would…		
		"The private sector has caused environmental degradation." Most of my coworkers would…		
		"Companies that do not adapt and address climate change now will become less successful over time." Most of my coworkers would…		
		"Policies and regulations are the best way to address climate change." Most of my coworkers would…		
		"Innovation and technology are the best way to address climate change." Most of my coworkers would…		
		"Individuals have a responsibility to protect the environment." Most of my coworkers would…		
Moderator	Organizational Support	My organization aims to contribute to the common good	Likert Scale (1–6)	(Jasinenko & Steuber, 2023)

Variable Type	Variable Name	Item	Response Type	Source
	Perceived Contribution	My organization seeks to create a positive change in the world		
		My organization aims to achieve something that goes beyond its own benefit		
	Organizational Sup-port *Perceived Authenticity*	My organization remains true to its core values even when conflicts occur		
		My organization is fully committed to its overarching goals		
		My organization credibly embodies its core values		
	Organizational Sup-port *Perceived Guidance*	My organization's overarching goals provide orientation in complex situations		
		My organization's higher goals guide decisions and actions		
		My organization's overarching goals provide stable guidance in times of rapid change		
	Organizational Sup-port *Perceived Inspiration*	My organization unites through inspiring higher goals		
		My organization conveys the idea of being part of something bigger		
		My organization inspires by providing a higher cause		
Dependent	Personal Environ-mental Attitudes (PEA)	"Companies should prioritize, equally, their workers, customers, suppliers, community, and the environ-ment." I…	Likert Scale (1–5)	(Composto, Constantino, and Weber 2023)
		"The private sector has caused environmental degradation." I…		
		"Companies that do not adapt and address climate change now will become less successful over time." I…		
		"Policies and regulations are the best way to address climate change." I…		
		"Innovation and technology are the best way to address climate change." I…		
		"Individuals have a responsibility to protect the environment." I		
Dependent	Personal Pro-Environmental Habits (PEB-Personal Habits)	I choose meatless meal options	Likert Scale (1–7)	-
		I use public transport of a bicycle to get around even if it takes longer to get to my destination		
		I wear recycled or second-hand clothing		
		I use reusable bags and packaging		
		I keep the air-conditioning and heating devices off even if it's uncomfortably hot or cold		
Dependent	Home Pro-Environ-mental	I give my time to help my friends and family take the environment into account in their daily activities		(Composto, Constantino, and Weber 2023)
		I actively participate in environmental events organized in and/or by my community		

Variable Type	Variable Name	Item	Response Type	Source
	Engagement (PEB-Home)	I voluntarily carry out environmental actions and initiatives in my daily life activities		
		I make suggestions to my friends and family about ways to protect the environment more effectively, even when it is not my responsibility		
Dependent	Work Pro-Environmental Engagement (PEB-Work)	I give my time to help my colleagues take the environment into account in what they do at work		
		I actively participate in environmental events organized in and/or by my company		
		I voluntarily carry out environmental actions and initiatives in my daily work activities		
		I make suggestions to my colleagues about ways to protect the environment more effectively, even when it is not my direct responsibility		

Appendix 2: Pearson correlation coefficients between all tested variables. * indicates two-tailed significance at the 0.05 level; ** indicates two-tailed significance at the 0.01 level.

	PEA	PEB - Personal Habits	PEB - Engagement at Home	PEB - Engagement at Work	Spiritual Beliefs	Political Beliefs	Political Activity	Purpose	Moral Disengagement	Climate Change Experience	Psychological Wellbeing	Entrepreneurial Spirit	Friends & Family Perception	Coworker Perception	Organizational Contribution	Organizational Authenticity	Organizational Guidance	Organizational Inspiration
PEA	1	0.382**	0.361**	0.311**	0.047	0.271**	-0.003	0.477**	-0.222**	0.230**	0.226**	0.261**	0.545**	0.424**	0.149**	0.134*	0.198**	0.175**
PEB - Personal Habits	0.382**	1	0.477**	0.391**	-0.030	0.233**	0.179**	0.321**	-0.212**	0.202**	0.171**	0.154**	0.296**	0.128*	0.150**	0.040	0.059	0.105
PEB - Engagement at Home	0.361**	0.477**	1	0.781**	0.217**	0.155*	0.197**	0.570**	-0.030	0.402**	0.398**	0.433**	0.311**	0.277**	0.297**	0.244**	0.274**	0.306**
PEB - Engagement at Work	0.311**	0.391**	0.781**	1	0.209**	0.094	0.136*	0.562**	0.048	0.406**	0.416**	0.436**	0.323**	0.347**	0.366**	0.281**	0.366**	0.397**
Spiritual Beliefs	0.047	-0.030	0.217**	0.209**	1	-0.079	0.096	0.305**	0.044	0.132*	0.175**	0.205**	0.095	0.192**	0.218**	0.191**	0.218**	0.235**
Political Beliefs	0.271**	0.233**	0.155*	0.094	-0.079	1	0.156*	0.078	-0.196**	0.149*	0.016	0.002	0.195**	0.034	-0.120	-0.143*	-0.115	-0.092
Political Activity	-0.003	0.179**	0.197**	0.136*	0.096	0.156*	1	0.164**	0.005	0.178**	0.078	0.077	0.101	0.030	0.008	-0.064	-0.015	-0.010
Purpose	0.477**	0.321**	0.570**	0.562**	0.305**	0.078	0.164**	1	-0.075	0.242**	0.519**	0.580**	0.388**	0.405**	0.427**	0.364**	0.397**	0.425**
Moral Disengagement	-0.222**	-0.212**	-0.030	0.048	0.044	-0.196**	0.005	-0.075	1	0.159**	-0.153**	-0.067	-0.046	0.088	0.018	0.027	0.079	0.075
Climate Change Experience	0.230**	0.202**	0.402**	0.406**	0.132*	0.149*	0.178**	0.242**	0.159**	1	0.144*	0.207**	0.149*	0.161**	0.070	0.047	0.131*	0.129*
Psychological Wellbeing	0.226**	0.171**	0.398**	0.416**	0.175**	0.016	0.078	0.519**	-0.153**	0.144*	1	0.693**	0.232**	0.269**	0.313**	0.367**	0.351**	0.368**
Entrepreneurial Spirit	0.261**	0.154**	0.433**	0.436**	0.205**	0.002	0.077	0.580**	-0.067	0.207**	0.693**	1	0.294**	0.248**	0.253**	0.279**	0.326**	0.363**
Friends & Family Perception	0.545**	0.296**	0.311**	0.323**	0.095	0.195**	0.101	0.388**	-0.046	0.149*	0.232**	0.294**	1	0.535**	0.175**	0.197**	0.201**	0.229**
Coworker Perception	0.424**	0.128*	0.277**	0.347**	0.192**	0.034	0.030	0.405**	0.088	0.161**	0.269**	0.248**	0.535**	1	0.352**	0.304**	0.359**	0.371**
Organizational Contribution	0.149**	0.150**	0.297**	0.366**	0.218**	-0.120	0.008	0.427**	0.018	0.070	0.313**	0.253**	0.175**	0.352**	1	0.716**	0.645**	0.699**
Organizational Authenticity	0.134*	0.040	0.244**	0.281**	0.191**	-0.143*	-0.064	0.364**	0.027	0.047	0.367**	0.279**	0.197**	0.304**	0.716**	1	0.767**	0.706**
Organizational Guidance	0.198**	0.059	0.274**	0.366**	0.218**	-0.115	-0.015	0.397**	0.079	0.131*	0.351**	0.326**	0.201**	0.359**	0.645**	0.767**	1	0.782**
Organizational Inspiration	0.175**	0.105	0.306**	0.397**	0.235**	-0.092	-0.010	0.425**	0.075	0.129*	0.368**	0.363**	0.229**	0.371**	0.699**	0.706**	0.782**	1

Uncertain futures of work: The perception of generative AI in knowledge professions

Florian Butollo, Jennifer Haase, Ann-Kathrin Katzinski, and Anne K. Krüger

Abstract: The application of generative AI (GenAI) tools has led to widespread speculation about the implications of technological change for the future of cognitive work. This article provides insights on how the use of GenAI affects work practices in the fields of IT programming, science and coaching based on expert interviews and a quantitative survey among users of GenAI. Specifically, we ask about perceptions on skills, creativity, and authenticity, which we regard as key qualities of cognitive work. Contrary to widespread expectations that AI use would hollow out or substitute aspects of cognitive work, we find that there is a strong awareness for the meaning of the professional core in each field. We conclude that the use of AI provokes reflections about the meaning of human work in operating AI tools adequately and taking on responsibility for their results, thereby rather reinforcing its relevance.

Keywords: artificial intelligence, cognitive work, digitalization, skills, creativity, authenticity

1. Introduction

Rapid progress in the development of AI systems has – once again – inspired speculations about the future of work. Unlike in prior periods of rapid automation, however, uncertainties and concerns about the impact of technology do not concern repetitive tasks of manual labour but knowledge work (Brynjolfsson et al., 2023; Dell'Acqua et al., 2023). Seemingly, the predictions in the much-acclaimed book *The second machine age* (2014) are becoming real. In this book, Eric Brynjolfsson and Andrew McAfee argue that AI technologies will have an impact on cognitive fields of work comparable to the effect of industrial automation on blue-collar work. As in similar texts, the perspective is one of a 'race against the machine', the title of another work by the same authors (Brynjolfsson & McAfee, 2012; see also Acemoglu & Restrepo, 2018), to avoid the substitution of human work by technology.

Only ten years ago, such statements very much seemed to address a distant future; now, they appear to become true in the present with the advent of generative AI (genAI), an incarnation of AI that progresses into fields that were thought to be the realm of human action. By writing poems, creating graphics, or composing music, it seems to possess skills that surpass those of many humans; genAI also seems to possess creativity in recombining and readjusting information, and the results appear to be authentic in the sense that it is hard to judge whether they result from human or artificial intelligence.

The presentation of ChatGPT by Open.AI was aptly described as the 'iPhone moment' of AI by the CEO of the chip producer Nvidia due to the degree of public attention it attracted and the extraordinarily high pace by which it was picked up by consumers. A new practical dimension began to enter the hitherto rather speculative discourse about AI, as millions of users began to experiment with ChatGPT and shared their experiences through social media, including ridiculous failures but also astonishing accomplishments of so far unprecedented capabilities.

Yet, even though everyone can now try and experiment with ChatGPT, we still do not have a more empirically grounded debate. On the contrary, it seems that expectations and speculations are rising to new peaks with the public excitement that follows each release of a new genAI application. On the one hand, much of the discussion follows the established patterns of a race between humans and machines. A study by Goldman Sachs projects a 'significant disruption' of labour markets and calculates that up to one-quarter of all jobs might be substituted through genAI-based automation (Hatzius et al., 2023). Another study that looks at the susceptibility of work tasks through genAI estimates that up to 49% of jobs contain tasks that could be taken over by genAI. The authors argue that genAI appears 'to be particularly impactful for highly-educated and highly-paid workers' and make policy suggestions about how to mitigate disruptive effects on the labour market (Felten et al., 2023). Here, genAI once again appears to be a job killer, but this time on steroids, endangering knowledge-intensive professions.

On the other hand, studies also predict job growth and discuss the potential of the 'augmentation' of human work through genAI. A study on the use of ChatGPT in mid-level professional writing tasks finds an increase in work productivity and a decrease in inequality among workers as ChatGPT benefits 'low-ability workers' more (Noy & Chang, 2023). A similar experimental investigation on the impact of ChatGPT in the field of customer support work similarly concludes that there is a significant improvement for novice and low-skilled workers but a minimal impact on experienced and highly skilled workers (Brynjolfsson et al., 2023).

While such studies help us formulate some hypotheses on the impact of genAI in the workplace, their empirical scope remains limited. Systematic research about the actual impact of genAI on professions is still in its infancy. In this chapter, we set out to explore how genAI is currently perceived and used in knowledge work. Focusing on **professional skills**, **creativity**, and **authenticity** as crucial aspects of knowledge work, we ask how genAI might change its professional core. Based on qualitative expert interviews and a quantitative survey, we investigate the perception and use of genAI in three fields of knowledge work that are likely to be particularly exposed: **academic work** in science, **IT programming** in industry, and the **coaching sector**.

Our results belie the expectation that human expertise and skills lose importance. Our study rather shows the contrary: debates and experiences with genAI help to sharpen and value the **core of the professional identity.** The debate on genAI thus helps people working in the respective sectors to see more clearly what constitutes the core of their profession (1) in terms of the **skills** needed not only to accomplish work tasks but also in designing tasks and the projects behind them, (2) in terms of **creativity** understood as original approaches to task fulfilment, and (3) in terms of **authenticity**, which highlights the need for the explicability of tasks and responsibility for results. Our study thus also highlights that professions consist of more than an addition of single work tasks. They contain experiential and tacit knowledge about how to frame, prepare, and interpret steps that are difficult for machines to replicate. However, there are also concerns among the interview partners that professions could be hollowed out and that the quality of products and services could deteriorate as automated 'good-enough-versions' of the former offers become commonplace.

The remainder of this contribution is structured as follows: in the next section, we briefly introduce a framework by which we approach the relationship between humans and genAI that connects insights from critical informatics with theoretical perspective, emphasising the complementarity between technology and (human work). In section 3, we present the methods for gathering and analysing our data. In section 4, we consecutively present the results from the fields of science, IT programming, and coaching. In the last section, we discuss the results comparatively and draw conclusions to understand the impact of genAI on knowledge work.

2. Research framework

In approaching the impact of genAI on knowledge work in the workplace, we connected two theoretical threads that caution against oversimplified expectations about the substitution of work through technology. The first are perspectives of **critical computer science** as pioneered by Joseph Weizenbaum and further developed by academics like Wolfgang Coy, Robert Kling, Katharina Zweig, and Gary Marcus. Weizenbaum was a protagonist of AI research in the 1960s–1980s. As a professor at MIT, he invented ELIZA, one of the first chatbots. At the same time, however, he was an outspoken critic of exaggerated expectations about AI and cautioned that society needed to set goals for and boundaries of technological development, in particular with regard to automated weapon systems. He was especially critical of the anthropomorphisation of technology that resulted in the term 'artificial intelligence'. In his seminal work *Computer power and human reason* (1976), Weizenbaum criticised the term AI because of its inherent analogy to human intelligence. Human intelligence, he argued, was categorically different to automated calculations: AI surpasses human intelligence in many aspects, but it lacks the intuitiveness, context-sensitivity, and reflectivity of human reasoning, among other deficits. While Weizenbaum was fascinated by the progress in AI research and, in fact, was one of its protagonists, he warned that the power of AI systems must not be overestimated and recognised a strong tendency in society to do so.

The work of Weizenbaum and other critical computer scientists (e.g. Coy & Bonsiepen, 1989; Marcus & Davis, 2019) thus intersect with a second stream of literature that we build upon. These contributions highlight the **complementarity** of artificial and human intelligence and the potential that AI can be used to augment human decision-making (Daugherty & Wilson, 2018; Davenport, 2018; Ramge, 2020). Decision-making from this perspective comprises various functions that involve data collection, analysis, and interpretation. Machines can be instrumental in supporting these functions and, indeed, open up unseen possibilities to provide the basis for more informed decisions. It remains the role of humans, however, to dynamically contextualise and interpret the respective material, which requires domain-specific and experiential knowledge that is often hard to codify.

These insights about the differences between artificial and human intelligence, as well as the possibilities for their interaction and the augmentation of human intelligence, are important cornerstones for the analysis of the use of genAI in knowledge work. From this brief discussion, we summarise the following aspects for our empirical analysis:

1. genAI should not be equated with human intelligence. Even if its results are fascinatingly similar to human communication, there is an epistemological abyss between AI and human reasoning. We, therefore, ask less about what human functions can be replicated and more about differences and possible complementarities.
2. How work changes through the use of genAI depends on the new ways by which human actors learn to use the tools at their disposal. Our research tackles questions of social innovation in the respective occupational fields and is particularly interested in the experimental mode by which actors identify new possibilities to fulfil tasks.

We are convinced that this perspective is particularly well-suited to analyse the specifics of genAI. While the above statements remain valid for many conventional automation technologies, the characteristic of genAI, and especially genAI chatbots, lies in the dialogical way in which their potential unfolds. We are not talking about a static set of machinery or algorithms that can be isolated and measured in its capabilities, but about a technology that evolves in interaction with human activity. For example, questions of skills, creativity, and authenticity are of particularly high relevance for the evolution of genAI, and the quality of its results most importantly lies in the prompts by which human actors operate it and the sensemaking needed to deal with its results.

3. Methods

Our main research question asks about the perception of genAI among employees in the field of knowledge work. In particular, we wanted to find out how they perceive (a) the impact on **skills** in terms of obsolete skills and newly emerging skills, (b) the relationship between genAI and **creativity**, and (c) the role and perception of **authenticity** with regard to the use of genAI. These terms were identified with a view to the theoretical discussion on the complementarity between human and artificial intelligence referred to above and public debates on the impact of genAI that particularly highlights issues of skills, creativity, and authenticity as particularly relevant.

In the context of this study, we limit our investigation to genAI systems that operate on the basis of large language models and provide the generation of written language. We selected three occupational fields in order to investigate these issues. These were supposed to represent relevant fields of knowledge work in which the application of genAI is imminent and in which we assume that the issues of (re)skilling, creativity, and authenticity matter. Based on the observation of public discourses on these matters in Germany and prior research experiences, we chose the fields of science, IT programming, and coaching for our study and used a mixed methods approach.

The **qualitative survey** consists of 14 expert interviews that we initiated in March of 2023, from which we derived insights that provided the basis for a broader investigation through a quantitative study that was conducted in December 2023. The qualitative study used inductive sampling following the snowball principle and aimed to provide detailed insights into the reasoning of employees in the surveyed professions about genAI. We also were able to reconstruct the manner and the current state of implementation in each field through qualitative data collection.

For the qualitative part of the analysis, we selected a sample of representatives of professional associations and other experts who possess insights about the general reception of genAI applications in their occupational field. We also contacted employees who used genAI in their work routine and thus were able to report about their experiences in working with genAI. We selected the interview partners according to the degree to which an organisation could be regarded as representing the perspectives of employees of a professional group and the ability of representatives to comment on the use of genAI in their field. In total, we conducted 14 semi-structured interviews that lasted between 60 and 90 minutes between May 2023 and March 2024. Questions were asked about their usage of genAI tools and perceived uptake of such tools in their occupational field, how such tool use changes work experiences and work outputs, and related queries.

The audio recordings of these conversations were transcribed and analysed using qualitative content analysis. As two coders, we worked iteratively and chose a deductive-inductive method of coding that departed from the three main categories 'skills', 'creativity', and 'authenticity' to which we added subcategories that were derived from the interview material.

The results were summarised and condensed to summaries for each main category in each sector and analysed comparatively.

The **quantitative online survey** (with different people) provides a broader perspective on the usage and perceptions of a greater number of employees and solidifies some of the assumptions that were developed in the qualitative analysis. The study also contained an **experiment** in a methods seminar at Humboldt University in Berlin that will constitute a third source of data. However, this data is currently being evaluated and has not been incorporated into this study.

To explore the perceptions of ChatGPT's impact across three distinct domains – coaching, IT, and science – we conducted three parallel yet separate studies. Participants were recruited through Prolific, specifically targeting individuals from the United States and Europe who regularly (and weekly) use AI technologies. To ensure relevance and expertise in each domain, we employed prescreeners tailored to each field: for IT, we required experience in computer programming, familiarity with various programming languages, and knowledge of software development techniques; for coaching, we sought individuals in roles such as consultant, coach, therapist, personal trainer, or well-being counsellor; and for science, we focused on those with a research function in their employment.

Across all studies, participants were compensated at a rate of £12 per hour. Initially, 385 individuals were recruited. However, participants who reported non-active use of ChatGPT were excluded from further analysis to maintain a focus on individuals with current and relevant experiences with it. This criterion led to a final participant pool distributed across the three domains as follows: 105 in IT, with 25.7% female, a mean age of 31.32 (SD = 10.35), and notable usage patterns of ChatGPT 3.5 (60%), and version 4 (40%); 120 in coaching, with 46.7% female, a mean age of 34.26 (SD = 10.22), and a preference for ChatGPT 3.5 (70%) over version 4 (30%); and 106 in science, with 54.7% female, a mean age of 30.15 (SD = 5.79), with 73% using ChatGPT 3.5 and 27% using version 4.

Participants came from diverse educational backgrounds, spanning all levels of education, which added a wide variety of perspectives to our analysis. The study aimed to assess their perceptions of the impact of ChatGPT on creativity, skills, and authenticity within their respective fields (compare Figures 1–3 for the precise items). This assessment was created for the purpose of this study and carried out using items rated on a 5-point Likert scale, the results of which are detailed in Figures 1–3, provided in the results section.

4. Findings: Skills, creativity, and authenticity

a) The case of science

The advent of ChatGPT and other genAI tools has generated much debate among researchers and research managers, editors and publishers, and funders. They discuss how the use of genAI might affect the quality of research and its evaluation, the publication and grant application system, and how it might introduce potential biases and inequalities in terms of research content, but also in terms of which research organisations are willing and able to pay for genAI tools. While people recognise the opportunities and possibilities for improving and speeding up the research process, the communication of results, and applications for further funding, people also fear a 'vicious circle' of genAI producing publications and grant proposals, which are then evaluated by genAI with fundamental effects on scientific knowledge production.

In the following, we discuss the results of our expert interviews focusing on (1) the (un)learning of old and new skills and expertise, (2) creativity as a central prerequisite of originality and innovation in scientific knowledge production, and (3) authenticity as the critical aspect for authorship in terms of taking responsibility for results, granting originality, and questions of intellectual property rights. These three topics present critical aspects of the ongoing technological developments and the debate on genAI in science.

aa) Skills

The debate on how genAI changes the social practices of doing science in terms of the (un)learning of skills and the gain or loss of expertise centres on two crucial topics: First, our interviewees discuss what skills are needed to properly use genAI as a tool to support research work. They address skills such as prompt engineering, the critical assessment of genAI results, and the need to understand how genAI works to know about its limitations (such as hallucinations). In the interviews, people also address the question of how to cope with the rapid development of new genAI tools. They highlight the need for constant adaptation to this development – either for using genAI or for regulating it – and indicate speed and permanent flexibility as other main aspects in terms of new skills for using genAI (see also Figure 1).

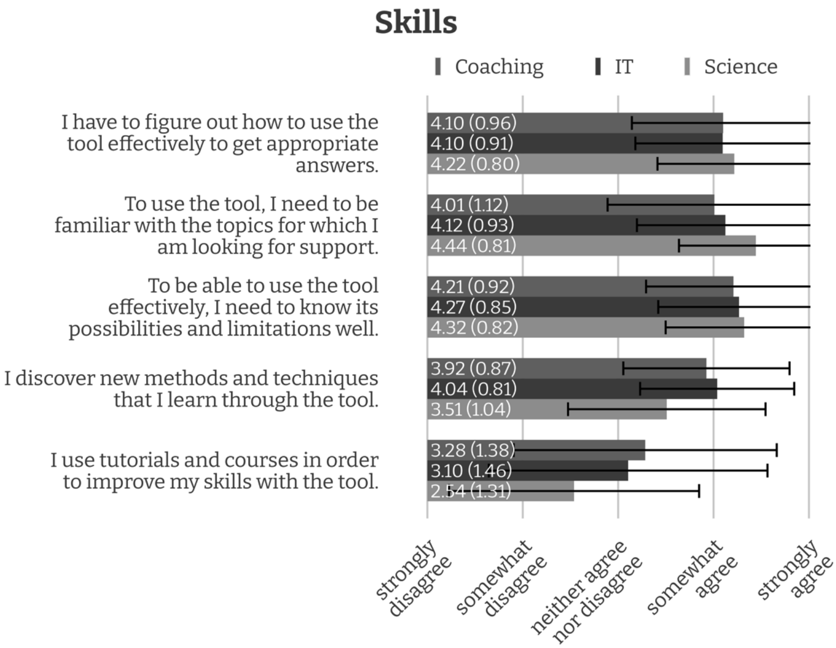

Figure 1. Comparative assessment of the perceived impact of ChatGPT on skill enhancement across coaching, information technology, and science domains

Taking all of this into account, they agree that genAI can help with analysing data in the research process and writing text. Nonetheless, the role of genAI in science is also regarded as differing between disciplines. While putting results into text might be considered a task which can be easily standardised and therefore automatised in the natural sciences, in the humanities and social sciences, the production of text is understood as an integral part of the research process and the produced text is considered to be the actual result. Some interviewees fear that the use of genAI as a tool to speed up the writing process could even lead to a dominance of disciplines more adaptive to integrating genAI over other disciplines, which then might fall behind, in particular, in producing grant proposals but also publications as the dominant currency in science to attain positive evaluations and funding.

Second, however, our interviewees also address the question of the limits of genAI and the limits of the *use* of genAI beyond the mere question of its capabilities. They ask what kind of scientific work can actually be automatised and which work should still be done by human experts. Some interviewees differentiate between standard tasks in text production and data analysis and tasks that require human sensemaking. The latter are regarded to be the sensemaking processes that guide the production of text and the analysis of data but also comprise the checking, weighing, and evaluation of results. Most of all, they argue that only humans can take responsibility for scientific results. They also highlight that it is necessary to grant the reproducibility and transparency of results, which limits the use of genAI to tasks where humans are able to understand on which grounds genAI is actually doing what it does. Yet, some people also question the use of genAI for standard tasks. They highlight instead that even if genAI can do standard tasks such as writing an abstract or doing a literature review, it is still necessary that scientists have the skills to do it in order to be able to judge the results of genAI.

This process leads to a new skill, which one interviewee addressed as being most urgently needed. It is described as the competence to 'manage' the use of genAI. This does not only include the question of who decides upon which software is bought and worked with at a research organisation but also who manages the possibilities for using genAI. As the capabilities of genAI and the possibilities of tasks for which to use genAI are rapidly increasing, this relates to questions of who decides upon rules for good scientific practice when working with genAI.

bb) Creativity

The interviews and the survey also address the question of how the notion of creativity in science might change through the use of genAI (see Figure 2). Understanding creativity as innovative and original and thus at the core of scientific practice, people discuss how far genAI can contribute to this creative process, particularly as genAI can already be used to develop research questions and review papers and proposals. A central insight from our interviews is that people recognise the capacity of genAI to discuss new ideas. However, they still regard the production of new research questions combined with its methodological operationalisation as unique to humans because of their capacity to meaningfully explain why something is innovative and original beyond statistical probabilities.

For the humanities and social sciences, the writing process is also considered a crucial element of the creative process that cannot be substituted by genAI. In the social and natural sciences, which apply quantitative methods, this is also discussed for the analysis of data. Beyond statistical calculations, which are done by using genAI – for instance, in pattern recognition software, data analysis is recognised as a creative process that needs original thought to raise innovative research questions, to operationalise them, and to discuss the results in an innovative way.

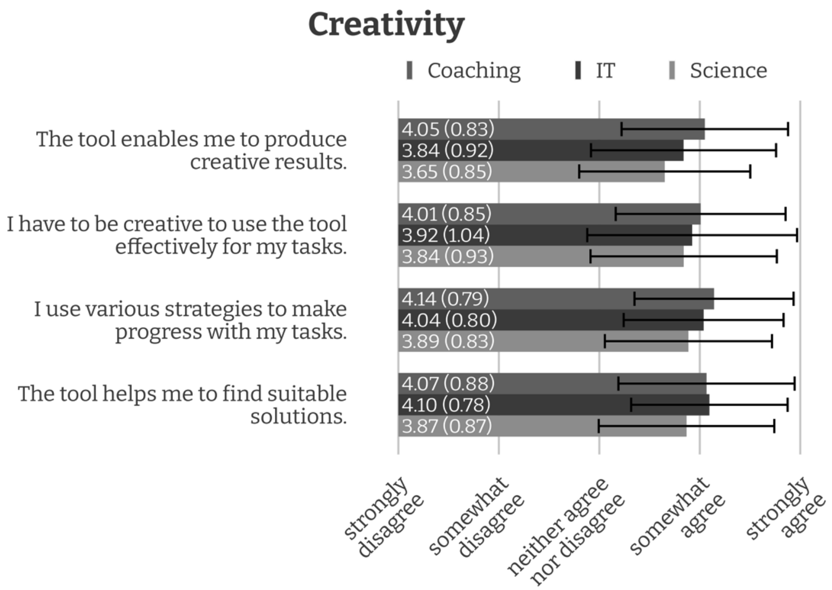

Figure 2. Comparative assessment of the perceived impact of ChatGPT on creativity across the coaching, information technology, and science domains

Creativity, however, is seen as crucial not only to the research process but also to reviewing and evaluating papers and grant proposals. Some of the interviewees argue that only humans are capable of judging scientific originality and novelty because this kind of judgement also draws on original thinking and the ability to critically reflect on what constitutes an original innovation beyond a mere recombination of existing ideas. Yet, creativity is also understood in terms of being creative when using genAI. Looking for the best prompts to achieve the best results is becoming recognised as part of the creativity needed to do good science.

cc) Authenticity

Authenticity, understood as a question of authorship, is a topic of central concern to all interviewees and was also part of the survey (see Figure 3). It is, on the one hand, discussed as a legal question in terms of good scientific practice in producing scientific results and evaluating research. Furthermore, people discuss whether genAI is a tool to assist people or if it can be regarded as a 'collaborator' with its own ideas. Central to this dilemma is the question of what happens to the reliability of scientific results as the crucial source for the legitimacy of science when genAI is used in an increasing number of ways during the research process. In any case, people agree that genAI cannot take responsibility for the results it produces, meaning it cannot attain the status of an author. Nonetheless, people also agree that the use of genAI should always be mentioned while being aware of the different degrees of involvement of genAI in the research process, from summarising existing literature to coming up with new research questions.

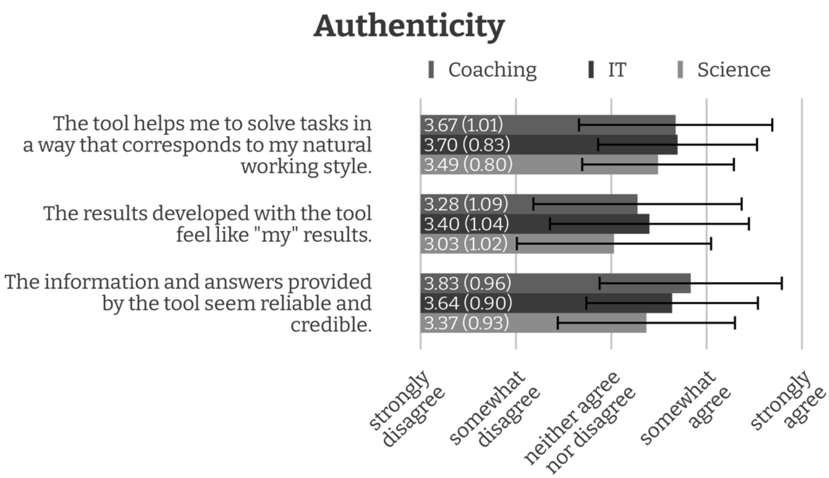

Figure 3. Comparative assessment of the perceived impact of ChatGPT on authenticity across the coaching, information technology, and science domains

Drawing on the aforementioned understanding of sensemaking and creativity as explicitly human capacities, the need for authenticity is understood as core to novelty and originality and thus to the reliability and legitimacy of scientific results, which must not be challenged by using genAI.

b) The case of coaching

In the evolving coaching landscape, the introduction of genAI (genAI) tools has sparked considerable discussion regarding their applicability and utility within the field. While there is an acknowledgement of the potential benefits that genAI can offer, the overall sentiment among coaching professionals reveals a cautious perspective. This apprehension stems from a clear recognition of the inherent limitations of current genAI technologies when applied to the core processes of coaching, which are deeply rooted in human interaction, empathy, and understanding. Coaching is a structured process to enhance an individual's personal or professional development through tailored guidance, support, and empowerment. It involves a collaborative relationship between the coach and coachee, wherein the coach employs various techniques to facilitate self-discovery, goal setting, and the achievement of desired outcomes. Rooted in principles of psychology, education, and human behaviour, coaching emphasises the coachee's active role in their growth journey, leveraging their strengths and potential to foster significant, sustainable change. The primarily informative nature of genAI, as exemplified by text-generation tools, contrasts with coaching's explorative approach, which focuses on eliciting personal insights through questioning. Thus, current genAI tools like ChatGPT are not suited for a direct coaching process. However, other potential areas are evident in which genAI tools can support the coaching practice, such as enhanced accessibility and efficiency in certain operational aspects of coaching.

aa) Skills

In the perception of the interviewed professionals, the integration of genAI in coaching is creating a divide between traditional coaching methods and technologically enhanced practices. Coaches recognise genAI's role in streamlining specific tasks, such as scheduling, initial assessments, and gathering basic information and common methods. Highly common and standardised coaching issues like time management are already outsourced into online courses enhanced by genAI. However, this seems only applicable to topics that are concise in scope and follow a somewhat structured approach. Furthermore, the exploration of genAI's potential in coaching has brought to light its utility in supplementary roles, such as collecting and analysing data during sessions. This application suggests a modest yet significant opportunity for genAI to contribute to the coaching process by offering additional information.

However, the interviews revealed a consensus on the irreplaceable nature of human qualities in coaching, highlighting that while genAI can augment certain aspects of coaching, emotional intelligence, empathy, and the capacity for deep human connection remain central to effective coaching practices. The capacity for emotional attunement, the development of trust, and the understanding of subtle contextual cues remain distinctly human attributes that genAI cannot replicate. Further, the survey results suggest that knowledge about specific topics is still necessary when working with ChatGPT, while the actual tool use is intuitive as it does not require any specific skills to use (Figure 1).

The skills associated with deploying genAI effectively in coaching – knowledge of the tools, understanding how they work, and efficiently leveraging them – are tempered by a critical awareness of the technology's boundaries. As the coaching profession continues to grapple with the implications of genAI, the prevailing view is one of cautious integration, where genAI's role is confined to augmentative and supportive functions, always secondary to the irreplaceable value of human insight and connection in the coaching process.

bb) Creativity

GenAI's capability to generate diverse ideas and perspectives can, at a superficial level, contribute to the creative toolkit available to coaches. For instance, AI-driven prompts can inspire novel approaches to standard coaching challenges or provide a range of options for tackling common issues like goal setting or motivation. Coaches appreciate genAI's ability to generate diverse ideas, aiding in brainstorming and problem-solving. The survey results enforce this notion, as coaches reported coming to creative and suitable solutions with ChatGPT (Figure 2) –whereby the bias of tool use should be taken into account here. Coaches who find the tool useful and meaningful continue to use it and thus could participate in our study. However, in the interviews, the depth and usefulness of these AI-generated contributions are often questioned. The nuanced understanding of a coachee's personal experiences, emotions, and the subtleties of human behaviour remain distinctly human facets that genAI cannot adequately capture or replicate. As such, the true essence of creativity in coaching, rooted in deep empathy and genuine human connection, appears beyond the reach of current genAI technologies. In fact, the survey results show that coaches need creativity to use ChatGPT effectively (Figure 2).

Moreover, the potential for genAI to support creativity in coaching raises significant considerations regarding the balance between technology and human intuition. While genAI can offer new insights, the critical judgement to discern which of these are truly valuable and applicable in a specific coaching context remains a distinctly human skill. The creative application of genAI in coaching, therefore, involves a discerning use of technology, where coaches must navigate between the benefits of AI-generated content and the irreplaceable value of human-driven creative insight.

cc) Authenticity

Authenticity in coaching is paramount, constituting the bedrock of the coach–coachee relationship. It involves genuine interactions, sincere emotional engagement, and mutual trust that enables profound personal growth. The use of genAI tools in coaching, while offering operational efficiencies and perceived viable answers (Figure 3), prompts a significant concern among the interview partners regarding the preservation of authenticity. It is held that these tools, by their very nature, operate on algorithms and patterns devoid of the capacity for genuine emotional intelligence or the nuanced understanding of human complexities. As such, while genAI can process and produce content at remarkable speeds, the depth of understanding and the empathetic connection that form the essence of an authentic coaching relationship cannot be algorithmically replicated. This limitation highlights the indispensable value of human presence in coaching, underscoring that the core of coaching – the authentic connection and understanding between coach and coachee – remains irreplaceably human.

Moreover, the challenge of maintaining authenticity extends to the ethical use of genAI in coaching. All interviewed coaches agree that transparency regarding the role and extent of genAI's involvement in the coaching process becomes crucial. Coaches are tasked with the responsibility of clearly communicating to coachees when and how genAI tools are used, ensuring that the coachees understand the nature of these tools and their limitations. This transparency is essential in maintaining the trust and integrity of the coaching relationship, allowing coachees to discern the origin of insights and advice – whether they are derived from their coach's expertise or augmented by genAI.

Amid the exploration of authenticity in the integration of genAI into coaching, an intriguing counterpoint emerges regarding the accessibility and comfort level some individuals may experience when using genAI tools. Specifically, anthropomorphised chatbots and AI systems present a unique potential for those who find it challenging to engage in the deeply personal and vulnerable process of coaching another human. For certain individuals, especially those grappling with issues surrounded by stigma or shame, initiating conversations and expressing their innermost thoughts and feelings can be significantly less daunting when interacting with an AI (see Waytz et al., 2014). The perceived lack of judgement and anonymity provided by genAI tools can lower barriers to entry for these individuals, offering them a preliminary step toward seeking help. In these instances, genAI does not detract from the authenticity of the coaching process but rather enables a form of engagement that might not have been possible otherwise.

Incorporating the various aspects into the broader discussion, it becomes evident that the role of genAI in coaching is multifaceted, offering both challenges and opportunities to preserve and enhance the coaching experience. The key lies in the mindful application of these technologies, recognising their potential to reach individuals who might otherwise remain outside the sphere of coaching support. As such, the integration of genAI into the coaching domain calls for a balanced approach, one that embraces the potential of technology to complement human-driven coaching processes while maintaining a steadfast commitment to the principles of authenticity and ethical practice. Nonetheless, the effective integration of these tools requires a deep understanding of their functions, limitations, and the ethical considerations involved. Coaches must discern the use of genAI and ensure that these technologies serve to support rather than overshadow the indispensable human elements of coaching.

c) The case of programming

The introduction of ChatGPT and other genAI tools raised great interest among programmers to test and utilise their potential and limitations for various tasks in a timely manner. The broad discussion about the work efficiency of the tools and the programmers' initial experience also quickly led to an increase in interest from some companies in financing the corresponding genAI tools for their IT departments. Data protection debates tended to take a back seat, or the topic was downplayed, and responsibility for its use was passed on to the individual programmer.

After all, the field of programming is constantly changing: Not only technical devices but also technical services and possibilities are constantly expanding. As a result, the demand for programmers is also increasing, and new jobs, such as data analyst and ML engineer, are becoming more and more in demand. At the same time, the requirements for every developer are constantly and gradually adapting in order to keep up with the latest trends and fulfil the quest for simplicity and efficiency. The most prominent example is programming languages, which themselves are constantly changing – not only in which context and by which type of developer they are used, but also to which further development they are subject. If they are used a lot, more and more functions and possibilities are added, while other languages lose their functionality. As one aspect of a programmer's job is to work with these languages in the programming process for about half of the workweek, this also raises the question of how the apparently necessary retraining and further training of skills in order to keep up is changing due to new tools such as ChatGPT.

aa) Skills

The discussion about the change in skills in the context of the use of genAI in programming encompasses a variety of aspects that were examined in the expert interviews. These include the identification of suitable areas of applicability of genAI, the skills required for its use, and the importance of human expertise in various contexts. In addition to the main coding, the tasks of programmers also include customer meetings, team meetings, code reviews, and the documentation of code and programs.

The potential use cases for genAI are primarily in the coding and creation of texts and documentation for code snippets and programs. In coding, genAI can support routine tasks and provide suggestions. However, it remains crucial that programmers have a sound understanding of the underlying processes in order to avoid potential errors and bugs, some of which only appear later (see also Figure 1). GenAI offers clear advantages, particularly in the area of text creation and documentation, which was previously neglected due to time constraints. The integration of genAI can, therefore, not only save time but also lead to better quality and traceability in, for example, code reviews.

However, the question of how the use of genAI changes forms of learning and knowledge is also raised. There is some concern that the experiential knowledge that is currently being acquired as a junior, for example, will be replaced by the use of genAI and that self-learning will fall by the wayside. Therefore, it seems essential that programmers continue to have the skills to recognise and identify problems and give instructions independently, as well as to acquire and test the knowledge required to use genAI tools effectively (see Figure 1).

bb) Creativity

The effective use of genAI in programming requires an in-depth analysis of the available tools. Programmers inform themselves in advance about the functions and possible uses of genAI tools through research and practical testing. This process of testing new tools is usually done through learning by doing, using simple tasks to evaluate the output and understand the potential of the tool. In addition, there is often a self-organised exchange between programmers to share experiences and best practices in dealing with genAI.

The discussion about creativity in programming is about finding innovative and original ways in the development process. The actual coding is only one part of the overall work, with the focus on the knowledge gained from experience and the ability to solve problems, also with the support of genAI. An interview partner points out: ‘Just that one doesn't only Google but also uses AI. So, the results are the same. The question is, just how quickly do you get to the result? Or how precise are the results?’

The use of genAI, such as ChatGPT, is only seen as an extension of the existing toolset. What is new is the wide range of possibilities that such tools now offer. For example, ChatGPT can be used as a research tool in specific cases or to simplify and visualise different solution approaches. This function provides effective added value if you have the relevant experiential knowledge to a) prompt in a targeted manner and b) be able to continue working with the output conscientiously. Our study results also underline the importance of human creativity and experiential knowledge, which are essential for critically evaluating the output of genAI and using it sensibly. It can be seen that human creativity plays a central role in the effective use of genAI in programming by enabling a meaningful integration of the tool into the development process (see Figure 2).

cc) Authenticity

Accordingly, the authenticity debates in the programming field remain rather quiet. Despite the fact that the latest tools and trends are usually tested in terms of their potential and possible use cases, questions about the origin of ideas are rarely discussed, even with the advent of freely accessible genAI. This lack of discussion may be due to the fact that genAI is not seen as more than just another tool, i.e. 'just the new Google' and seems to correspond more or less with their own natural working style (see Figure 3). However, when it comes to the question of the correct use of the tools and the responsibility for the product presented to the customer, one encounters recurring threads of debate.

While the occurrence of errors and bugs in code is nothing new, the use of genAI also leads to new sources of errors that the developers are not familiar with from their 'own code'. In particular, genAI-generated code that exceeds the developer's own experiential knowledge leads to an increased number of subsequent bugs, similar to untested genAI code. A comprehensive understanding of typical sources of errors, which often are more profound and only occur in a few cases, is often only the domain of senior developers. Junior developers would not even recognise them. Consequently, some senior programmers offer junior developers the opportunity to work together a) through their concepts, b) their prompting, and c) through the code created with genAI in order to increase junior developers' experiential knowledge. They hope that by sharing their experiences, their juniors will avoid future errors and subsequently reduce their own time-consuming rework. Otherwise, there is always a risk that sources of error will not be recognised due to a lack of knowledge at the junior level and a lack of time at the senior level, meaning the product will be delivered to customers with bugs. Nevertheless, this is also part of the business: it is always possible to optimise and expand a code, so before genAI, the products often had to be revised and adapted after a certain time anyway.

Overall, genAI tools are primarily described as support tools that may be able to take over a large part of the tasks of junior developers in the future. Nevertheless, even experienced developers find them to be a supportive tool rather than a collaborative partner with their own ideas. More central seems to be the question of how the qualification, learning, and knowledge processes of developers will evolve if the genAI is already a junior developer at their side.

5. Conclusion

Our study provides evidence for new requirements in terms of **skills**, **creativity**, and the ability to verify the **authenticity** and reliability of automatically generated inputs in a dialogical relationship between workers and genAI tools. This view is clearly expressed in the numerous statements by interview partners in all three professions on requirements for **learning** how to operate and master genAI tools, and it is also reflected in our quantitative study. Figure 1 displays a particularly high awareness of the need to acquire skills to do so, which so far is only partially backed up by education and learning opportunities in formalised courses. Acquiring skills on genAI to date mostly seems to be a matter of trial and error.

The awareness of participants in our study about requirements for creativity is equally high and present in all surveyed professions. Clearly, the use of genAI is not a zero-sum game in which human actors lose capabilities when genAI systems grow stronger and more sophisticated. Accordingly, the attitude of most interview partners could be described as a concerned embrace of the new technical options. There are some concerns about a hollowing out of professions and especially about a deterioration of quality, e.g. in peer review processes in science and in services provided by semi-automated coaching platforms. However, there is clearly an acknowledgement of the new possibilities and benefits that could be derived from them.

A fascinating result of the qualitative study that we did not expect is the amount of reflection about the **professional core** of the occupations that we encountered in each field. Rather than weakening the professions, genAI triggered an intensive process of reflection and self-assurance about what it means to be a scientist, coach, or programmer. The participants of this study, by and large, agree that the meaning of these professions transcends providing single tasks that could be replicated or substituted by AI. There is a growing awareness of the skills to combine, contextualise, and dynamically adjust work tasks that clearly lie beyond the boundaries of the punctual responses and calculations of large language models. There are also indications that the professions **dynamically adjust** to new technological (and societal) realities by enlarging and updating the skills and competencies in each profession. This is most clearly expressed in the case of programming, where dynamic adjustment to new programming languages and software developments has constituted a fundamental characteristic of the profession for many years.

Among the confident statements in favour of the complementarity of humans and genAI, we could also detect **uncertainties** about the boundaries of what AI could and should do in the future. When interview partners expressed convictions about what AI would never be able to do, we also detected underlying fears behind such strong statements that sometimes also remained contradictory. The self-assurance about the core content and values in each profession is accompanied by a desire to delimit the reach of AI. This self-clarification process is characterised by tensions that revolve around questions of 'authenticity' – questions that not only concern what AI could do but also how society should effectively control and regulate the usage of constantly evolving AI.

References

Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, *108*(6), Article 6. https://doi.org/10.1257/aer.20160696

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work* (Working Paper 31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Brynjolfsson, E., & McAfee, A. (2012). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Digital Frontier Press.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. W. W. Norton & Company.

Coy, W., & Bonsiepen, L. (1989). *Erfahrung und Berechnung* (Bd. 229). Springer. https://doi.org/10.1007/978-3-642-75217-9

Daugherty, P. R., & Wilson, H. J. (2018). *Human + machine: Reimagining work in the age of AI*. Harvard Business Review Press.

Davenport, T. H. (2018). *The AI advantage: How to put the artificial intelligence revolution to work*. MIT Press. https://mitpress.mit.edu/9780262538008/the-ai-advantage/

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality*. SSRN Scholarly Paper 4573321. https://doi.org/10.2139/ssrn.4573321

Felten, E. W., Raj, M., & Seamans, R. (2023). *How will language modelers like ChatGPT affect occupations and industries?* SSRN Scholarly Paper 4375268. https://doi.org/10.2139/ssrn.4375268

Hatzius, J., Briggs, J., Kodnani, D., & Pierdomenico, G. (2023). *The potentially large effects of artificial intelligence on economic growth*. Goldman Sachs.

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon.

Noy, S., & Zhang, W. (2023). *Experimental evidence on the productivity effects of generative artificial intelligence*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375283

Ramge, T. (2020). Augmented intelligence. Wie wir mit Daten und KI besser entscheiden: Reclam, Philipp, jun. GmbH, Verlag.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Weizenbaum, J. (1976). *Computer power and human reason: From judgement to calculation* (new edition). W. H. Freeman & Co Ltd.

Sustainability: Energy-efficient intelligent bicycle sensors to promote mobility transition

Paula Scharf, Benjamin Karic, Luca Hesse, and Thomas Bartoschek

Abstract: Cycling plays a key role in pushing towards more environmentally friendly forms of mobility and ensuring the safety of cyclists, which is crucial to motivate more people to choose the bike as their preferred mode of transportation. The still car-centred infrastructure calls for innovative methods to ensure a safe environment for cyclists. In Germany, despite the legal requirement of a minimum distance of 1.5 metres when cars overtake bicycles, in practice, compliance is often inadequate. In this work, we propose and prototype a mobile, low-cost, and low-energy system that allows cyclists to monitor the occurrence of dangerously close takeovers. This system employs a time-of-flight sensor that measures distances in an 8x8 pixel matrix and uses tiny machine-learning models to detect dangerous takeovers. When included in citizen science campaigns, the collected data can be used for urban mobility planning and safety strategies.

Keywords: sensors, citizen science, machine learning, urban mobility, sustainability

1. Introduction

The transformation of transportation is a major societal challenge in achieving climate neutrality. Technological innovations like improved renewable energy production and battery technologies are critical steps toward this goal. But just as important is reducing the use of private fuel-powered vehicles and increasing the use of more climate-friendly modes such as bicycles, pedelecs, and public transportation whenever possible to promote green transportation.

While almost half of all Germans would prefer to ride their bicycles more often (Sinus Marktund Sozialforschung GmbH, 2023), cars are still predominantly used – even for short commutes under 5km (Statistisches Bundesamt (Destatis), 2022). The low usage of bicycles, even for short distances, can be attributed to the lack of proper infrastructure: 40% of cyclists state that they do not feel safe while cycling due to excessive traffic, reckless driving by motor vehicles, and a lack of designated bike lanes (Sinus Marktund Sozialforschung GmbH, 2023). Cars performing takeover manoeuvres on streets without separated bike lanes pose a typical safety threat to cyclists. Despite the German Road Traffic Regulations (StVO) stipulating a minimum passing distance of 1.5 metres between cars and cyclists in urban areas (§ 5 Abs. 4 S. 3 StVO), compliance with this requirement is often lacking in practice. Such situations not only pose an immediate safety risk but also have a significant impact on people's willingness to use bicycles as an everyday means of transportation. In fact, in 2022, a comprehensive study found that 42% of German respondents would choose a car over a bicycle, even for short journeys, due to the unsafe cycling infrastructure (Pyhel, 2022).

To improve cycling infrastructure, it is crucial to identify high-risk areas. Subsequently, implementing safety measures and promoting cycling as a safe and sustainable mode of transportation can help overcome the negative safety perceptions associated with cycling. However, identifying high-risk locations in Germany's extensive transportation infrastructure presents a significant challenge to municipalities. The task at hand presents a two-fold challenge: 1) detecting high-risk locations in cycling infrastructure and 2) ensuring the scalability of the approach. To address this challenge, we propose a concept using a citizen science approach to identify safety gaps in urban cycling infrastructure that result in dangerous overtaking manoeuvres.

We present a prototype based on the proposed concept that combines cutting-edge time-of-flight sensor technology, neural networks, and a senseBox (Wirwahn et al., 2015) to automatically identify high-risk locations. The method uses low-cost and energy-efficient technologies that can be implemented in a citizen science project to generate large-scale data on the safety of cycling infrastructure while engaging cyclists in improving cycling safety. The objective is to establish a sustainable data collection method to support proactive accident prevention, enhance road safety, and improve urban cycling infrastructure. Research in this area is highly relevant as it positively impacts the sustainability of urban mobility planning and safety strategies.

Existing efforts in this field often rely on manual, qualitative incident reporting. Prominent examples of this approach include bikemaps.org (Nelson et al., 2015) and the Bike Barometer (Storme et al., 2022). However, data collection is often exhaustive and sporadic for such solutions, leading to a lack of broad spatial and temporal coverage.

CycleAI[1] has taken a different approach to bicycle road safety. They have trained an AI algorithm on how 'bikeable' a road is based on static images (from Google Street View) of that stretch of road. Based on these ratings, a routing algorithm suggests safer routes. This algorithm considers only the infrastructure depicted in the images and does not take into account additional environmental data or a temporal component. Despite achieving better spatial coverage, it does not consider varying safety at different times of the day or with changing weather conditions.

The OpenBikeSensor project[2] offers a semi-automated solution for recording violations of the minimum takeover distance. They use an ultrasonic proximity sensor to measure the distance to passing cars when the cyclist presses a button (e.g. Hauenstein et al., 2023). However, it can be challenging to trigger the device reliably in hazardous traffic situations, which may result in forgetting to capture data or recording inaccurate data.

The projects 'Sicher überholt!' (Blees, 2021) and 'Radmesser' (Lehmann et al., 2018) propose a fully automatic system that uses a combination of a smartphone camera and an ultrasonic proximity sensor. The smartphone automatically captures images to confirm a car takeover triggered by ultrasonic sensor measurements. Nonetheless, the use of cameras raises privacy concerns and increases energy consumption.

The senseBox:bike[3] project involves using a senseBox to measure distances to the left side of a bicycle with an ultrasonic proximity sensor. However, the primary purpose of the project is to serve as a mobile environmental monitoring station for bicycles. Although distance values are captured, they are not processed to further determine if a car has actually overtaken the bicycle.

2. Concept

In this work, we propose a mobile, low-cost, and low-energy takeover detection system as schematised in Figure 1. At the heart of our system is a microcontroller unit (MCU) – a compact unit of processor, memory,

1 https://cycleai.net/ (last visited 22.04.2024)
2 https://www.openbikesensor.org/ (last visited 22.04.2024)
3 https://sensebox.de/de/products-bike (last visited 22.04.2024)

and periphery, combining all essential parts of a computer on a small scale with low requisites in terms of cost and power. Specifically, an MCU of the senseBox family is chosen. The senseBox is a versatile electronics kit designed specifically for citizen science projects and educational initiatives, focusing on environmental monitoring and data collection (Witte et al., 2023). Originally backed by the German Federal Ministry of Education and Research, it was designed to promote understanding of environmental science and technological applications. In addition to providing hardware and software components, the senseBox enables users to gather, analyse, and disseminate their own environmental data. Its anchoring in open-source technology and compatibility with the popular Arduino platform[4] for programming and prototyping electronics make it easily accessible and adaptable for a wide range of users and use cases (Bartoschek et al., 2019).

4 https://www.arduino.cc/en/Guide/Introduction (last visited 25.04.2024)

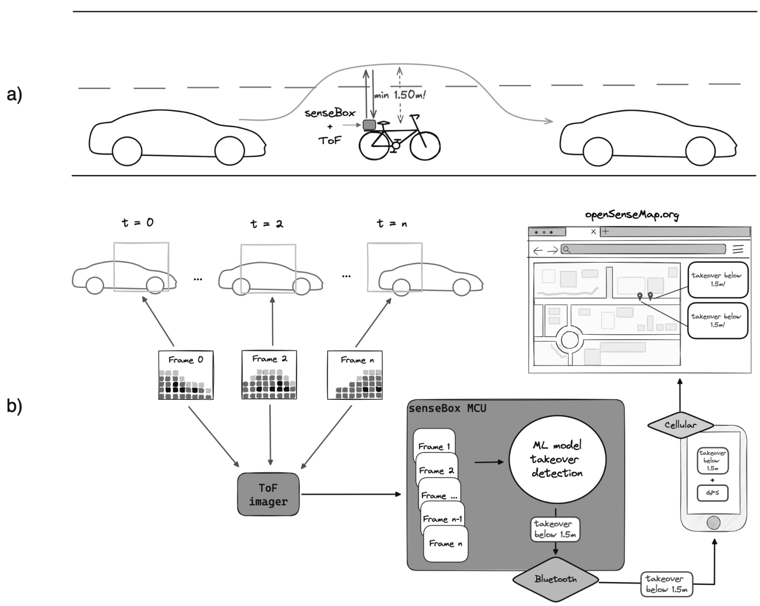

Figure 1. A conceptual model of the takeover detection system. a) A car taking over a bicycle monitored by an automatic detection system; b) Architecture and flow of the detection system. A time series of depth images captured by a Time-of-Flight (ToF) imager is classified on a senseBox MCU using machine learning. Results are transmitted via Bluetooth to a mobile application for geo-location and mobile data transfer to the open data platform openSenseMap (Pfeil et al., 2015).

A proximity sensor is utilised to sense passing traffic, particularly one that can capture very low-resolution depth images. Depth images, as opposed to single distance readings, enable the system to perceive not only the presence of objects but also their approximate shape and dimensions. In contrast to cameras, they offer a high level of privacy (no images of faces or licence plates can be acquired) and consume less power. Especially in light of the reduced processing power of MCUs, it is sensible to keep the amount of captured information low, including the image resolution. Additionally, since determining the distance of a passing car is crucial for assessing the safety of a manoeuvre, depth images are preferred over standard colour images. In this context, the VL53L8CX multizone time-of-flight (ToF) sensor[5] poses a perfect fit for our use case. ToF sensors capture distances/depths by emitting infrared light and measuring the time it takes for the light to be reflected. Compared to other options for recording depth images, the VL53L8CX sensor has particularly low requisites in terms of size, cost, and energy consumption. More detailed specifications of the sensor will be given in the following section.

The low-resolution depth images captured by the ToF sensor are well-suited for machine learning applications, being decently complex and often noisy. Furthermore, the usage of machine learning algorithms allows for flexibility in problem definition, for example, by detecting the type of the passing vehicle (car, truck, bicycle) and its speed and distance all in a single algorithm. To detect takeover manoeuvres, sequences of recorded depth images and their temporal order need to be considered. Recurrent neural networks (RNNs) offer a suitable machine learning methodology for this combination of classification problems and available data. They allow current input data (e.g. a single depth image) to be put in the context of previously recorded input data. This ability enables the analysis of the entire movement process of an object, such as a car, over time, which is crucial for distinguishing takeovers from other situations, such as oncoming traffic.

To put the detected takeovers into a geographic context and upload the resulting data, we adopt existing solutions built within the senseBox:bike project. Previous iterations of the senseBox:bike relied on a dedicated GPS module for georeferencing and utilised WiFi for data upload whenever a signal from preconfigured networks was available, typically the cyclist's home network. While functional, newer versions of the senseBox:bike improve this workflow by utilising a smartphone app through Bluetooth connection, in which both the position is determined and data is uploaded to the open data platform openSenseMap (Pfeil et al., 2015). The app streamlines the process, minimises code and configuration requirements, reduces the risk of data loss, and allows cyclists to begin recording immediately rather than waiting for the dedicated GPS module to establish a signal. The data uploaded to the openSenseMap is publicly available and can be explored through a web application or accessed for further analysis.

5 https://www.st.com/en/imaging-and-photonics-solutions/vl53l8cx.html (last visited 22.04.2024)

3. Prototype

As a demonstration of our proposed concept for a mobile, low-cost, and low-energy takeover detection system, we built a prototype, which entailed the assembly of hardware components and the development and deployment of its software. The latest iteration of the senseBox features an MCU based on an ESP32S2,[6] which boasts a single-core processor with 4 MB of flash memory and 320 KB of RAM, which is enough for small machine-learning applications.

To perceive passing traffic, the VL53L8CX multizone ToF sensor is chosen. The 'multizone' attribute indicates that the sensor measures distance not just at a single point but across 8x8 zones. Under perfect conditions without any ambient light whatsoever, the VL53L8CX can measure distances up to 400 cm. However, in regular ambient daylight, the maximum measurable distance is reduced to approximately 200 cm, which is just enough to detect when a car does not comply with the 150 cm minimum distance. Figure 2 provides illustrations of what the sensor detects compared to a camera capturing visible light.

a)

b)

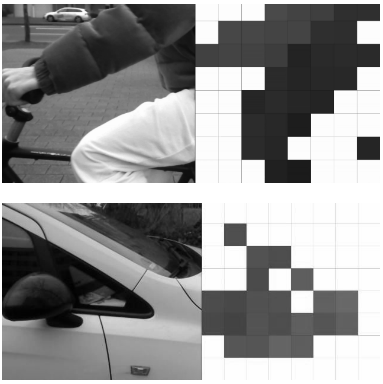

Figure 2. Two examples of what a normal webcam sees (top) and what the VL53L8CX sees (bottom). One of a cyclist (a) and one of a car (b), both at close distance

6 https://www.espressif.com/sites/default/files/documentation/esp32-s2_datasheet_en.pdf (last visited 22.04.2024)

To mount the sensor and the microcontroller on a bicycle, we 3D printed a simple case that can be attached to the carrier of a bicycle (see Figure 3). Alternatively, the case could also be attached underneath the saddle of the bicycle if it does not have a carrier.

Figure 3. Sensor (in black) in a 3D printed case together with the senseBox MCU S2 on a bicycle. The sensor is oriented to the left side of the bicycle.

As this prototype serves as a proof-of-concept rather than a finalised production system, the classification problem to be solved is deliberately simplified: sequences of distance recordings shall be either classified as takeover or no takeover, without any regard for distance, speed, or type of vehicle. To solve this problem, a simple RNN is trained and deployed.

Given the straightforward nature of the chosen classification problem, we maintain a simplistic structure for the neural network employed. The network comprises a sequence of only two layers: one consisting of long short-term memory (LSTM) neurons and the other of dense neurons. LSTM neurons are a type of recurrent neural network unit capable of capturing long-range dependencies in sequential data and retaining information over extended time periods. The network is configured to take a series of 20 frames as input for each classification task. Given that the sensor records data at a rate of 15 Hz, this equates to approximately 1.25 seconds of data, in which a takeover manoeuvre should fully take place.

For the purpose of training the proposed neural network, it was implemented in Python using TensorFlow,[7] a widely used open-source machine learning framework developed by Google. Given the high compatibility and support of the TensorFlow framework, the resulting trained model can be directly ported to and run on an MCU in the Arduino programming language. The model is first converted from the original TensorFlow version to a smaller and more efficient TensorFlow Lite version. This is then converted to a byte array and reconstructed in Arduino with TensorFlow Lite for Microcontrollers.

The network undergoes supervised training, relying on manually classified training data to learn effectively. As a matter of fact, a significant portion of the prototype's development effort was dedicated to recording and formatting suitable training data. We decided on a realistic approach for recording data by mounting the ToF sensor on a bicycle and driving around the streets of Münster. Data collection was impeded by the already well-developed bicycle infrastructure in this city, leaving only a few suitable streets with flowing traffic but without dedicated bike lanes. Data was recorded in those places by driving up and down the street during peak rush hour. To simplify data collection and not further endanger the cyclist recording the data, it was chosen to augment data by driving by parked cars and reversing the recorded video. Additionally, sequences in which other cyclists passed at short range were also considered valid takeovers.

A normal webcam was also mounted on the bicycle and configured to record simultaneously with the ToF. Using timestamps, the webcam videos were later used to label the sequences of depth frames from the ToF sensor as takeover or no takeover. The resulting labelled data was then divided into subsets for training, validating, and testing the neural network.

The data and the code for training and deployment are openly accessible on GitHub, along with more detailed descriptions and explanations of the procedure: https://github.com/TinyAIoT/tof-takeover-detection.

7 https://www.tensorflow.org/ (last visited 22.04.2024)

4. Applicability

The ultimately trained model performs incredibly well on the recorded data, with an accuracy of 99% or 95.5% for the model converted to TensorFlow Lite. However, considering that the recorded no-takeover sequences lack variety (they are mainly composed of empty frames), this is not immediately translatable to positive performance in the real world. The trained model struggles, especially by classifying many false positives, meaning that it detects takeovers when there are none. Nonetheless, the results are valuable as a proof-of-concept of the proposed method.

The size of the model comes out at a little over 15 KB, while memory reservations of a little over 8 KB are necessary to store the data going through the model. Meanwhile, its inference time on the senseBox based on ESP32S2 reaches 26 microseconds on average. Due to this incredibly low inference time, the model can easily be applied to each new sequence of frames in real time. The VL53L8CX ToF sensor can, at maximum, be configured to 15Hz, meaning it records a new frame every 66.7 milliseconds. These frames are continuously recorded in a ring buffer, and the most recent 20 frames are fed to the model whenever a new frame comes in.

Regarding the overall applicability of the VL53L8CX sensor, this study offers a very promising outlook on its performance. While manually reviewing the recordings that were made with the sensor attached to a bike, discernible differences among passing vehicles – whether cars, bicycles, scooters, or pedestrians – become immediately evident. These differences in dimensions and textures, clearly perceptible to the human eye, suggest the feasibility of similar distinctions by neural networks. Consequently, the cost-effective time-of-flight technology has the potential to contribute to traffic technologies. Although the lighting conditions impaired the quality to a certain extent, the captured images still resulted in clearly recognisable structures of vehicles and other objects. However, a notable issue persists in accurately capturing dark or highly reflective objects. Dark-coloured vehicles usually only yield patchy images containing mostly noise, a recognised limitation of ToF sensors (Baek et al., 2020; Zieliński, 2021).

In terms of future work, the available memory and RAM on the senseBox MCU leave room to increase the model size and complexity. It could, for example, be interesting to add convolutional layers to the network. Measurements of the accelerometer onboard the senseBox MCU could also aid the model performance and enable the determination of the speed of both the cyclist and passing vehicle at the moment of takeover. Priority for improving the model performance should, however, be the enhancement of the training data. Primarily, efforts should focus on diversifying the dataset by incorporating a wider array of non-takeover sequences to mitigate the prevalent issue of false positives. Additionally, existing data can be enriched through artificial augmentation techniques, such as introducing random noise, translating or shifting sequences, or adjusting the speed of recordings. Such augmentation strategies generally contribute to the robustness and generalisation capabilities of a model.

The georeferencing, uploading, and provision of collected data was not implemented in the prototype but has already been deployed in the senseBox:bike project, as previously mentioned. The existing workflow can be largely mirrored for our use case, with the exception that the interface of the smartphone app would have to be adjusted in line with the uploaded data. Depending on the detection solution adopted, uploaded data could, for example, include the distance, time of incident, type of vehicle, and also the current speed of the cyclist and passing vehicle, in case readings from the onboard accelerometer are added. To facilitate further analysis and provide additional visualisation tools, it would also be interesting to develop our own web interface tailored to the specific data instead of using the openSenseMap.

5. Conclusion

In this work, we present a concept and an initial prototype demonstrating the detection of dangerous overtaking manoeuvres using low-cost ToF sensor technology combined with machine learning methods. By employing ToF technology instead of visible light cameras and running the classification on-device, we effectively reduce the size of collected data and communication, thus saving energy. The two-fold significance of sustainability in this work stems from the fact that the proposed system is motivated to support a mobility transition towards more sustainable forms of transportation overall but also intrinsically aims for sustainability through the energy and resource-efficient, on-device processing of data.

The trained model showcases its capability to generate valuable classifications of takeover manoeuvres. It proves the feasibility of the concept while demonstrating the necessity for further collection of training data to offer a more robust solution. An unsolved challenge persists in detecting black or highly reflective cars with the chosen proximity sensor. Previous projects with the senseBox have already paved the way for georeferencing and uploading data, which have not yet been implemented in this described prototype. When combined with location information, the data can show spatial and temporal accumulations of dangerous overtaking manoeuvres, and consideration can then be given to adjusting the local cycling infrastructure. Ultimately, the result of this work is not intended to represent a final system but rather to serve as an essential preliminary step for further research and development with the knowledge gained and the software developed.

Looking ahead, our work holds promise for applications in the field of traffic and transport planning. For instance, the detected and localised dangerous overtaking manoeuvres allow for the identification of high-risk areas on roads, which can then be addressed by local authorities through targeted infrastructure improvements such as adding physical barriers between bicycles and cars on shared road segments. By analysing broad datasets collected by citizens, traffic planners can gain valuable insights into traffic patterns and the frequency of dangerous manoeuvres, allowing for data-driven decisions to enhance road safety. This participatory approach not only increases the amount of data available for analysis but also fosters community engagement in traffic safety initiatives.

Further use of devices applying our approach can significantly impact the field of education. By employing the device in workshops and initiatives in schools, students can learn about machine learning and sensor technology through hands-on experience. Schools can participate in data acquisition and analysis projects, enabling students to contribute to real-world datasets as part of citizen science initiatives. This involvement promotes spatial citizenship, as described by Gryl (2012), by empowering students to understand and influence their local environment. This educational approach not only enriches the learning experience but also instils a sense of civic responsibility and technological literacy in young learners.

Acknowledgements

This research has been partly funded by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection through the funding programme KI-Leuchttürme in the project 'TinyAIoT Energie und ressourceneffiziente künstliche Intelligenz für moderne IoT-Anwendungen' and partly by the Federal Ministry for Education and Research through the funding programme EUREKA-Globalstars-Kooperation

mit Brasilien in the project ‘atrAI Bikes Sichere Fahrradstädte durch Datenwissenschaft und intelligente Fahrräder’.

References

Baek, E.-T., Yang, H.-J., Kim, S.-H., Lee, G., & Jeong, H. (2020). Distance error correction in time-of-flight cameras using asynchronous integration time. *Sensors*, *20*(4), 1156–1156. https://doi.org/10.3390/s20041156

Bartoschek, T., Fehrenbach D., Fehrenbach T., Pesch M., & Steinmann, L. (2019). *Das senseBox-Buch*. dpunkt.verlag

Blees, V. (2021) Sicher Überholt. Radverkehrsinfrastruktur – Baustein Der Verkehrswende. Gemeinsame Abschlusspublikation Des NRVP-Forschungsbegleitkreises ‘Förderschwerpunkt Infrastruktur’, 24–31.

Lehmann, H., Meidinger, D., Wittlich, H., & Gegg, M. (2018, August 20). Fahrradverkehr in Berlin: Das Projekt Radmesser. *Der Tagesspiegel Online*. https://www.tagesspiegel.de/gesellschaft/medien/das-projekt-radmesser-3980056.html.

Statistisches Bundesamt (Destatis). (2022, January 31). Erwerbstätige nach Stellung im Beruf, Entfernung, Zeitaufwand und benutztem Verkehrsmittel für den Hinweg zur Arbeitsstätte 2020 in %. https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Tabellen/pendler1.html.

Gryl, I., & Jekel, T. (2012). Re-centering geoinformation in secondary education: Toward a spatial citizenship approach. *Cartographica*, *47*(1), 18–28. https://doi.org/10.3138/carto.47.1.18.

Hauenstein, J., Eckart, J., Zeile, P., & Merk, J. (2023). The effect of overtaking distances on the stress occurrence of cyclists in urban areas. *LET IT GROW, LET US PLAN, LET IT GROW. Nature-based Solutions for Sustainable Resilient Smart Green and Blue Cities. Proceedings of REAL CORP 2023, 28th International Conference on Urban Development, Regional Planning and Information Society* (pp. 699–708). https://doi.org/10.48494/REALCORP2023.9045.

Jurczok, F., Gensheimer, T., & Specht, F. (2023). *Fahrrad-Monitor Deutschland 2023*. Sinus Marktund Sozialforschung GmbH. https://bmdv.bund.de/SharedDocs/DE/Anlage/StV/fahrradmonitor-langfassung.pdf?__blob=publicationFile.

Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., & Winters, M. (2015). BikeMaps.Org: A global tool for collision and near miss mapping. *Frontiers in Public Health*, 3. https://doi.org/10.3389/fpubh.2015.00053.

Nobis, C. (2019). *Mobilität in Deutschland – MiD Analysen Zum Radverkehr Und Fußverkehr*.infas, DLR, IVT and infas 360 on behalf of the German Ministry of Transport. https://www.mobilitaet-in-deutschland.de/archive/pdf/MiD2017_Analyse_zum_Rad_und_Fussverkehr.pdf

Pfeil, M., Bartoschek, T., & Wirwahn, J. A. (2015). OPENSENSEMAP: A citizen science platform for publishing and exploring sensor data as open data. *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*, *15*(39). https://doi.org/10.7275/R56971SW.

Pyhel, J. (2022). *Studie Zum Weltfahrradtag: Sicherheitsbedenken Halten Menschen Vom Radfahren Ab*.Ipsos GmbH. https://www.ipsos.com/sites/default/files/ct/news/documents/2022-05/Ipsos-PI_Cycling_2022-05-24.pdf.

Storme, T., Benoit, S., Van De Weghe, N., Mertens, L., Van Dyck, D., Brondeel, R., Witlox, F., Zwartjes, L., & Cardon, G. (YEAR). Citizen science and the potential for mobility policy – Introducing the bike barometer. *Case Studies on Transport Policy*, *10*(3), 1539–1549. https://doi.org/10.1016/j.cstp.2022.05.013.

Wirwahn, J. A., & Bartoschek, T. (2015). Usability Engineering For Successful Open Citizen Science. *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*, *15*(1),. https://doi.org/10.7275/R54J0C9W.

Witte, V., Schwering, A., Bartoschek, T., & Pesch, M. (2023). Zukunftsweisender MINT-Unterricht Mit Dem senseBox-Ökosystem Die Plattform Für Partizipative Data Science Mit Physical Computing. *MNU Journal / Verband Zur Förderung Des MINT-Unterrichts*, *76*(4), 296–301.

Zieliński, B. (2020). A comparison of proximity sensors for a bicycle-to-car distance rangefinder. *International Journal of Electronics and Telecommunications*, 277–282. https://doi.org/10.24425/ijet.2021.135976.

Green coding – bridging the gap from theory to practice

Verena Majuntke, Lena Hoffmann, Felix Kronlage-Dammers, Grischa Beier, and Nikolas Becker

Abstract: This article summarises key points from the interdisciplinary panel discussions held at the Weizenbaum Conference 2024 titled "Green Coding – Bridging the Gap from Theory to Practice", where experts shared diverse perspectives on the importance of Green Coding (GC) in sustainable software development. As digital technologies and services expand, their ecological costs rise, with data centres consuming 1.0–1.5% of global electricity in 2022 and information and communication technologies (ICT) contributing about 1.5–4% of global greenhouse gas (GHG) emissions. The panel highlighted the role of software in influencing the energy and resource consumption of hardware. GC focuses on enhancing energy and material efficiency across the software lifecycle, addressing the growing demands of ICT driven by cloud computing, artificial intelligence and video streaming. Despite some companies recognising the importance of GC, implementation remains low and is often not part of established practices in software development. The discussion emphasised the need for open-source initiatives, standardised metrics, industry integration, stakeholder awareness and robust governance to align digital transformation with sustainability goals. This article highlights the main arguments and insights from the panel discussions.

Keywords: sustainable software, energy-efficient software, green coding, software-induced resource consumption, open source

1. Green coding – bridging the gap from theory to practice

The transition towards sustainable development requires a critical examination of the environmental emissions caused by digital technologies and services, as well as counteracting strategies to mitigate these impacts. **Green Coding (GC)** has emerged as a vital approach to enhancing the energy efficiency of software. Based on an interdisciplinary panel discussion at the Weizenbaum Conference 2024, this article elaborates on the current state of GC, the challenges in its measurement and the role of open-source methodologies in promoting sustainable software development.

2. Why is GC important for the transformation towards sustainability development?

Digital technologies and services are a vital part of increasingly numerous areas of social life; however, digital innovations also come with increasing ecological costs. In 2022, the operation of data centres already accounted for approximately 1.0–1.5% of global electricity demand (IEA, 2023). However, estimates regarding global energy consumption by information and communication technologies (ICT) vary greatly, particularly due to differing definitions of ICT (Gelenbe, 2023). Studies such as Hilty et al. (2009) argue that there is no clear definition of the scope of ICT; for example, they consider three building blocks of ICT: servers, end-user devices and the networking infrastructure typically used to communicate either between two or more end-user devices or among end-user devices and servers. As a result of differing underlying assumptions, estimates regarding ICT's share of global greenhouse gas (GHG) emissions also vary greatly, with systematic reviews estimating emissions at 1.8%–2.8% (Freitag et al., 2021) and 1.5%–4% (Bieser et al., 2023). One major part of ICT-related emissions is determined in the production phase of ICT hardware. For end-user devices, these emissions often exceed 50% of the device's overall life cycle GHG emissions, while for data centres, it is estimated that most of the life cycle-related GHG impacts are caused during the use phase (Bieser et al., 2023).

In contrast to hardware, concepts aimed towards making software usage more resource efficient have historically seen lower prioritisation. However, software is the core of ICT, and ICT is continuously growing and evolving to make software more efficiently available for users, i.e. industry and private consumers. While technologies become more efficient, the gained efficiency tends to be utilised to transmit greater volumes of data rather than making the transmission of existing data more resource-efficient, a phenomenon known as the **Jevons Paradox**. It is necessary to focus more extensively on the energy and material efficiency of data infrastructure and data centres while also looking to **make software operations as resource-efficient as possible**. Combining these approaches is an important factor and a crucial prerequisite if we want to be successful in achieving a more general transition towards sustainable development.

GC is one approach aimed at increasing the energy efficiency of software usage. GC is defined as "**the act of designing, developing, maintaining, and (re-)using software systems in a way that requires as little energy and natural resources as possible.**" Green Coding methods or practices thus mean any action or use of technology intended and suitable to further this (Junger et al., 2024). Moreover, GC concerns itself with the entire lifecycle since software may have a long history, with its resource consumption not limited to the version currently in use.

Against the backdrop of the rapidly growing number of data centres and volumes of data being hosted – reasons for which include but are not limited to the increasing usage of cloud computing, artificial intelligence and video streaming – there is an urgent demand to optimise the resource consumption of the respective applications and the data operation processes of data centres. GC may be a promising concept for both of these issues, as it helps guide practitioners to become more conscious about the material and energy consumption of the systems they design and use. All hardware and software are tied to inevitable resource demands, and as long as they are associated with an environmental footprint, their usage needs to be carefully considered.

3. What is the current situation concerning GC?

There exists a substantial body of scientific research addressing the concept of sustainable software. Research in the context of GC assesses a broad range of aspects, including the energy requirements of different programming languages (Pereira et al., 2017), the holistic energy consumption throughout the software development process and the criteria for ensuring sustainable software development practices (Gröger et al., 2018). Additionally, there has been an increased focus on the energy consumption of machine learning models at various stages of their lifecycle (Kaack et al., 2022; Luccioni et al., 2022; Strubell et al., 2019). However, despite the focus on sustainability in a variety of other sectors, the IT industry has largely been overlooked and has not yet prioritised sustainable software development. A study evaluating the extent to which GC practices are used in the IT industry revealed that, while half of the surveyed software companies consider GC a strategic goal, less than 17% of these companies have consciously integrated GC strategies into their operations (Henze et al., 2023).

To advance towards more sustainable software, it is essential to educate developers. Higher education institutions must ensure that graduates are aware of sustainability challenges and know how to address them. Moreover, sustainability should be treated as a cross-sectional issue. Even with well-trained developers, resource-efficient software development will not be achieved unless it becomes a (sub-)target within companies. If companies and customers prioritise new features over addressing technical debts that impact quality – as can be seen in many software projects – practices aimed at improving resource efficiency will not be implemented.

A significant challenge also concerns the current lack of transparency regarding software resource consumption. In order to optimise software towards resource efficiency, resource consumption needs to be measured. Since software typically comprises and utilises a variety of services, it is crucial for all components to disclose their resource consumption publicly. Politicians could push the issue further by implementing regulations that require transparency in resource consumption.

Currently, nearly every initiative to reduce software resource consumption faces a similar challenge: it is **methodologically demanding** to precisely measure (in the sense of directly attribute) emissions for a specific application or service. Thus, an essential first research step is to establish an empirically sound basis for measurement with widely trusted and reliable reference standards and public benchmarks. Such a foundation is crucial for the subsequent development of broadly accepted labels for software and services. Some projects, such as ECO:DIGIT, address this challenge by developing an automated assessment environment that provides transparent metrics and data on the resource consumption and CO_2 emissions of software applications. However, even though increasing resource efficiency in every resource-consuming context is absolutely desirable, efficiency gains alone will not be sufficient to achieve the necessary transformation towards circular and emission-free methods of production and consumption.

The societal perspective also needs to be cautiously considered. GC must not be used to shift the focus of public debate away from the (oftentimes critical) materiality of digital transformation. Additionally, there are various issues related to the openness and inclusivity of digital infrastructure. Another important topic in the current era of AI concerns the size of data centres; the larger the data centre, the more efficiently it can process large volumes of data, but the quest for ever-greater efficiency may create the risk of concentrating power in very few hands. Careful reflection is needed to determine the extent to which we want our access to the internet to depend on a few international actors. It is apparent that adequate governance is needed to ensure that digital transformation remains inclusive and reliable while balancing the conflicting paradigms of increasing efficiency and openness.

4. What role do open-source approaches play regarding GC?

Considering the four pillars of open source – ***Freedom to use*, *Freedom to study*, *Freedom to modify* and *Freedom to redistribute*** – the advantages of its utilisation become evident. The availability of source code not only facilitates the easy adoption of well-established patterns in related projects but also creates a basis for thorough and widespread analysis. An example of such analysis is the work done by Eugenio Capra and colleagues, who examined the use of application development environments in relation to energy efficiency (Capra et al., 2012). This analysis allows for the refinement of development paradigms and a shift towards more sustainable ways of development. Furthermore, it provides the needed transparency regarding the methods applied.

It is unsurprising that a variety of methodologies and available datasets originated from projects within the open source community, such as Boavizta. Various efforts have arisen within the broader community over the past several years, with opensustain.tech and the OSS for climate podcast providing fitting examples. Still, the power of the decentralised nature of open source is also its flip side: the vast amount of software and approaches lead to the need for standardisation and the development of best practices.

5. Advancing research and practical implementation of GC – what key steps and strategies are needed?

Following the principle "you can't manage what you can't measure", it is crucial to encourage companies to measure the environmental impact of their software. However, as stated above, the process of acquiring accurate measurements is challenging for many companies. In response to increased public awareness of climate change, international digital corporations have started to include digital emissions in their annual sustainability reports and equip their software services with tools that help customers track and assess their sustainability. These approaches, however, are based on varying assumptions and often lack **transparency** if the underlying methodologies remain undisclosed and the results cannot be verified or replicated by external parties. To address this challenge, there is an urgent need for **trusted and reliable reference standards and public benchmarking** to allow for comparability and transparency (which in turn can also help to prevent greenwashing). These standards should enable companies to measure their software's environmental impact based on transparent and scientifically sound methods. In addition, we need to advance environmental accounting and measurement methodologies, not only for the resource consumption of software but also for communication (network) and computing (data centre) infrastructure. Since the application of standards is generally voluntary and standards only become binding when included in contracts or mandated by law, policy measures are also necessary to effectively promote GC practices. Furthermore, transparency regarding applied methods can be achieved by releasing building blocks as open source, which would allow an even wider range of adoption as other methods can be built upon the same foundations and be improved together.

From a societal perspective, we need to raise awareness in all relevant stakeholder groups (developers, industries, political actors), as well as develop **governance concepts** to cope with the further centralisation (resp. concentration) of power, as environmental targets for the operation of digital technologies and ICT infrastructure can more easily be met at scale.

From an industry perspective, it is essential to examine how GC can be broadly integrated into software development. Successful case studies could demonstrate to enterprises that investing in sustainable software development is not only an ethical choice but also a cost-saving measure. While such metrics are necessary to inform economic decisions, no comparative data is currently available.

From an education perspective, sustainability must be integrated into curricula and education. Awareness that sustainability is a cross-sectional topic that needs to be addressed is necessary not only for those who actually build the software but especially for those at C-level.

In summary, to reduce the environmental impact of software, it is essential for companies to **measure** accurately, despite the challenges doing so poses. Transparent, **scientifically sound reference standards** and public **benchmarks** are urgently needed to ensure comparability and prevent greenwashing. **Policy** measures and the adoption of **open-source** methodologies can drive the broader implementation of GC practices. By integrating sustainability into **education,** industry practices and governance frameworks, we can create an **inclusive and resource-efficient digital future**.

Acknowledgements

Grischa Beier would like to thank his colleague Malte Reißig for the critical suggestions and reflections on the topic, which have enriched his work.

References

Bieser, J. C. T., Hintemann, R., Hilty, L. M., & Beucker, S. (2023). A review of assessments of the greenhouse gas footprint and abatement potential of information and communication technology. *Environmental Impact Assessment Review*, *99*, 107033. https://doi.org/10.1016/j.eiar.2022.107033

Capra, E., Francalanci, C., & Slaughter, S. A. (2012). Is software "green"? Application development environments and energy efficiency in open source applications. *Information and Software Technology*, *54*(1), 60–71. https://doi.org/10.1016/j.infsof.2011.07.005

Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G. S., & Friday, A. (2021). The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. *Patterns*, *2*(9), 100340. https://doi.org/10.1016/j.patter.2021.100340

Gelenbe, E. (2023). Electricity Consumption by ICT: Facts, trends, and measurements. *Ubiquity*, *2023*(August), 1–15. https://doi.org/10.1145/3613207

Gröger, J., Köhler, A., Naumann, S., Filler, A., Guldner, A., Kern, E., Lorenz, H., & Maksimov, Y. (2018). *Entwicklung und Anwendung von Bewertungsgrundlagen für ressourceneffiziente Software unter Berücksichtigung bestehender Methodik*. Umweltbundesamt.

Henze, C., Junger, D., Schauermann, E., & Weber, S. (2023). *Potentials of Green Coding: Perspectoves on implementation*. https://gi.de/en/aktuelles/projekte/en-green-coding

Hilty, L. M., Coroama, V., de Eicker, M. O., Ruddy, T. F., & Müller, E. (2009). *The Role of ICT in Energy Consumption and Energy Efficiency*. Report to the European Commission, DG INFSO, Project ICT ENSURE: European ICT Sustainability Research.

IEA. (2023). Tracking Clean Energy Progress 2023. IEA. https://www.iea.org/reports/tracking-clean-energy-progress-2023

Junger, D., Westing, M., Freitag, C. P., Guldner, A., Mittelbach, K., Obergöker, K., Weber, S.,Naumann, S., & Wohlgemuth, V. (2024). *Potentials of Green Coding—Findings and Recommendations for Industry, Education and Science—Extended Paper* (arXiv:2402.18227). arXiv. http://arxiv.org/abs/2402.18227Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change*, *12*(6), 518–527. https://doi.org/10.1038/s41558-022-01377-7

Luccioni, A. S., Viguier, S., & Ligozat, A.-L. (2022). *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model* (arXiv:2211.02001). arXiv. http://arxiv.org/abs/2211.02001

Pereira, R., Couto, M., Ribeiro, F., Rua, R., Cunha, J., Fernandes, J. P., & Saraiva, J. (2017). Energy efficiency across programming languages: How do energy, time, and memory relate? *Proceedings of the 10th ACM SIGPLAN International Conference on Software Language Engineering*, 256–267. https://doi.org/10.1145/3136014.3136031

Strubell, E., Ganesh, A., & McCallum, A. (2019). *Energy and Policy Considerations for Deep Learning in NLP* (arXiv:1906.02243). arXiv. http://arxiv.org/abs/1906.02243

Social justice, governance, and citizen participation

Where will they settle? On the role of uncertainty and choice of algorithm for humanitarian decisions

Tina Comes, Meyke Nering Bögel, and Martijn Warnier

Abstract: Migration is among the most uncertain and contested topics for policymaking. The increasing number of migrants and refugees globally necessitates effective planning and management, particularly in addressing infrastructure needs such as access to healthcare. While efforts to accommodate a surge of refugees prioritise primary needs, improving structural access to essential infrastructure becomes imperative over time. However, the path-dependent nature of the expansion of refugee settlements poses challenges for infrastructure development. Existing facility location models for infrastructure planning overlook the interplay of infrastructure growth and human behaviour. This chapter presents a study on the interplay between the settling preferences of refugees (behaviour) and the location of healthcare facilities as essential infrastructure. We develop a data-based approach that combines an agent-based model representing decision behaviour with facility location optimisation models for infrastructure planning. Through a case study of Cox's Bazar, Bangladesh, home to over 1 million Rohingya refugees, we demonstrate the implications of different optimisation approaches and thereby explore how and in how far digital tools influence policymaking on one of the most contested and uncertain topics in the current policy landscape. Our findings underscore the importance of integrating uncertainty about human behaviour in infrastructure decisions.

Keywords: refugee settlement, healthcare, optimisation, complex systems, agent-based model, adaptation, digitalisation

1. Introduction

Migration is a substantial driver of population change and, thereby, policy change. It has been shown that the uncertainties associated with migration have a major impact on population forecasts and the assumptions underlying many policies (Azose et al., 2016). One of the most important drivers of migration is geopolitical conflicts and wars. The worldwide refugee crisis has reached unprecedented levels, with a record number of more than 108 million people who have been forced to flee their homes in 2022.[1] Sure enough, forced migration and refugee streams pose important challenges to host communities and countries. A surge of refugees owing to geopolitical conflict presents major uncertainty for planning and policymaking. For instance, the war in Ukraine has led to about 6 million refugees, mostly seeking protection in the EU.[2] Similarly, after more than a decade of conflict, more than 14 million Syrians have been forced to flee their homes.[3]

The example of refugee camps serves here to demonstrate the interplay of behavioural uncertainty and planning decisions, paradigmatic for many policy decisions in the migration space. By using the case study of the Rohingya refugee camps in Bangladesh, we show how digital tools can be used to conceptualise and capture behavioural uncertainty. Given the increasing push for digital tools and automated decisions, especially in highly complex and uncertain conditions, we also discuss how digital tools for planning and policymaking influence and drive the emergence of structural patterns in migration (crises) and how the different objectives that are embedded in the algorithm shape the emergence of different access conditions.

2. Case study: Rapidly expanding refugee camps

Today, about 23 million refugees live in camps, where refugees are provided with essential services and goods, including water, food, shelter, emergency relief items, and healthcare. Humanitarian responders and NGOs alike strive to address these primary needs but often face resource limitations, especially if there is a rapid and massive influx of refugees. Host states, especially developing countries with limited resources, struggle to provide essential services, leading to environmental degradation and strained infrastructure (Fransen et al., 2024).

Given the rapid influx of people in need, refugee settlements often emerge and grow ad-hoc, driven by the initial settlement choices of refugee groups. Several studies use remote sensing and satellite imagery to understand how camps grow (Bjorgo, 2000), including infrastructure growth

1 see UNHCR Global Trend Report 2023
2 https://data.unhcr.org/en/situations/ukraine
3 https://www.unrefugees.org/news/syria-refugee-crisis-explained/

(Tomaszewski et al., 2016). For instance, as refugees settle close to those with similar backgrounds, new settlements often emerge where they cluster together. Much less is known, however, about how planning algorithms and digital tools, combined with the accessibility of infrastructure, influence the growth of refugee settlements and how, in turn, this growth impacts the expansion of infrastructure, see Figure 1.

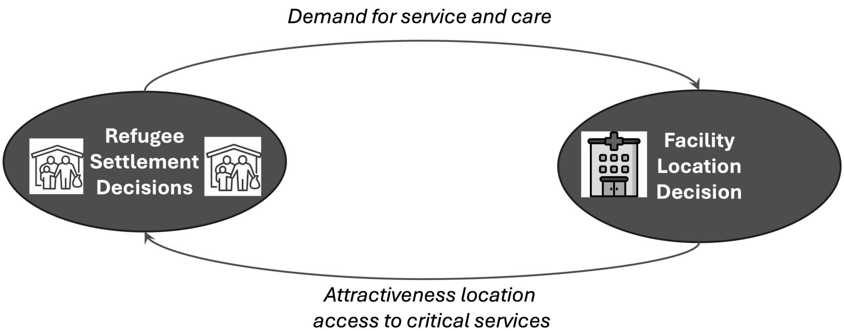

Figure 1. The interplay between settling decisions of refugees and facility location decisions

Traditionally, refugee camps have been viewed as temporary locations providing emergency shelter and services (Jahre et al., 2018). Standards such as the Sphere standards[4] provide guidance on the space requirements and service levels. However, many refugee camps are becoming permanent homes for their inhabitants (Hermans et al., 2017), and it has been argued that these settlements should rather be considered 'cities' instead of 'camps' (Montclos & Kagwanja, 2000). Nevertheless, because of the assumed short-term nature of the camps, the need for further growth or adaptation is neglected when camps are established. Initial efforts for camp design and management prioritise meeting urgent and primary needs. This is in line with other findings around crisis response (Comes et al., 2020) or resilient urban planning (Krishnan et al., 2024).

Over time, however, the demand for services and infrastructure has expanded. For example, in the Rohingya refugee camps in Cox's Bazar, Bangladesh, the percentage of refugees that indicated more attention should

4 https://www.spherestandards.org

primarily be given to healthcare increased from less than 6% in 2018[5] to 44% in 2019.[6] However, the high population density and limited space in camps limit the available locations for healthcare facilities. The lack of guidelines for long-term adaptation and the uncertain impacts of climate change further compound the challenges.

We argue that refugee settlements can be understood as complex adaptive systems in which the decisions of individual actors and the infrastructural decisions co-evolve (cf. Figure 1). In the following, we investigate the role of the digital tools and algorithms that are used to support the planning decisions in this interplay.

5 https://data.humdata.org/dataset/iom-bangladesh-needs-and-population-monitoring-npm-round-13-site-assessment

6 https://reliefweb.int/report/bangladesh/iscg-situation-report-rohingya-refugee-crisis-cox-s-bazar-10-january-2019-covering

Figure 2. The layout of camps 14, 15, and 16 in Cox's Bazar, including healthcare facilities. Data retrieved from IOM Geoportal (https://iom.maps.arcgis.com/home/index.html)

As one of the central infrastructures that exhibits a growing need, we focus on healthcare. To illustrate our approach, we develop a case study for the camps Hakimpara, Jamtoli, and Potibonia (camps 14, 15, and 16) in Cox's Bazar (see Figure 2), one of the world's largest refugee settlements. The camps appeared after the outbreak of the Rohingya crisis in August 2017, when the government of Bangladesh allocated undeveloped forest land for these camps in September 2017. Camps 14 and 15 were first mentioned in site assessments in September 2017, followed by camp 16 in October of the same year.[7] The two main camps (Kutupalong and Balukhali) already existed, and many NGOs had been providing healthcare in these camps for years before 2017. Therefore, by the time the situation reached emergency level 3, the necessary infrastructure was there to perform monthly needs assessments for the population in all camps. This also happened in the new camps 14, 15, and 16, resulting in regular situation reports. As camps 14 and 15 were first assessed in September 2017, this research uses data from September 2017 until June 2019.

3. Research approach and models

We start from the premise that migration and refugee settlements are both planning and social problems. To simulate the interplay of human settlement behaviour and infrastructure planning, we combine an agent-based simulation model to explain camp expansion patterns with optimisation models for facility location. Figure 3 provides an overview of three main modelled processes: camp expansion, healthcare use and the creation of new facilities (along with their location), and the interplay of behaviour and planning. The camp expansion and healthcare use cover refugee behaviour, whereas the facility creation falls under the realm of planning and policy problems. By connecting both realms, we can analyse how they interact. The source code of all models and the underlying datasets are available on GitHub: https://tinyurl.com/RefugeeSettlement.

The refugee decisions on shelter locations and healthcare provider decisions on healthcare facility locations are implemented in an agent-based model to study the emergent camp expansion. Various scholars acknowledge that the emergence of camps should be regarded as a process (Augustijn-Beckers et al., 2011; Ligmann-Zielinska & Sun, 2010). Therefore, the design of a camp should also be viewed as a process with no definite end state (Kennedy, 2008), which is reflected in the process diagram in Figure 3. Okyere et al. (2021) add that this process can be understood in terms of refugee settling choices. Hence, the emergence of the camp, resulting from the collective behaviour of all agents, can be researched through experiments in an agent-based model (Anderson et al., 2006). Our model follows earlier studies on the formation of informal settlements (Collins &

7 https://reliefweb.int/report/bangladesh/bangladesh-humanitarian-response-plan-september-2017-february-2018-rohingya

Frydenlund, 2016; Hofmann et al., 2015; Suleimenova et al., 2017). These models are effective in imitating settlement behaviour but neglect how the settlement choices were affected by the presence or lack of access to infrastructure, such as healthcare facilities (Hofmann et al., 2015).

To understand the impact of digital tools for planning and decision support, two different optimisation approaches are tested as tools to support the location decision health facilities: facility location optimization problems, and p-median and maximum coverage problems. Increasingly, approaches to support or even automate decisions are introduced in crisis management and humanitarian crisis response (Comes, 2024; Coppi et al., 2021). While there is a discussion on the use of principles, trustworthy, transparent, or explainable tools or AI, the implications of using different optimisation approaches have barely been explored.

Determining the (best) location of healthcare facilities falls under the category of facility location problems (FLPs), optimisation problems that involve a location decision for a facility that serves a number of demand centres with a certain demand level for the lowest possible effort or cost (Melo et al., 2009). Facility location decisions can be used to assess the location of a camp as well as the specific location of separate facilities within a camp, such as medical or educational facilities (Cilali et al., 2021).

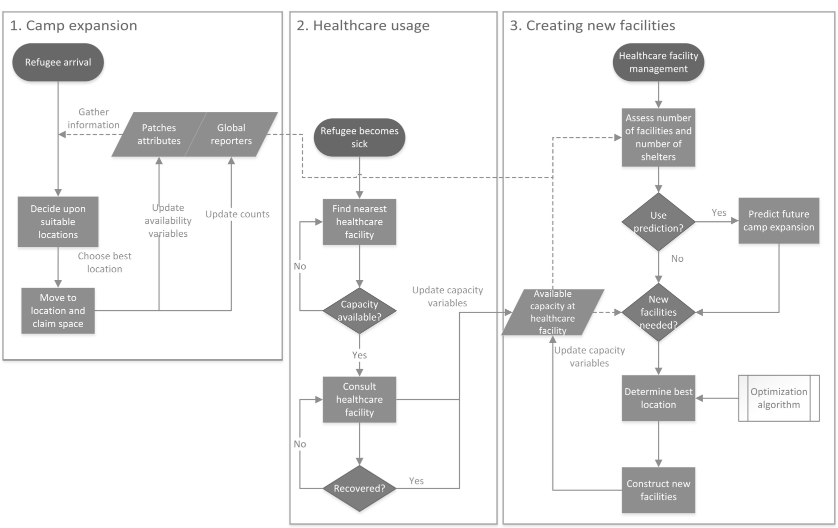

Figure 3. Flow diagram of the three main processes: camp expansion, healthcare usage, and creation of new healthcare facilities

To analyse the implications of using different digital tools, we here use the p-median and the maximum coverage problems, which are prominent approaches in location planning for humanitarian logistics (Ayudhya, 2020; Balcik & Beamon, 2008; İbrahim Miraç & Eren, 2020). For a recent review of humanitarian facility location problems under uncertainty, see Dönmez et al. (2021). Therefore, we investigate their implications to understand the choice of algorithm for infrastructure planning under uncertainty. The p-median approach aims to minimise the average demand-weighted travel distance, in our case, between shelters and healthcare facilities. The maximum coverage approach aims to maximise the number of shelters covered by the capacity of a healthcare facility. Here, we use a radius of 400 m.

To understand the influence of uncertainty, we test whether using predictions about future camp expansion based on the settling preferences of refugees improves the results of the planning algorithms. If future predictions are used, healthcare providers can adapt their location decisions to the settling preferences of refugees. Simultaneously, the placement of new facilities designed for expected camp expansion can affect the settling choices of refugees. This interplay is analysed in this research and applied in an approach for decision-making on healthcare facility locations.

The effect of location decisions is measured in the accessibility of healthcare facilities. While we acknowledge that there are many social, cultural, and financial factors that determine access to healthcare (Dawkins et al., 2021), we focus here primarily on physical factors: capacity of the facility and waiting and travel times as primary barriers that are especially relevant in refugee settlements (Aylett-Bullock et al., 2022). The accessibility is determined using four indicators: (i) the travel distance between shelters and healthcare facilities, (ii) the ratio of refugees allocated to a healthcare facility that has sufficient capacity to cover all allocated demand, (iii) capacity shortages across all facilities, and (iv) the ratio of patients who are waiting for treatment over the unused capacity. These first two indicators measure the accessibility of healthcare, and the latter two provide insight into the distribution of healthcare facilities throughout the camp.

4. Results: The interplay of refugee behaviour and location decisions

Regarding the location decisions of healthcare providers, both location optimisation methods have been found to be effective in achieving accessible healthcare for the camp inhabitants in terms of Sphere standards. For the p-median problem, refugees are mostly not adapting their settlement choices successfully: their settling choice to the located healthcare facilities does not improve the overall accessibility of healthcare facilities. For the maximum coverage problem, adapting settling choices by refugees increases the accessibility of healthcare facilities.

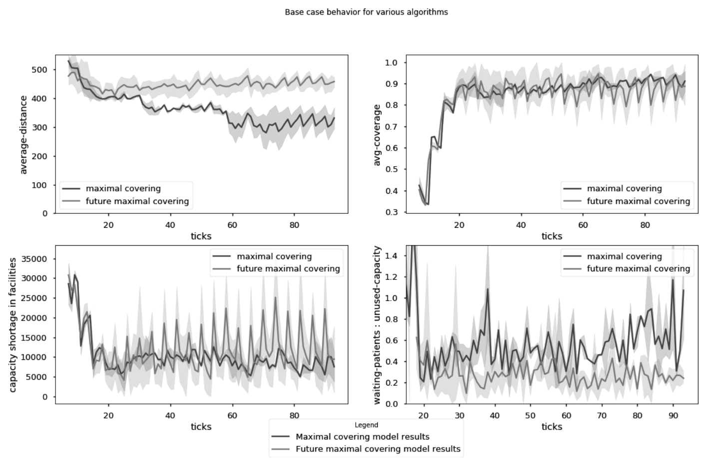

Figure 4. Impact of future predictions on KPIs when the focus lies on maximising equal allocation of healthcare for all shelters (max coverage problem)

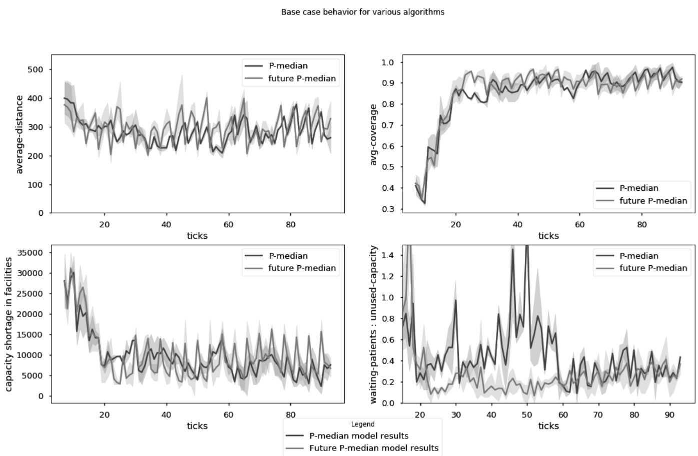

Figure 5. Impact of future predictions on KPIs when the focus lies on minimising the average travel distance to healthcare facilities (p-median problem)

Figures 4 and 5 show the difference between the results when distinguishing between the models that take future camp expansion into account while placing facilities (orange lines), the models that do not consider future expansion (blue lines) for the p-median problem (Figure 5), and the maximum coverage problem (Figure 4). For each algorithm, the indicator that is subject to the optimisation shows bigger fluctuations when future predictions are included. In Figure 5, this can be seen in the average distance results, and in Figure 4, this can be seen in the average coverage results.

Interestingly, including future predictions does not seem to lead to a significant improvement in the results. Moreover, the average distance between facilities and shelters is much higher with predictions for the p-median problem: the distance stabilises between 400 and 500 metres, while the model *without* future predictions returns results between 250 and 400 metres. For the maximum coverage problem, there is no such significant difference found.

This finding implies that when refugees can choose a location to settle upon arrival in a refugee camp, facility location optimisations improve when taking expected camp expansion into account. Thus, the expected camp expansion should be determined based on the settling preferences of refugees. However, if refugees *cannot* choose their settling location, the inclusion of expected camp expansion in the facility location optimisation does not improve the resulting accessibility of healthcare facilities. In other words, *if behavioural uncertainties exist and shape the expansion of the camp, the results improve when these uncertainties are integrated into the model.*

Naturally, behaviour is associated with uncertainty. Therefore, our next experiment explores the role of prediction accuracy in the results. Figure 6 shows the impact of predictions about future camp expansion with high or low prediction accuracy. The number of predicted shelters always equals 20 in these runs. It appears that a higher prediction accuracy can decrease the average distance in both algorithms. This effect is small for the p-median problem but is considerable (around 10%) for the maximum coverage problem. Simultaneously, the ratio of waiting patients over the unused capacity is slightly higher in the models with 100% prediction accuracy. This result means that the number of waiting patients is slightly higher or that the unused capacity is lower in these models.

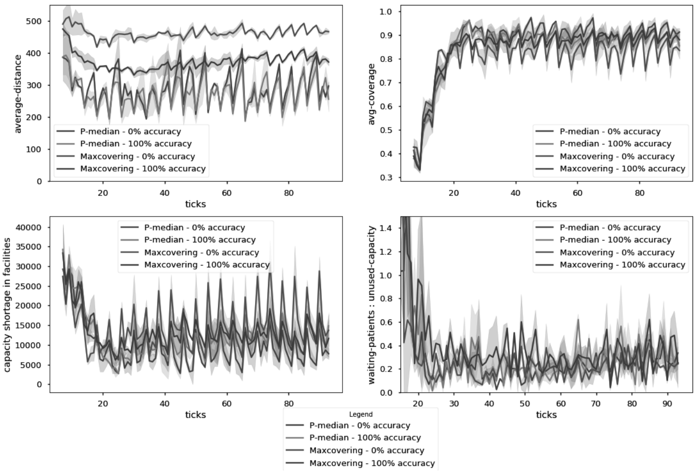

Figure 6. Effect of varying the prediction accuracy of new shelters while optimising facility locations for both algorithms and varying levels of accuracy and uncertainty

5. Conclusions

This research represents a first attempt to explore the interplay between refugee behaviour and health facility location decisions in refugee settlements and review the role of digital tools and planning algorithms and the impact of foresight. We first showed that the choice of algorithm, especially the ambitions and objectives that the algorithm represents, has a major impact on access to infrastructure and also on the expansion and growth of a settlement (Figures 4 and 5). Second, we show that taking into account behavioural uncertainty matters, especially if populations have freedom of choice. Third, Figure 6 shows that prediction accuracy has a significant influence, especially on travelling distances.

The findings directly highlight the importance of considering settlement choices made by refugees in optimising healthcare accessibility. By using two of the most popular facility location algorithms in humanitarian problems (p-median and maximum coverage), we also show that the choice of the specific optimisation algorithm has an important impact on performance. In sum, these results show that the digital tools (algorithm chosen and foresight accuracy) impact the quality of access and the emergence of camp infrastructures, as highlighted in Figure 7. This result is especially important given the current move in the humanitarian sector to increasingly use digital tools to support or fully automate decisions (Comes, 2024).

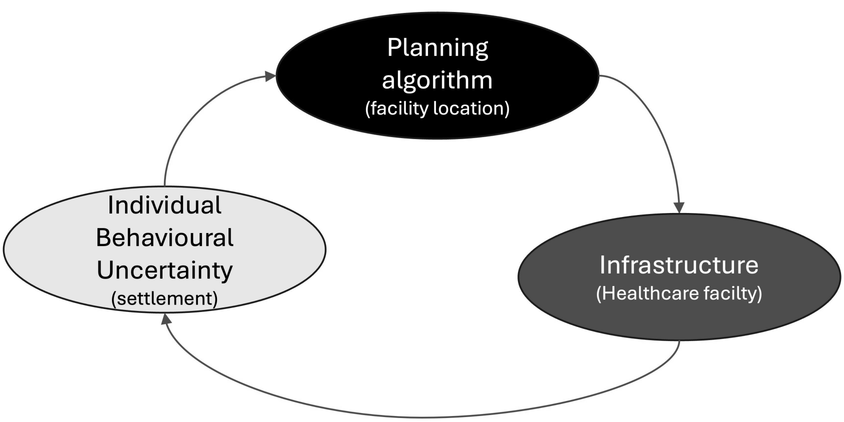

Figure 7. On the role of digital algorithms and planning tools in social-technical systems

This study has some limitations that warrant further investigation. Future work could explore scenarios and contexts and examine alternative strategies for healthcare facility location optimisation. Additionally, incorporating a broader range of healthcare quality and performance indicators, exploring dynamic capacity considerations, improving the accuracy and complexity of predictions for camp expansion, and adding additional vital infrastructures would enhance the modelling framework.

References

Anderson, J., Chaturvedi, A., Lengacher, D., & Cibulskis. M. (2006). Modeling the health of refugee camps: An agent-based computational approach. *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems*. https://dx.doi.org/10.1109/CBMS.2006.118

Augustijn-Beckers, E. W., Flacke, J., & Retsios, B. (2011). Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. *Computers, Environment and Urban Systems*, *35*(2), 93–103.

Aylett-Bullock, J., Tucker Gilman, R., Hall, I., Kennedy, D., Evers, E. S., Katta, A., Ahmed, H., Fong, K., Adib, K., Al Ariqi, L., Ardalan, A., Nabeth, P., von Harbou, K., Hoffmann Pham, K., Cuesta-Lazaro, C., Quera-Bofarull, A., Maina, A. G. K., Valentijn, T., Harlass, S., Krauss, F. … Luengo-Oroz, M. (2022). Epidemiological modelling in refugee and internally displaced people settlements: Challenges and ways forward. *BMJ Global Health*, *7*(3), e007822. https://doi.org/10.1136/bmjgh-2021-007822

Ayudhya, W. S. N. (2020). Flood shelters location using p-median model. *2023 8th International Conference on Business and Industrial Research (ICBIR*), 765–768. https://dx.doi.org/10.1109/ICBIR57571.2023.10147525

Azose, J. J., Ševčíková, H., & Raftery, A. E. (2016). Probabilistic population projections with migration uncertainty. *Proceedings of the National Academy of Sciences*, *113*(23), 6460–6465. https://doi.org/10.1073/pnas.1606119113

Balcik, B., & Beamon, B. (2008). Facility location in humanitarian relief. *International Journal of Logistics: Research and Application*, *11*, 101–121.

Bjorgo, E. (2000). Using very high spatial resolution multispectral satellite sensor imagery to monitor refugee camps. *International Journal of Remote Sensing*, *21*(3), 611–616.

Cilali, B., Barker, K., & González, A. D. (2021). A location optimization approach to refugee resettlement decision-making. *Sustainable Cities and Society*, *74*, 103–153.

Collins, A. J., & Frydenlund, E. (2016). Agent-based modeling and strategic group formation: A refugee case study. *WSC '16: Proceedings of the 2016 Winter Simulation Conference*, 1289–1300.

Comes, T. (2024). AI for crisis decisions. *Ethics and Information Technology*, *26*(1), 1–14.

Comes, T., Van de Walle, B., & Van Wassenhove, L. (2020). The coordination-information bubble in humanitarian response: Theoretical foundations and empirical investigations. *Production and Operations Management*, *29*(11), 2484–2507.

Coppi, G., Moreno Jimenez, R., & Kyriazi, S. (2021). Explicability of humanitarian AI: A matter of principles. *Journal of International Humanitarian Action*, *6*, 1–22.

Dawkins, B., Renwick, C., Ensor, T., Shinkins, B., Jayne, D., & Meads, D. (2021). What factors affect patients' ability to access healthcare? An overview of systematic reviews. *Tropical Medicine & International Health*, *26*(10), 1177–1188.

de Montclos, M. A. P., & Mwangi Kagwanja, P. (2000). Refugee camps or cities? The socio-economic dynamics of the Dadaab and Kakuma camps in Northern Kenya. *Journal of Refugee Studies*, *13*(2), 205–222.

Dönmez, Z., Kara, B. Y., Karsu, Ö., & Saldanha-da-Gama, F. (2021). Humanitarian facility location under uncertainty: Critical review and future prospects. *Omega*, *102*, 102393.

Fransen, S., Werntges, A., Hunns, A., Sirenko, M., & Comes, T. (2024). Refugee settlements are highly exposed to extreme weather conditions. *Proceedings of the National Academy of Sciences*, *121*(3), e2206189120. https://doi.org/10.1073/pnas.2206189120

Hermans, M. P. J., Kooistra, J., Cannegieter, S. C., Rosendaal, F. R., Mook-Kanamori, D. O., & Nemeth, B. (2017). Healthcare and disease burden among refugees in long-stay refugee camps at Lesbos, Greece. *European Journal of Epidemiology*, *32*, 851–854.

Hofmann, P., Taubenböck, H., & Werthmann, C. (2015). Monitoring and modelling of informal settlements: A review on recent developments and challenges. *2015 Joint Urban Remote Sensing Event (JURSE)*, 1–4.

Eligüzel, I. M., & Eren, Ö. (2020). P-median and maximum coverage models for optimization of distribution plans: A case of United Nations Humanitarian response depots. In P. Golinska-Dawson, K. M. Tsai, & M. Kosacka-Olejnik (Eds.), *Smart and sustainable supply chain and logistics–trends, challenges, methods and best practices, volume 1* (pp. 225–246). Springer Nature. https://doi.org/10.1007/978-3-030-61947-3_15

Jahre, M., Kembro, J., Adjahossou, A., & Altay, N. (2018). Approaches to the design of refugee camps: An empirical study in Kenya, Ethiopia, Greece, and Turkey. *Journal of Humanitarian Logistics and Supply Chain Management*, *8*(3), 323–345.

Kennedy, J. (2008). *Structures for the displaced: Service and identity in refugee settlements* [Doctoral dissertation, Delft University of Technology]. TU Delft Research Repository. https://repository.tudelft.nl/islandora/object/uuid%3A398c20b1-5bb8-4464-be47-b556b5899050

Krishnan, S., Aydin, N. Y., & Comes, T. (2024). TIMEWISE: Temporal dynamics for urban resilience-theoretical insights and empirical reflections from Amsterdam and Mumbai. *NPJ Urban Sustainability*, *4*(1), 4.

Ligmann-Zielinska, A., & Sun, L. (2010). Applying time-dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change. *International Journal of Geographical Information Science*, *24*(12), 1829–1850.

Melo, M. T., Nickel, S., & Saldanha-da-Gama, F. (2009). Facility location and supply chain management – A review. *European Journal of Operational Research*, *196*(2), 401–412. https://doi.org/10.1016/j.ejor.2008.05.007

Okyere, S. A., Kusi Frimpong, L., Diko, S. K., Abunyewah, M., & Kita, M. (2021). Situating everyday urban struggles within the context of the SDGs in an informal settlement in Accra, Ghana. In M. Addaney & P. Cobbinah (Eds.), *Sustainable urban futures in Africa* (pp. 265–287). Routledge.

Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific Reports*, *7*(1), 13377.

Tomaszewski, B., Tibbets, S., Hamad, Y., & Al-Najdawi, N. (2016). Infrastructure evolution analysis via remote sensing in an urban refugee camp–Evidence from Za'atari. *Procedia Engineering*, *159*, 118–123.

Digital sovereignty as an ill-structured (or wicked?) problem

Christian Herzog, Daniela Zetti*, and Robin Preiß*
**Both the first and second authors contributed equally*

Abstract: In this chapter, we discuss digital sovereignty as an ill-structured problem negotiated within democratic and participative discourses in public and private organisations. We argue that even though ill-structured problems describe challenges that cannot be solved in a formally structured way, tensions between knowledge and practice within discursive attempts at solutions and theoretical foundations may gradually lead to (increasingly) well-structured problem formulations. We first invoke Herbert Simon's analytical take on ill-structured problems to then apply the concept to digital sovereignty, the conception of which oscillates between the individual and the collective (inter-)national level. In light of issues and transgressions related to digital practices in violation or absence of digital sovereignty, however, we voice a call to resiliently pursue and engage in spaces of negotiations rather than succumbing to defeatism. Identifying digital sovereignty as an ill-structured problem can only emphasise the relevance of attempts to determine whether a transformation into a well-structured problem is possible. Accordingly, we contend that the discourse about digital sovereignty confirms an 'ill-structured problem' as a timely analytical term that helps enhance our understanding of problems shaped by the conditions of (digital) societies.

Keywords: digital sovereignty, ill-structured problem, wicked problem, democratic discourse, digital societies, digital transformation

1. Introduction

Digital sovereignty is often expressed as a nation or community's capacity for digital self-determination (Pohle & Thiel, 2020). Digital sovereignty is, as Keysar et al. (2012, p. 371) put it:

> a prescriptive normative term to describe various forms of autonomy, self-determination and independence in relation to digital infrastruc-

> tures, technologies and data. While territorial perspectives on the term are prevalent, understandings of digital sovereignty as digital self-determination and autonomy through collective control are increasingly articulated by civil society entities and social movements.

However, protecting citizens from the repercussions of digital transformation – especially in terms of preserving the capacity for self-determination – needs framing within contexts associated with caring for individual citizens' digital sovereignty. The individual and the territorial concepts are intimately and intricately connected (cf. Floridi, 2020). It is not yet entirely clear how approaches to preserve one contribute to safeguarding the other – and even if it would be – possible enactments involve several public and private stakeholders and powerful actors. Floridi (2020) portrays citizens and residents as being caught in the middle:

> Today, the fight is ... over corporate and political power over the digital, yet the roots of this clash are very old. But most importantly ... whoever will win the fight for digital sovereignty will determine the lives of all people on both sides of the digital divide. (p. 377)

Digital sovereignty is a complex topic involving power plays among resourceful multinational enterprises, public administration officials, and citizens, including political minorities and marginalised and disproportionally affected, yet mostly powerless, communities (Birhane, 2021; Bory & Zetti, 2022; Heikkilä, 2022). The topic is difficult to understand, explain, and contribute to by academia, non-government actors, industry, and – perhaps above all – individual citizens and residents. Debates on linguistic terms structure discourses favouring one party or another (cf., e.g. Fetzer, 2013). Emphases on the territorial meaning of digital sovereignty may hence draw attention from the representational deficit suffered by minorities and marginalised communities despite their epistemic privilege to voice legitimate concerns (cf. Bar On, 1992).

This chapter contributes to dissolving a potentially dichotomous interpretation, where one side may claim superior significance and discursive dominance over the other. Digital sovereignty can neither be solved as an exclusively (inter-)national nor as an individual problem, leading us to conceptualise the issue of digital sovereignty as a **wicked** – an **ill-structured** – **problem**. Ill-structured problems lack a clear definition. They are open-ended, and epistemic iterations are required to arrive at a more precise formulation of the problem or related sub-problems. From an interdisciplinary, ethical-historical perspective, we will attempt to render this conceptual understanding of the digital sovereignty discourse conducive to addressing and mitigating epistemic injustice (cf. Fricker, 2007). The notion of the ill-structured problem bears options for making all voices heard inclusively. We understand digital sovereignty as an issue that requires constant and shared deliberation, and we ask how actors can be empowered to contribute.

By evoking digital sovereignty as an ill-structured problem that requires an **iterative approach** to epistemically advance in understanding its nature and potential solution, we would like not to be mistaken as stating that a proper way to advance is to just go ahead by trial and error. There is too much at stake, as irretrievable attempts at implementations could lead to sustained power asymmetries. Instead, we offer a hopeful perspective: By understanding digital sovereignty as an ill-structured problem, we argue that pursuing digital transformation guided by proper inclusive deliberation will gradually lead to an understanding. We will primarily be adopting the notion of an 'ill-structured problem' as originally proposed by Herbert Simon (1973), rather than referring to 'wicked problems', which may be deemed synonymous but carry with them the idea that they are problems so severe that any attempt at a solution is futile. Or as Sandra Batie (2008, p. 1,176) puts it: 'wicked problems tend to be intractable'.

The paper is structured as follows: Section 2 introduces a historical perspective on so-called ill-structured problems, the concept's emergence, and recent adaptations. This section elucidates the core characteristics of ill-structured problems, which we will then apply to analysing the discourse around digital sovereignty in section 3. Section 4 draws conclusions and points to future work on understanding the strife to enact digital sovereignty as an ill-structured problem.

2. Ill-structured problems

From the beginning, Herbert A. Simon's 'ill-structured problem' (ISP) has been an analytical concept that allows for observing social problems in a world shaped by technology. Simon was trained in social sciences in the 1930s, which led him to work in such diverse fields as public administration, political science, computer science, psychology, sociology, and economics. Being a social scientist, Simon was also the founder of the Department of Computer Science at Carnegie Mellon University. He published extensively on the topic of decision-making and problem-solving in 'modern' organisations. In 1973, he published a paper entitled 'The structure of ill structured problems' in a journal called *Artificial Intelligence* (Simon, 1973). At this time, Simon had been co-developing computer programs for 20 years and had developed many close contacts with computer scientists (Crowther-Heyck, 2015).

In his 1973 seminal paper, he described a concept he witnessed emerging as a by-product of the '**well-structured problems**' **(WSPs)** that mathematicians and computer scientists were producing. They were eager to define problems that then could be handed over and processed 'autonomously' by computers in nicely structured ways. Yet, according to Simon, something else always remained – something was left over while structuring problems for automated processing, a residual. The '**ill structured problem**' **(ISP)** is a residual concept. An ISP is usually defined as a problem whose structure needs definition in some respect. More simply, a problem is an ISP if it is not a WSP (Simon, 1973, p. 181).

The observation that there is such a thing as a residual of the unstructured problem led Simon to conclude that it was impossible to draw a sharp boundary between ISPs and WSPs. The computer's way of solving problems fundamentally affected the (human) perception of human problem-solving: 'The boundary between well structured and ill structured problems is vague, fluid, and not susceptible to formalization. Any problem-solving process will appear ill structured if the problem solver is a serial machine' (Simon, 1973, 181). Simon emphasised that this was good news. Computing sciences were obviously capable of handling the emerging **fluid boundaries** by readjusting to the 'external real world', by testing feasibilities, and by elaborating the details of a plan. There was no need to 'introduce mechanisms that are qualitatively different from the ones already introduced in artificial intelligence schemes' (Simon, 1973, p. 200).

Simon was convinced that technological progress would not inhibit computer scientists from actually – if unknowingly – solving ISPs. 'There appears to be no reason to suppose that concepts as yet uninvented and unknown stand between us and the fuller exploration of those problem domains that are most obviously and visibly ill structured' (Simon, 1973, p. 200). In the end, Simon disproved the assumption that pure WSPs existed: computer scientists always had worked and would work on ISPs.

Since 1973, computing technology has developed in hitherto unprecedented ways. Standalone machines in computing centres soon would give way to computing networks. So-called personal computers entered households, and – last but not least – artificial intelligence (AI) has become a ubiquitous technology. While we should still agree that pure scientific and technological formalisation is fiction, we should also ask: Is Simon's **optimism** about future developments still appropriate? There are good reasons why it could be **outdated**. Perhaps computer scientists did not follow Simon's advice to rely on problem-solving skills to compensate for computing's by-effects? Maybe human problem-solving skills stagnated? Are computers still organised in the same way they once were? It might even be that the 'real world' in which computers operate and adapt has changed dramatically.

How has Simon's concept of the ISP developed? For some decades now, it has been adapted and refined by social scientists. Here are three recent definitions. In 2011, ethicists of technology van de Poel et al. built on and departed from the concept of WSPs and thus from the same point as Simon. But their interpretation also emphasises a) the definition of the ISP and b) the **process of problem-solving** as an open condition that iteratively defines and transforms the ISP into a WSP:

> Whereas well-structured problems (such as basic arithmetical calculations) usually have clear goals, fixed alternatives to choose from, usually maximally one correct answer and rules or methods that will generate more or less straightforward answers, ill-structured problems have no definitive formulation of the problem, may embody an inconsistent problem formulation, and can only be defined during the process of solving the problem. (Van de Poel et al., 2011, p. 136)

Ge et al. recently highlighted the process of learning and the **cognitive and communicative skills** involved in iterative problem-solving:

> Although ill-structured problem solving and well-structured problem-solving share similarities, problem solvers need to rely more on metacognitive skills in solving ill-structured problems, especially in interpreting the state or essence of the problem, defining the goals of a problem, comparing and selecting solutions, and generating defensible arguments. In addition, solving ill-structured problems takes more iterations between various processes in comparison to solving well-structured problems. (Ge et al., 2022, p. 97)

In another article, Love (2015) asks 'Why Herbert Simon matters for policymaking'. Ill-structured problems, or similarly, 'wicked problems', refer

> to questions like climate change, immigration, international trade and so on. … no agreement exists about what the problem is. Each attempt to create a solution changes the problem. The solution is not true or false – the end is assessed as 'better' or 'worse' or 'good enough'. The problem changes over time. … Solution(s) to the problem is (are) based on 'judgments' of multiple stakeholders. The problem is associated with high uncertainty as to system components and outcomes. There are no shared values with respect to societal goals. (Love, 2015, p. 930–931)

Love heavily emphasises the difficulty of even defining (i.e. agreeing on) the problem, the **openness of problem-solving processes**, and, last but not least, the multiple perspectives involved in processes that can result in agreements but do not have to. Love fuses Simon's ISP with the similar notion of the 'wicked problem', first introduced by Horst Rittel and Melvin Webber in 1973 (Gruendel, 2002; Hou et al., 2022; Oswald, 2022; Reznich, 2017; Rittel et al., 1973). Rittel and Webber reflected on problems that characterised planning processes and expert knowledge vis-à-vis the general public. For professionals, there were no more 'easy problems' to solve:

> The streets have been paved, and roads now connect all places, houses shelter virtually everyone, the dread diseases are virtually gone, clean water is piped into nearly every building, … and so on. … But now that these relatively easy problems have been dealt with, we have been turning our attention to others that are much more stubborn. The tests for efficiency, that were once so useful as measures of accomplishment, are being challenged by a renewed preoccupation with consequences for equity. (Rittel et al. 1973, p. 156)

In parallel to Herbert Simon, Rittel and Webber developed a concept to reflect on and optimise the relationship between technology and society. Their account today reminds us that history is sometimes a **wicked process**: problems solved bear new problems. Some problems persist. ISPs and wicked problems – analytical categories to describe social problems – are such persisting problems. There is no 'solution' to social problems, Rittel and Webber insist: 'Social problems are never solved. At the best, they are only re-solved – over and over again' (Rittel et al. 1973, p. 160). While these problems are not well-structured, in the words of Herbert Simon, they are nevertheless intimately linked with acts of formulation, communication, and negotiation: 'The formulation of a wicked problem is the problem.

The process of formulating the problem and of conceiving a solution (or re-solution) are identical' (Rittel et al. 1973, p. 161).

What can we take away from these conceptual works? ISPs and wicked problems are both analytical terms for social problems in societies characterised by technology. The focus is on the fact that only the treatment of a problem makes a definition possible. Authors today point out that the definition and treatment of a wicked problem or ISP does not have to be based on unity to resolve the problem. It is Herbert Simon, sitting in the shadow of a computer that back then, in the mid-20th century, was conceived to be accurate to the highest degree, who insisted that well-structured and ill-structured problems are neither identical nor do they resemble antagonistic poles. Instead, they share a 'boundary' that 'is vague, fluid, and not susceptible to formalization' (Simon, 1973, p. 181). Acknowledging this fluidity seems to be the precondition for further thoughts, negotiations, and acts.

3. Digital sovereignty as an ill-structured problem?

Digital transformation is one of the most pressing concerns of Western societies today. Many quantitative and qualitative advances are expected and hoped for, especially by introducing digitalisation into public administration systems (cf. Twizeyimana & Andersson, 2019). In addition, a shortage of skilled workers (cf. Müller, 2023), reduced public spending capacities (Gründler et al., 2023), and an obligation to responsibly allocate taxpayers' money may all seem to necessitate an increase in digitally supported and automated processes – at least in Germany. However, it has become clear that the introduction of digital means (AlgorithmWatch & Bertelsmann Stiftung, 2019), such as artificial intelligence (AI)-based algorithms in social welfare allocation (Heikkilä, 2022; Oravec, 2019), digital interfaces for interacting with public administration (Preiß et al., 2023), and predictive analytics for supporting police work (Amnesty International UK Section, 2018) all have the potential to put citizens' autonomy and flourishing at risk (Mittelstadt et al., 2016). Protecting citizens from these repercussions of digital transformation – especially in terms of preserving the capacity for self-determination – is often framed within contexts that are also associated with caring for individual citizens' digital sovereignty.

Digital sovereignty is commonly interpreted as referring to a state, an economy, or an individual's autonomous control over digital means and data, as well as over the structures and practical ways of governing them (Floridi, 2020; Moerel & Timmers, 2021). A recent scoping review by Müller et al. (2022) highlights the **various interpretations and aspects of digital sovereignty** and their interaction with individuals, collective actors, and structures. Their analysis of almost 50 scientific papers and non-scientific position papers shows that many interpretations of digital sovereignty focus on competencies such as self-determination and control but also on decision-making ability, security, autonomy, independence, participation, and co-creation (Müller et al., 2022, pp. 17–22).

Generally speaking, digital sovereignty is about being able to make concrete decisions based on, about, or with the digital. Even though this may be an apt synthesis of the different descriptions found in the literature (e.g. Floridi, 2020; Lambach & Oppermann, 2022; Moerel & Timmers, 2021; Pohle & Thiel, 2020; Wessel et al., 2023), it is also insufficient, for it is precisely the distinction between a state and an individual's digital sovereignty which appears to be fluid. For instance, by introducing the concept of cyberspace sovereignty, Couture and Toupin (2019) have indicated that fixing digital sovereignty to a nation's boundaries appears inherently complex, perhaps even problematic. They have further disambiguated the digital sovereignty of social movements, Indigenous communities, and individuals, suggesting that these taxonomies – though indeed useful in discourses to highlight the needs of disproportionally and often negatively impacted marginalised groups – may amount to almost arbitrary stratifications of the concept.

To be clear, in this chapter, we are not pursuing a single, unifying perspective on digital sovereignty out of the motivation to consolidate and structure its associated discourse in some particular way. This may be regarded as a fruitful undertaking (cf. e.g. Pohle & Thiel, 2020, 2021), but it is not ours. Rather, we would like to take an outside view, comment on the digital sovereignty discourse, and hopefully suggest an insightful perspective that is primarily inspired by Herzog and Zetti's (2022, p. 10) observation that 'often sovereignty *about* the digital is, in fact, a prerequisite for the realisation of sovereignty *via* the digital'. In fact, we would like to even regard **sovereignty *about* and *via* the digital** as co-requisites, meaning that we deem sovereignty in deciding over and sovereignly engaging with digital matters to be intimately connected. The latter, of course, also includes the need for individuals to be empowered to aptly apply digital tools as means conducive to their political sovereignty. Both Pohle and Thiel (2020) and Bekkers and Zouridis (1999) have criticised how modern approaches to the digital transformation of public administration often treat citizens as consumers, undermining their political engagement. In turn, Funke (2022, p. 190) surmises that 'public services in the digital era require that citizens as customers become the focal point of administrative activities, supplanting administrative traditions that were centered around the state's own needs and intrinsic logics'. This begs the question: What is digital sovereignty about – the state or the citizen's needs? Or is – and should – there be a difference? We suggest that this discourse and many others highlight the intricate but subtle correlations between digital and political sovereignty both on an individual and national scale.

Especially when going further along the lengths of the digital transformation, we postulate that, at least to some degree, we cannot and should not have one without the other: **national and individual concepts** of digital sovereignty are intimately connected. The fact that emphases in addressing digital sovereignty oscillate between the national and individual interpretation (Pohle, 2020) adds to this suspicion. In addition, scientific and political communities are discussing an aspect that is both highly relevant and is to be enacted in an ever-changing environment. The digital transformation more than just casts its shadow – it already has had a tangible impact and, at the same time, remains inherently open-ended. For example, people can only decide to share their data if they understand the importance and possibilities of general data use. However, it should be noted that managing your data protection is only possible to a limited extent, following Solove's (2020) observation, for example, that 'even totally rational people can't succeed at privacy self-management' (p. 27). Therefore, infrastructure regulations must ideally be congruent with citizens' demands, digital skills, and needs and should, hence, be designed accordingly.

Perhaps it is only in an idealised setting that individual and national digital sovereignty can be genuinely separate. For instance, at the very outset of digitally transforming some aspect of our lifeworlds, we may be in a position to autonomously decide whether to pursue the introduction of some specific digital technology or not. However, even then, society – and by this, we refer to non-expert citizens most commonly – will not be entirely sure about what it is getting into. Vallor (2016) denotes this kind of inability to assess, and perhaps even envision, the socio-ethical impact of technologies as ***socio-technical opacity*** and attests to our 21st-century societies suffering from it. While Vallor takes this to be valid grounds for refuting utilitarianism in favour of a virtue ethics approach more broadly, we would like to acknowledge this assessment in light of the digital sovereignty discourse. Society's inability to predict the implications of its own transformation basically prompts it to establish ways to continually assess what kind of digital transformative activities best suit its democratic ideals. Now that societies have already arrived at digital realities, we may have to adjust our conceptual and societal understanding of the precise meanings and implications that digital sovereignty carries with it. Digital sovereignty as an ill-structured problem refers to uncertainties and openness created through and shaped by **decision-making processes** with stakeholders that 'are likely to have differing ideas about what the "real" problem is and what its causes are' (OECD, 2017, p. 29).

The quest for digital sovereignty hence shows the basic features of an Ill-structured problem – we can only begin to fathom a more concise nature of the problem as we attempt to solve it. Disappointing as this is from an ethical perspective adhering to a rational foundationalist approach (cf. Whitbeck, 2011), it is also a call to action to continuously and tirelessly address the most pressing and current issues – and adopt a design approach to the digital transformation that carefully weighs the implications on questions of sovereignty (cf. Dorst and Royakkers, 2006). As national conceptions of digital sovereignty have guided the introduction of algorithmic techniques – mainly under the auspices of more efficient public spending (Dobrolyubova et al., 2019; Finger & Montero, 2021), disastrous failures affecting the most vulnerable, e.g. children (Heikkilä, 2022), the unhealthy (Obermeyer et al., 2019), and ethnic or other minority groups (Bélanger & Carter, 2006; Rosenberg, 2019), highlight the need to realign practices in advancing the digital transformation and promoting digital sovereignty towards solving issues for minorities or even for the disproportionally affected (e.g. Birhane, 2021). Without individual digital sovereignty that could amount to the effective notification of the algorithmic processing at play, the right to object to these practices, the right to demand human reevaluation on principle, as well as proper rights to an explanation of the conditions and parameters that influenced an algorithmic decision – and how, vulnerabilities amount.

The case documented extensively by Hadwick and Lan (2021) is cautionary: administrative practice made it possible for the algorithmically mediated detection of benefit frauds to harm thousands of families. This case – the so-called '*toeslagenaffaire*' – shows how a deterministic stance on techno-solutionism challenges citizens' digital sovereignty in multiple ways: The scandal did not just involve a reliable algorithm corrupted by administrative malpractice. Nor did it simply involve a biased algorithm.

Instead, the scandal illustrates how a strong but ultimately unjustified belief in algorithmic efficiency encouraged less tolerance towards presumptive fraudsters, including 'a lack of transparency of the norms regulating the algorithm' and 'the impossibility for welfare recipients to access the data retained by the administration' (Hadwick & Lan, 2021, p. 625). Hadwick and Lan (2021, p. 614) conclude: 'The assumption that it makes no difference to individual taxpayers whether a task is carried out by a human agent or an AI system, as often claimed by EU tax administrations is proven false by the *toeslagenaffaire*'.

Still, regional governments may focus on covering as much ground as possible in the pursuit of digitalisation with proclamations 'to leave nobody behind' being but shallow lip service as concrete actions are lacking. Take, for instance, Germany's call in Schleswig-Holstein for digital – first and only – as the principle in the e-government transformation (AöR & Schleswig-Holstein, 2023). This **digital strategy** firmly interprets digital sovereignty as independent of big-tech multinationals. However disappointing this may seem for big-tech multinationals and, more importantly, considering caring for those at risk of being negatively affected, it may be understandable from the perspective of a region that is trying to make headway and has, admittedly – to our knowledge, avoided failings anywhere near as dramatic as the Dutch children's welfare scandal (Heikkilä, 2022). A country's ability to commission, develop, maintain, and modify its digital solutions is an essential prerequisite to keeping it this way. Reviving the concept of sovereignty in this light can be called a minimum programme that allows the state as a sovereign to uphold basic functionalities.

We do not intend to suggest that we should make up our minds as we go along by analysing digital sovereignty as an ill-structured problem. However, only experience from implementing technological advances allows us to understand the **potentialities** both in the negative and positive sense, forcing us to realign our ethical priorities. Hence, we believe that by embracing digital sovereignty as an ill-structured problem, we are guided to do three things:

We should look for **experiences** made with digital technologies beyond our regional, national, and cultural boundaries, evaluate similarities and risks, and learn from them.

By considering the limitations of our own experience with the matter, we should recognise the need for **inclusive deliberation** involving a wide range of communities.

Acknowledging the endeavour's ill-structuredness keeps one alert. It raises the significance of a **continuing debate** about the direction actors are taking and reminds one to be wary of adhering to overly reductionist concepts that risk diverting attention from the more intricate and detailed socio-ethical impacts that are likely to occur.

4. Conclusions

Our analysis of digital sovereignty discourses supports Gruendel (2022, p. 205), who observes that the term 'wicked problem' historically 'points us to moments when the relationship between technocratic and democratic elements of modern institutions was at stake'. She notes that today, the concept continues to alert 'us to the unsettled roles of expertise and of public participation in political decision-making, particularly in ostensibly democratic systems' (Gruendel, 2022, p. 205). Gruendel, therefore, sees knowledge, understood non-exclusively as **knowledge** in its different forms, in the centre of 'wicked' governmental processes.

To this, we can add that reflections on digital sovereignty can gather momentum whenever they tackle the principle-to-practice gap, i.e. when they address the challenge of working in the **trading zone** that opens between (inter-)national and individual negotiations. Knowledge and practice need not go hand in hand, but they are not opposite poles per se. Herbert Simon's highly analytical work can help researchers and practitioners remain attentive and sensitive to those fluid and **vague boundaries** between ill-structured and well-structured problems. Our contribution shows that the exploration of these boundaries and the future openness they bring about is a research desideratum. Acknowledging that digital societies have a history that includes individual and collective experiences can eventually help explain uncertainties and shape the future.

Finally, a thought on the **wicked or ill-structured problem** is tested and sharpened using the example of discourses on digital sovereignty. Both concepts are analytical, intended as heuristic tools that are nonetheless social and real. Neither an ISP nor a wicked problem is a 'tricky thing'. Their inventors were concerned with issues that were not easy to formulate, and they believed that this was precisely why social opportunities opened up. The term ill-structured seems more suitable for us to convey this than the term wicked problem, which is close to defeatism. This is (again) not a problem per se, but it does run the risk of opening additional gaps.

In light of the issues and transgressions we highlighted that are related to digital practices in violation or the absence of digital sovereignty proper, we voice a call to resiliently pursue and engage in spaces of **negotiations** rather than succumbing to defeatism. Identifying digital sovereignty as an ill-structured problem can only emphasise the relevance of attempts to determine whether a transformation into a well-structured problem is possible. Ill-structured problems may even require transdisciplinary approaches due to individual actors or groups' limited epistemic privileges and horizons. It appears that this insight is close to being a truism in sustainability science – the science and practice of digital transformations can only learn from that. Accordingly, we purport that the discourse about digital sovereignty confirms the 'ill-structured problem' as a timely analytical term that helps enhance our understanding of problems shaped by the conditions of (digital) societies.

References

AlgorithmWatch, & Stiftung, B. (2019). Automating society: Taking stock of automated decision-making in the EU. https://algorithmwatch.org/en/automating-society-2019/

AöR, Dataport, & Schleswig-Holstein, S. (2023). Digitalstrategie Schleswig-Holstein. https://www.schleswig-holstein.de/DE/landesregierung/themen/digitalisierung/digitalisierung-zukunftsthema/Digitale-Verwaltung/digitalstrategie/digitalstrategie_node.html

Bar-On, B. A. (1992). Marginality and epistemic privilege. In L. Alcoff & E. Potter (Eds.), *Feminist Epistemologies*. Routledge, 83–100.

Batie, S. S. (2008). Wicked problems and applied economics. *American Journal of Agricultural Economics*, *90*(5), 1176–1191. https://doi.org/10.1111/j.1467-8276.2008.01202.x

Bekkers, V. J. J. M., & Zouridis, S. (1999). Electronic service delivery in public administration: Some trends and issues. *International Review of Administrative Sciences*, *65*(2), 183–195. https://doi.org/10.1177/0020852399652004

Bélanger, F., & Carter, L. (2006). The effects of the digital divide on e-government: An empirical evaluation. *Proceedings of the 3^{9t}h Annual Hawaii International Conference on System Sciences (HICS"06)*, Kauai, HI, USA, 1–7. https://doi.org/10.1109/HICSS.2006.464

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Bory, P., & Zetti, D. (2022). Digital federalism: Balancing autonomy, authority and automation. Itinera. *Schweizerische Zeitschrift für Geschichte*, 49, 6–17.

Couture, S., & Toupin, S. (2019). What does the notion of 'sovereignty' mean when referring to the digital? *New Media & Society*, *21*(10), 2305–2222. https://doi.org/10.1177/1461444819865984

Crowther-Heyck, H. (2005). *Herbert A. Simon: The bounds of reason in modern America*. Johns Hopkins University Press.

Dobrolyubova, E., Klochkova, E., & Alexandrov, O. (2019). Digitalization and effective government: What is the cause and what is the effect? In D. A. Alexandrov, A. V. Boukhanovsky, A. V. Chugunov, Y. Kabanov, O. Koltsova, & I. Musabirov (Eds.), *Digital transformation and global society* (Vol. 1038, pp. 55–67). Springer International Publishing. https://doi.org/10.1007/978-3-030-37858-5_5

Dorst, K., & Royakkers, L. (2006). The design analogy: A model for moral problem solving. *Design Studies*, *27*(6), 633–656. https://doi.org/10.1016/j.destud.2006.05.002

Fetzer, A. (2013). The structuring of discourse. In M. Sbisà & K. Turner (Eds.), *Pragmatics of speech actions* (pp. 685–712). de Gruyter. https://doi.org/10.1515/9783110214383.685

Finger, M., & Montero, J. (2021). Digitalization, efficiency and convergence. In J. Montero & M. Finger (Eds.), *A modern guide to the digitalization of infrastructure*. Edward Elgar Publishing. https://doi.org/10.4337/9781839106057.00017

Floridi, L. (2020). The fight for digital sovereignty: What it is, and why it matters, especially for the EU. *Philosophy & Technology*, *33*(3), 369–378. https://doi.org/10.1007/s13347-020-00423-6

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Funke, C. (2022). *Digitization, fast and slow – Comparing the creation of digital public services in Denmark, France and Germany* [Doctoral dissertation, European University Institute]. https://hdl.handle.net/1814/74971

Ge, X., Muftuoglu, A. C., & Brickell, S. (2022). Instructional design from the lens of self-regulated ill-structured problem solving. In J. Stefaniak & R. Reese (Eds.), *The Instructional Design Trainer's Guide* (pp. 77–89). Routledge. https://doi.org/10.4324/9781003109938-9

Gruendel, A. (2022). The technopolitics of wicked problems: Reconstructing democracy in an age of complexity. *Critical Review*, *34*(2), 202–243. https://doi.org/10.1080/08913811.2022.2052597

Gründler, K., Potrafke, N., & Weber, L. (2023). Ökonomenpanel zur 'fiskalischen Zeitenwende': Wie wollen Ökonomen die Zeitenwende im Haushalt realisieren? *Ifo Schnelldienst*, *76*(7), 68–71.

Hadwick, D., & Lan, S. (2021). Lessons to be learned from the Dutch childcare allowance scandal: A comparative review of algorithmic governance by tax administrations in the Netherlands, France and Germany. *World Tax Journal*, *13*(4), 609–645. https://doi.org/10.59403/27410pa

Heikkilä, M. (March 29, 2022). *Dutch scandal serves as a warning for Europe over risks of using algorithms*. Politico. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/#

Herzog, C., & Zetti, D. (2022). Digitally aided sovereignty – A suitable guide for the e-government transformation? *Practicing Sovereignty*. Weizenbaum Conference, Berlin, Germany. https://doi.org/10.34669/wi.cp/4.1

Hou, X., Li, R., & Song, Z. (2022). A bibliometric analysis of wicked problems: From single discipline to transdisciplinarity. *Fudan Journal of the Humanities and Social Sciences*, *15*(3), 299–329. https://doi.org/10.1007/s40647-022-00346-w

Keysar, H., Calderón Lüning, E., & Unteidig, A. (2021). Prototyping digital sovereignty. *Design*, *1*, 371–396. https://doi.org/10.14361/9783839457603-021

Lambach, D., & Oppermann, K. (2022). Narratives of digital sovereignty in German political discourse. Governance, *36*(3), 693–709. https://doi.org/10.1111/gove.12690

Love, P. (2021). Why Herbert Simon matters for policymaking. *Computational Economics*, *57*(3), 923–933. https://doi.org/10.1007/s10614-018-9800-2

Love, P., & Stockdale-Otárola, J. (Eds.). (2017). *Debate the issues: Complexity and policy making*. OECD. https://doi.org/10.1787/9789264271531-en

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Moerel, E. M. L., & Timmers, P. (2021). *Reflections on digital sovereignty. EU Cyber Direct, Research in Focus series 2021*, 1–33. https://ssrn.com/abstract=3772777

Müller, J., Tischer, M., Thumel, M., & Petschner, P. (2022). Unboxing digitale souveränität. Ein scoping review zu digitaler souveränität von individuen. *Medienimpulse*, *60*(4), 1–37. https://doi.org/10.21243/mi-04-22-19

Müller, M. (2023). *KfW-ifo skilled labour barometer December 2023* (KfW Research). https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-KfW-ifo-Fachkräftebarometer/KfW-ifo-Fachkraeftebarometer_2023-12_EN.pdf

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Oravec, J. A. (2019). Artificial intelligence, automation, and social welfare: Some ethical and historical perspectives on technological overstatement and hyperbole. *Ethics and Social Welfare*, *13*(1), 18–32. https://doi.org/10.1080/17496535.2018.1512142

Oswald, D. (2022). Cybernetics, operations research and information theory at the Ulm School of Design and its influence on Latin America. *AI and Society*, *37*(3), 1045–1057. https://doi.org/10.1007/s00146-021-01339-1

Pohle, J. (2020). Digital sovereignty – A new key concept of digital policy in Germany and Europe. Konrad Adenauer Stiftung. https://www.kas.de/de/einzeltitel/-/content/digitale-souveraenitaet

Pohle, J., & Thiel, T. (2020). Digital sovereignty. *Internet Policy Review*, *9*(4), 1–19. https://doi.org/10.14763/2020.4.1532

Pohle, J., & Thiel, T. (2021). Digitale Souveränität: Von der Karriere eines einenden und doch problematischen Konzepts. In *Digitale Gesellschaft* (Vol. 36, pp. 319–340). Transcript Verlag. https://doi.org/10.14361/9783839456590-014

Preiß, R., Zetti, D., & Herzog, C. (2023). Belonging as a relevant success factor for e-government? In *Proceedings of the STS Conference Graz 2023*, Graz, Germany, 2024, 200–220. https://doi.org/10.3217/978-3-85125-976-6-12

Reznich, C. (2017). Horst Rittel. Design Methodology / Wicked Problems Theory. *Medium*. https://medium.com/@creznich/1973-7921763949fd

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/BF01405730

Rosenberg, D. (2019). Use of e-government services in a deeply divided society: A test and an extension of the social inequality hypotheses. *New Media & Society*, *21*(2), 464–482. https://doi.org/10.1177/1461444818799632

Section, A. I. U. (2018). Trapped in the matrix – Secrecy, stigma, and bias in the Met's gangs database. Amnesty International United Kingdom Section.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, *4*(3–4), 181–201. https://doi.org/10.1016/0004-3702(73)90011-8

Solove, D. J. (2020). The myth of the privacy paradox. *SSRN Electronic Journal*, *89*(1), 1–51. https://doi.org/10.2139/ssrn.3536265

Twizeyimana, J. D., & Andersson, A. (2019). The public value of e-government – A literature review. *Government Information Quarterly*, *36*(2), 167–178. https://doi.org/10.1016/j.giq.2019.01.001

Vallor, S. (2016). *Technology and the virtues*. Oxford University Press.

van de Poel, I., & Royakkers, L. (2011). *Ethics, technology, and engineering: An introduction*. Wiley-Blackwell.

Wessel, D., Hanke, K., Preiß, R., Herzog, C., Wegener, C., Claudy, S., & Heine, M. (2023). *Improving citizens' digital sovereignty by supporting transparency in public administration*. 31st European Conference on Information Systems, Kristiansand, Norway.

Whitbeck, C. (2011). *Ethics in engineering practice and research* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511976339

Shaping AI (more) sustainably. Socio-technological perspectives on AI infrastructures

Anne Mollen and Sigrid Kannengießer

Abstract: In the field of Machine Learning, a growing body of research reveals and analyses the negative ecological effects of generative AI, especially its contribution to climate change via carbon dioxide emissions that result from generative AI training. Moreover, such analyses also consider partly how to reduce these effects. However, beyond carbon emissions, the field rarely acknowledges further ecological, economic and social consequences of the production and use of generative AI. In the social sciences and humanities, the socio-ecological-economic effects of generative AI and sustainable ways of shaping AI infrastructures thus far constitute only a minor research area. But the question of how generative AI infrastructures can be sustainably shaped demands a social science perspective. AI infrastructures include not only technological devices, such as servers, cables, data centres and consumer apparatuses on which generative AI is used. Also, all actors engaging in social practices that shape AI infrastructures must be considered. Focusing on the practices that shape AI infrastructures, therefore, can not only help to understand how these structures are currently shaped and why they cause tremendous socio-ecological-economic effects, it also has the potential to recognise attempts to shape AI infrastructures (more) sustainably and to conceptualise practices that would allow for sustainable production and use of generative AI. Therefore, using a sustainability perspective, our theoretical reflection develops a *socio-technological concept of AI infrastructures*. Firstly, it points to the socio-ecological effects of the production of AI infrastructures. Secondly, it also discusses how AI infrastructures can be shaped more sustainably. The underlying theoretical assumption is that AI infrastructures come into being through social practices that can be altered towards sustainability.

Keywords: artificial intelligence, sustainability, social practices, digital infrastructures, social sciences

1. Introduction

In the field of Machine Learning, a growing area of research reveals and analyses the negative ecological effects of generative AI, especially its contribution to climate change via carbon dioxide emissions that result from the training of the large language models (LLM) that form the basis for generative AI applications. Moreover, this research considers how these effects can be reduced.

However, the ecological, economic and social consequences of the production and use of generative AI, beyond carbon emissions, are rarely acknowledged in this field. In the social sciences and humanities, the socio-ecological-economic effects of generative AI, as well as ways of addressing these consequences and thereby shaping AI infrastructures more sustainably, thus far represent only a minor research area. But a sustainability perspective—one that understands the term in the tradition of the Brundtland report as social-ecological-economic justice (UN WCED, 1987) with a clear social dimension to it—becomes crucially important in shaping AI infrastructures. Thus, the question of how infrastructures of generative AI can be shaped in a sustainable way is one that also demands a social science perspective.

AI infrastructures include not only technological devices, such as servers, cables, data centres and consumer apparatuses on which generative AI is used, but also actors whose social practices shape these devices and AI infrastructures, in general. Therefore, focusing on the practices that shape AI infrastructures cannot only help to understand how these structures are currently shaped and why they cause tremendous socio-ecological-economic effects, it also has the potential to recognise attempts to shape AI infrastructures (more) sustainably and to conceptualise practices that would allow for producing and using generative AI sustainably. Therefore, the theoretical reflection presented here develops a **socio-technological concept of AI infrastructures** by using a sustainability perspective. Pointing to the socio-ecological effects of the production of AI infrastructures; it also discusses how AI infrastructures can be shaped in a more sustainable way. The underlying theoretical assumption is that AI infrastructures come into being through **social practices** that can be altered towards sustainability.

To develop this argument, we will first sketch the interdisciplinary research field of sustainable AI. Second, we will develop a socio-technological concept of AI infrastructures, stressing the relevance of all actors and practices shaping these structures. In doing so, we will use sociological practice theory, which has been used and further developed in media and communication studies. Third, we argue that a socio-technological perspective on AI infrastructures is necessary not only to understand generative AI, but especially so when applying a sustainability perspective to AI infrastructures. It allows an understanding of the socio-ecological-economic effects caused by the production and use of generative AI and the attempts to shape AI in a more sustainable way, as well as conceptualising practices that would allow for producing and using generative AI more sustainably.

2. The interdisciplinary research field of sustainable AI

Currently, two strands of discussion on sustainable AI can be differentiated (van Wynsberghe, 2021). One concerns the **use of automation**, especially Machine Learning technologies, to reach sustainability objectives. The other concerns the **socio-ecological-economic impacts of AI systems** themselves and the question of how to shape AI more sustainably. As part of the first discussion, researchers are, for instance, concerned with analysing the environmental benefits of applying AI systems to achieve the Sustainable Development Goals (Vinuesa et al., 2020), to save resources (Rolnick et al., 2022), to perform complex distributional tasks (Klobasa et al., 2019), as an element in environmental governance (Nishant et al., 2020), to give some examples. With a strong political tail wind (UN DESA, 2021; World Economic Forum, 2021), the discussion of the use of AI technologies for sustainability objectives seems to be much more well-established than the discussion on the sustainability impacts of AI technologies.

Nonetheless, in the field of Machine Learning, questions of sustainability are increasingly taking shape as a distinct field of inquiry (Kaack et al., 2022). However, it must be noted that environmental concerns, in particular, are addressed here. Strubell and colleagues (Strubbel et al. 2019) published a seminal analysis of the possibly immense emissions resulting from the development and training of so-called transformer LLMs, which currently form the basis for generative AI. Following this work, research interest in sustainable AI increased, focusing mainly on environmental impacts of LLMs such as BERT (Google, 2018), GPT-3 (OpenAI, 2020), LLAMA (Meta, 2023) or similar models.

Publications have emerged that attempt to understand and quantify the environmental impact of LLMs. In this regard, there have been publications on the energy usage of different LLMs (Chowdhery et al., 2022; Patterson et al., 2021), statistics related to water consumption (Li et al., 2023; Mytton, 2021), emissions calculated over the life cycle of LLMs, including in the production of hardware and inference/application (Luccioni, Viguier, et al., 2022), attempts to measure system-level impacts of AI-related emissions, especially in comparison with their potential sustainability benefits (Kaack et al., 2022) and so on.

Based on these numbers and attempts to understand the environmental impacts of AI systems, some publications have begun to address how to improve environmental impacts, for instance by choosing training times and locations based on environmental concerns (Dodge et al., 2022; Li et al., 2023) or minimising datasets to reduce computing time (Menghani, 2023). Learning about the environmental impacts of Machine Learning systems through this kind of research has led to the practical application of ecological sustainability-oriented approaches in Machine Learning development (Luccioni, Mueller et al., 2022).

Thus, sustainability of AI so far has primarily been understood based on Machine Learning heuristics. Van Wynsberghe (2021, p. 214), for instance, describes sustainable AI as 'a field of research that applies to the technology of AI (the hardware powering AI, the methods to train AI, and the actual processing of data by AI) and the application of AI while addressing issues of AI sustainability and/or sustainable development' over the life cycle of an AI.

Here, the emphasis primarily falls on technical components and production logics (e.g. computing hardware, data sets, Machine Learning algorithms, computing) and the role they play in sustainability questions. Although helpful as a structuring element, this approach has little analytical value for understanding processes and agency behind (un)sustainable AI; it thus lacks the foundations for profound, socially oriented analyses. Therefore, we argue for a **socio-technological perspective on the life cycle of generative AI**, one that acknowledges not only the technological components of the infrastructures of generative AI but also the actors and practices that shape these technologies.

3. Socio-technological perspectives on AI infrastructures

Although the relationship between AI infrastructures and sustainability has primarily been acknowledged as a technological matter within Machine Learning discourses, it is crucial to acknowledge how AI infrastructures are shaped. This is not only to understand the socio-technological-economic impacts of generative AI, but also attempts to shape AI (more) sustainably, as well as to conceptualise and call for practices that allow for more sustainable AI infrastructures.

In recent years, digital technologies and their infrastructures have become an established subject of study in communication and media research (Flensburg & Lai, 2020; Gillespie et al., 2014; Parks & Starosielski, 2015; Plantin & Punathambekar, 2019; Starosielski, 2014), engaging with perspectives of science and technologies studies on information systems, as well as information and communication technologies (Bowker et al., 2009; Star & Ruhleder, 1996) and social constructivist approaches to media technology and communicative infrastructures (Andersen, 2018; Christmann et al., 2022). Within **Human–Computer Interaction**, such socio-technical approaches have been reflected in the integration of human actors into the analysis of digital technologies.

Although definitions of digital infrastructures still stress the relevance of the backbone of the internet and understand infrastructures as a combination of consumer devices, network structures and server farms (Constantinides et al., 2018; Fox & Hao, 2017; Greenstein, 2020), defining infrastructures as a

'substrate' (i.e. something on which something else 'runs' or 'operates') has long been criticised (Star & Ruhleder, 1996, p. 112). Infrastructure is not a 'thing' that forms the basis for the internet; rather **infrastructure is a relational concept** that becomes real in and through practices (Star, 2002, p. 116).

These practices of 'infrastructuring' (Pipek & Wulf, 2009) are conducted by designers, builders, maintainers or users of information systems who shape infrastructures through their practices. A practice approach to infrastructures accordingly enables an analysis of how infrastructures come into being and what kinds of manifestations (e.g. regarding inequalities) they entail. This understanding applies also to infrastructures of generative AI that come into being through social practices.

Sociological practice theory defines social practices as:

> A routinised type of behaviour which consists of several elements, interconnected to one another: forms of bodily activities, forms of mental activities, 'things' and their use, a background knowledge in the form of understanding, know-how, states of emotion and motivational knowledge (Reckwitz, 2002, p. 249-250).

In contrast to habits, practices are patterns of actions that are based on knowledge and become routinised. Looking at AI systems, the question then is: which practices design, build and maintain AI systems? These practices can either be performed by human actors or executed by technologies, but looking at the question of who is responsible for shaping AI infrastructures more sustainably, human actors become more relevant as 'there can be no machines without humans to make them' (Jasanoff & Kim, 2015, p. 8).

In communication and media studies, practices have been subdivided into those in which actors are 'acting with media', meaning that media are used as mediators, and those in which actors 'act on media', meaning that media (as organisations, content and technologies) themselves are the focus of action (Kannengießer & Kubitschko, 2017). 'Acting on' media refers to practices in which 'people consciously and actively seek to transform [media] technologies and in doing so try to change not only the devices, but also society' (Kannengießer, 2020, p. 178). This conceptualisation, which was formally developed for media, can also be applied to other technologies, including AI, meaning that 'acting with AI' refers to those practices that use AI for whatever purpose (e.g. sustainability) and 'acting on AI' refers to those practices through which actors consciously and actively (try to) shape AI (e.g. shaping AI systems more sustainably). When examining the sustainability of AI, it is these practices of 'acting on AI' that become relevant.

4. Shaping AI infrastructures (more) sustainably

The proposed socio-technological perspective on AI infrastructures, based on a practice approach, thus allows for answering questions on how AI infrastructures can be shaped more sustainably. We can see such practices already emerging in the very different ways of researching, developing and using generative AI – partly reflecting the highly contested views that currently exist on the technology. Two fundamentally different positions are emerging in the discussions. On the one hand, there are those who perceive existential but thus far **hypothetical AI risks** as the greatest danger to humanity (**AI safety/alignment**). On the other hand, there are those who give top priority to **existing risks**, such as **bias and discrimination** (**AI ethics**), and criticise the focus on hypothetical risks as indifference to the detrimental effects AI has on groups of people who are often already marginalised (Sætra & Danaher, 2023).

Particularly with regard to these current risks and harms, we can see practices emerging that attempt to shape AI infrastructures in more sustainable ways (Stanusch et al. 2024). Although dominant generative AI players (e.g. OpenAI or Meta) align with the approach of building ever larger LLMs, other initiatives (e.g. Bloom, Eleuther AI and HuggingFace) have begun to address specific, sustainability-relevant issues: Examples include experimenting with **new modes of data management practices** as well as **output curation** to address issues of bias, discrimination and racism in AI outputs. Others try to reduce resource consumption while maintaining high model performance, as well as developing **more efficient hardware**. An important element across such attempts are matters **of openness and access**, which are intended to '**democratise**' generative AI development, as such initiatives often like to stress.

All these attempts to shape AI more sustainably, through corporations, technology experts, non-profits, policy actors and so on, have further implications relevant to matters of sustainability. Investing in more efficient hardware, for instance, might not lead to less resource consumption by generative AI systems, if these efficiency gains are again offset by LLMs that are ever-growing in number as well as in size. Furthermore, what defines non-biased datasets and outputs is equally debatable. Hence, we see negotiations, power dynamics and the normalisation of routines into practices within these attempts to create more sustainable AI systems, and thus a rich field for empirical analysis of the socio-technical emergence of AI infrastructures.

While some of the challenges regarding sustainable AI infrastructures are unique to generative AI, for example the size of data sets or the amount of compute and energy needed for creating generative AI, others have already been addressed in discourses on the sustainability of digital technologies in general. The results of research that analyse ways of shaping digital infrastructures in a more sustainable way thereby become relevant (Kannengießer, 2020), as generative AI is embedded in digital infrastructures that then integrate specific AI aspects (e.g. training with relevant data).

When looking at the actors trying to shape AI more sustainably, questions of **agency** become relevant.

Agency can be defined:

> As a temporally embedded process of social engagement, informed by the past (in its "iterational" or habitual aspect) but also oriented toward the future (as a "projective" capacity to imagine alternative possibilities) and toward the present (as a 'practical-evaluative' capacity to contextualise past habits and future projects within the contingencies of the moment) (Emirbayer & Mische, 1998).

Regarding AI and its sustainability, the question is: which actors have the capacity to shape current AI systems? This question presupposes that such actors possess knowledge about past systems and are able to imagine future possibilities, in other words that these actors have agency and have the power to either potentially shape AI systems more sustainably – or fail to do so. Reconstructing agency then helps to reveal the power structures in which AI-related actors are positioned. The question of agency directly relates to the question of practices. When applying a socio-technological perspective to the study of AI systems, it is important not only to identify actors, but also to consider the way they shape AI systems. In other words, their practices need to be an object of research. Here, we identify a research gap in the interdisciplinary academic discourse on sustainable AI.

Using practice theory for the analysis of sustainable AI helps us to move beyond understanding the sustainability of AI as an engineering problem. It allows us to integrate 'the social' into a perspective on sustainable AI. This means acknowledging the role of actors and their practices in shaping AI infrastructures – that is to say examining not only the role of AI designers, but also that of the actors involved in building and maintaining AI infrastructures, including networks, data centres, datasets, policies and, eventually, users of AI and the digital technologies that are needed for the everyday appropriation of AI in people's everyday usage.

5. Conclusion

In this article, we argue for understanding generative AI infrastructures as socio-technological phenomena that consist not only of technological aspects, but also come into being through diverse actors and their social practices. Using practice theory and investigating how actors 'act on' and thereby shape AI infrastructures allows for an understanding of the reasons underlying the socio-ecological-economic effects of generative AI. At the same time, it investigates attempts to shape AI (more) sustainably while conceptualising sustainability-oriented practices. Applying such a comprehensive concept of infrastructures makes it possible to integrate social and technological aspects of infrastructures in an empirical analysis and to analyse actors and practices involved in infrastructures in order to understand the relationship between sustainability and generative AI. At the same time, it also allows for stressing the need for shaping AI infrastructures more sustainably and identifying opportunities to do so.

References

Andersen, J. (2018). Archiving, ordering, and searching: Search engines, algorithms, databases, and deep mediatization. *Media, Culture & Society, 40*(8), 1135–1150. https://doi.org/10.1177/0163443718754652

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). Association for Computing Machinery*, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022). The values encoded in machine learning

research. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). Association for Computing Machinery*, New York, NY, USA, 173–184. https://doi.org/10.1145/3531146.3533083

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., . . .Fiedel, N. (2022). *Palm: Scaling language modeling with pathways.* arXiv. https://arxiv.org/abs/2204.02311

Christmann, G. B., Löw, M., & Knoblauch, H. (2022). *Communicative constructions and the refiguration of spaces: Theoretical approaches and empirical studies.* Routledge. https://doi.org/10.4324/9780367817183

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*(4), 962–1023. https://doi.org/10.1086/231294

Flensburg, S., & Lai, S. S. (2020). Mapping digital communication systems: Infrastructures, markets, and policies as regulatory forces. *Media, Culture & Society, 42*(5), 692–710. https://doi.org/10.1177/0163443719876533

Gillespie, T., Boczkowski, P. J., & Foot, K. A. (Eds.). (2014). *Media technologies: Essays on communication, materiality, and society*. The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.001.0001

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff, & S. -H. Kim (Eds.), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power* (pp. 1–33).

University of Chicago Press. https://doi.org/10.7208/9780226276663-001

Kaack, L. H., Donti, P. L., Strubell, E., Kamiya, G., Creutzig, F., & Rolnick, D. (2022). Aligning artificial intelligence with climate change mitigation. *Nature Climate Change, 12*(6), 518–527. https://doi.org/10.1038/s41558-022-01377-7

Klobasa, M., Plötz, P., Pelka, S., & Vogel, L. (2019). *Artificial intelligence for the integrated energy transition.*

Fraunhofer ISI. https://doi.org/10.24406/publica-fhg-300027

Kannengießer, S. (2020). Consumption-critical media practices: acting on media for sustainability. In H. Stephansen & E. Treré, E (Eds.), *The turn to practice in media research: implications for the study of citizen- and social- movement media* (pp. 176–188). Routledge.

Kannengießer, S., & Kubitschko, S. (2017). Acting on media: influencing, shaping and (re)configuring the fabric of everyday life. *Media and Communication, 5*(3), 1-4. https://doi.org/10.17645/mac.v5i3.1165

Li, P., Yang, J., Islam, M. A., & Ren, S. (2023). Making AI less "thirsty": Uncovering and addressing the secret

water footprint of AI models. *arXiv*. https://doi.org/10.48550/ARXIV.2304.03271

Luccioni, A. S., Mueller, Z., & Raw, N. (2022, April 22). CO2 emissions and the ● Hub: Leading the charge.HuggingFace. https://huggingface.co/blog/carbon-emissions-on-the-hub

Luccioni, A. S., Viguier, S., & Ligozat, A. -L. (2022). Estimating the carbon footprint of BLOOM, a 176B parameter language model. *arXiv*. https://arxiv.org/abs/2211.02001

Menghani, G. (2023). Efficient deep learning: A survey on making deep learning models smaller, faster, and better. ACM Comput. Surv. 55, 12, Article 259 (December 2023), 37 pages. https://doi.org/10.1145/3578938

Mytton, D. (2021). Data centre water consumption. *npj Clean Water, 4*(1), 11. https://doi.org/10.1038/s41545-021-00101-w

Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities,

and a research agenda. *International Journal of Information Management, 53.* https://doi.org/10.1016/j.ijinfomgt.2020.102104

Parks, L., & Starosielski, N. (Eds.). (2015). *Signal traffic: Critical studies of media infrastructures.* University of Illinois Press.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L. -M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon emissions and large neural network training. *arXiv*. https://arxiv.org/abs/2104.10350

Pipek, V., & Wulf, V. (2009). Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems, 10*(5), 306–332. http://dx.doi.org/10.17705/1jais.00195

Plantin, J. -C., & Punathambekar, A. (2019). Digital media infrastructures: Pipes, platforms, and politics. *Media, Culture & Society, 41*(2), 163–174. https://doi.org/10.1177/0163443718818376

Reckwitz, A. (2002). Toward a theory of social practices: A development in culturalist theorizing. *European Journal of Social Theory, 5*(2), 243–263. https://doi.org/10.1177/13684310222225432

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., & Waldman-Brown, A. (2022). Tackling climate change with machine learning. A*CM Computing Surveys (CSUR), 55*(2), 1–96.

Sætra, H. S., & Danaher, J. (2023). Resolving the battle of short- vs. long-term AI risks. In J. Maclntyre, & L. Medsker (Eds.), *AI and ethics.* Springer. https://doi.org/10.1007/s43681-023-00336-y

Shalf, J. (2020). The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 378*(2166), 20190061. https://doi.org/10.1098/rsta.2019.0061

Stanusch, N., Rogers, R., Baym, N., Liu, C., Shaw R., Katzenbach, C., Richter, V., Kannengießer S., Mollen, Glawatzki, A., A., Brause, S. R., Yang, H., Schäfer, M., Zeng, J. (2024). AI Industry Expectations and Underperforming Imaginaries. Panel presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Starosielski, N. (2014). The materiality of media heat. *International Journal of Communication*, *8*(2014), 2504–2508. https://ijoc.org/index.php/ijoc/article/view/3298/1268

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *arXiv preprint* arXiv:1906.02243. https://arxiv.org/abs/1906.02243

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, Automation, and News. *Digital Journalism, 7*(8), 980– 992. https://doi.org/10.1080/21670811.2019.1685395

UN DESA. (2021). *Artificial intelligence saving the natural world. United Nations Department of Economic and Social Affairs.* https://www.un.org/en/desa/artificial-intelligence-saving-natural-world

UN WCED. (1987). *Our common future.United Nations World Commission on Environment and Development.* https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics, 1*(3), 213–218. https://doi.org/10.1007/s43681-021-00043-6

Velkova, J. (2023). Retrofitting and ruining: Bunkered data centers in and out of time. *New Media & Society, 25*(2), 431–448. https://doi.org/10.1177/14614448221149946

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications, 11*(1), 1–10. https://doi.org/10.1038/s41467-019-14108-y

World Economic Forum. (2021). *Harnessing artificial intelligence to accelerate the energy transition.* https://www.dena.de/fileadmin/dena/Dokumente/Meldungen/Harnessing_Artificial_Intelligence/WEF_Harnessin g_AI_to_accelerate_the_Energy_Transition_2021.pdf

‘Encryption doesn’t matter’: Pitfalls in cybersecurity communications

Florian Meissner, Jan Magnus Nold, Martina Angela Sasse, Rebecca Panskus, and Alexander Wilke

Abstract: Due to digitalisation, cybersecurity is becoming increasingly important for citizens. In Germany, the Federal Office for Information Security (BSI) is the authority responsible for ‘digital consumer protection’. Its aim is to use social media to communicate with the public about cybersecurity. Precisely, this area is an uncharted scientific territory. Theoretical approaches such as **protection motivation theory** (PMT) and **framing** provide useful guidelines for effective communication on protective behaviours. Our study explores the basic characteristics of BSI’s social media communication and analyses to what degree BSI’s posts on Twitter (X) published in 2021 and 2022 correspond with these guiding principles. Based on a computational analysis of n = 3,058 tweets and a qualitative in-depth analysis of the most prominent n = 34 tweets, the results show that BSI’s social media communication is often self-referential and discusses current events related to digital security only to a limited degree. When mentioned, cyber threats and countermeasures are typically presented in a vague manner. Similarly, it is often unclear who might be (potentially) affected by a threat. We conclude that applying a model for designing risk messages that draw on the dimensions of PMT could help cybersecurity-related social media communication.

Keywords: cybersecurity, cybersecurity communication, risk communication, protection motivation theory, framing

1. What are the objectives of communication on cybersecurity?

With the increasing number of digital devices and services, there is value to be had – not least for attackers. This growth has a significant impact on the daily lives of people and businesses, who increasingly need to know about the cybersecurity risks and steps they should take to protect themselves.

Cyberattacks, data leaks, and other security incidents can have serious consequences, ranging from financial losses and a loss of trust in digital systems to the impairment of critical infrastructures. Public communication plays an important role in raising awareness of the risks and available protection measures. The goal is not only to provide details of technical aspects but also to sensitise humans who use digital systems daily. However, awareness-raising measures are not enough to enable secure behaviour.

Research on human behaviour in cybersecurity has concluded that many of the prerequisites for the adoption of secure behaviours are not in place (Sasse et al., 2022). Many humans do not fully understand the risks associated with the use of digital devices and services. They are sometimes unaware of relevant countermeasures or doubt they are effective (Dechand et al., 2019). Such misunderstandings are partly based on media reports that misrepresent the technical reality (Fulton et al., 2019). Furthermore, humans are often overwhelmed by the many different – and sometimes contradictory – pieces of information about cybersecurity (Reeder et al., 2017). At the same time, cybersecurity risks should not be exaggerated. Florencio et al. (2014) identify the negative impact of exaggerated risks by cybersecurity vendors on businesses, and Menges et al. (2023) observe that trainees become dejected and passive after being exposed to worst-case online risks.

When it comes to communicating about IT security, national security agencies need to take a leading role. Effective public communication should create awareness of relevant risks and offer citizens concrete instructions for action to protect themselves against them. To achieve this, communication measures must reach and appeal to the target group and be understandable as well as practicable for them. In Germany, the Federal Office for Information Security (BSI) is the central authority for cybersecurity, which has the task of informing and sensitising citizens (as well as cybersecurity experts and industry professionals). The present work aims to gain insights into what and how the BSI communicates and how effective this is in helping citizens keep safe.

We use the BSI's public communication via Twitter (today X) to examine whether and how this communication can be improved. First, postings by the BSI on Twitter over a period of 20 months were analysed quantitatively. In the second step, the messages with the greatest reach were analysed qualitatively to see if they follow scientific principles for effective communication and are thus likely to be effective. Based on our results, we derive recommendations for more effective public communication on cybersecurity.

2. State of research on public cybersecurity communication

Communication science in general, and risk and crisis communication research in particular, has so far rarely dealt with cybersecurity. The research that has been carried out reveals an increase in the media coverage of cybersecurity (Alagheband et al., 2020; Boholm, 2021; Buse & Meissner, 2023), which is often driven by key events that have increased public awareness, such as the Snowden revelations or the Cambridge Analytica case. Regarding the use of different media channels for information on cybersecurity, online media have become the primary source of information on cybersecurity for a large proportion of the population (Das et al., 2018). However, differences in demographic characteristics, affinity for technology, and gender must be considered (Herbert et al., 2022).

Another research interest is the analysis of reporting on specific incidents in the context of cybersecurity (Griesbacher & Griesbacher, 2020), as well as the portrayal of the actors, especially hackers, in the media (Buse & Meissner, 2019). Here, we examine how these actors are presented by the media and what effects these representations could have on the formation of public opinion in the field of cybersecurity.

In a prior study, the communication of cybersecurity issues in social media was examined by Vogler and Meissner (2020), who find that people affected by a data leak at a large ticketing provider communicated on Twitter primarily about service aspects and not about security aspects, which could indicate a lower prioritisation of the topic of data security. Bada et al. (2019) uncovers no evidence for the effectiveness of awareness campaigns developed to promote cybersecurity. Reasons for this are that the campaigns depend too heavily on fear-based appeals or are not adapted to the cultural circumstances of their target groups (Sasse et al., 2022). So far, however, there has been a lack of scientific studies analysing the factors for the success or failure of such campaigns to develop suggestions for an improved communication strategy in cybersecurity.

3. Theoretical background: Basic principles for effective security communication

On a theoretical level, a risk and crisis communication perspective offers suitable and instructive approaches that can be applied to communication about cybersecurity. These were developed to support the self-protective behaviour of people who are exposed to certain risks or emergency situations. The present case study is based on the **protection motivation theory** (PMT) approach by Floyd et al. (2000) and the **framing** approach by Entman (1993). Initially proposed by Rogers (1983), PMT posits that messages highlighting personal threats and countermeasures can prompt protective behaviour (Floyd et al., 2000). This theory is particularly relevant when users require added incentives for secure behaviour (Boss et al., 2015). PMT suggests two prerequisites for protective actions: **threat** and **coping appraisals**, each comprising various dimensions.

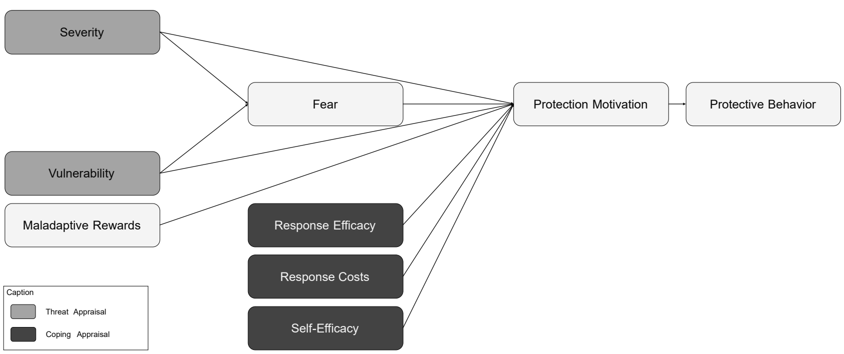

Figure 1. PMT Model, own illustration based on Boss et al. (2015, p. 854)

The threat appraisal involves assessing threat severity and personal vulnerability, with greater weight given to fear over maladaptive rewards (reward for the reaction of not protecting oneself, such as time or cost savings) to stimulate motivation for protection. Coping appraisals include perceived efficacy, self-efficacy, and the response costs of protective measures (Boss et al., 2015). The knowledge that these are key factors in convincing people to implement a certain security behaviour can also be applied to design security communication, for example, in cybersecurity. Prior PMT research highlights self-efficacy's pivotal role in influencing behaviour (Branley-Bell et al., 2022; Feltz & Öncü, 2014; Mwagwabi & Hee Jiow, 2021; Sasse et al., 2023).

This study also includes the concept of framing, which is widely used in communication science. According to Entman (1993), so-called frames – patterns of interpretation – consist of four elements: a problem description, a causal interpretation, a moral evaluation, and a recommendation for action. For the context of this study, the categories of problem description (here in the sense of a cyber threat) and recommended action (here in the sense of a specific recommended protective behaviour) are particularly relevant. We propose that both approaches – PMT and framing – can be applied to cybersecurity communication.

4. Methodology of data collection and analysis

This case study conducts a two-stage analysis of BSI's Twitter communication on cybersecurity, aiming to assess risk communication practices and identify areas for improvement. BSI's Twitter activity serves as a representative sample of its broader online presence, including platforms like Facebook and LinkedIn, where similar content is shared.

The data for this analysis was extracted from the overall timeline of BSI's Twitter account and retrieved via the Academic Twitter API shortly before it was closed in spring 2023. The period under investigation was 21 April 2021 (two days before a new IT security law was passed in Germany) to 31 December 2022. In total, we analysed 3,058 tweets, looking at the frequency of hashtags, @-mentions, tweet frequency and the most successful (most liked) tweets. The analysis was carried out using RStudio. In addition to the BSI dataset, datasets entailing all tweets in the same time frame by the ethical hacker group Chaos Computer Club (1,145 tweets) and netzpolitik.org (2,854 tweets), a journalistic website on digital policy, were used for comparative purposes.

Then, the study conducted a qualitative analysis of BSI tweets with over 100 likes, a pragmatically set threshold indicating public attention. A category system, aligned with PMT and including problem definition and recommended action, was developed. Following Boss et al. (2015), all PMT dimensions were operationalised and coded as 'High', 'Ambivalent', 'Low', or 'Not available'. Response costs and maladaptive rewards could also be labelled 'Neutral'. Severity and vulnerability were coded only in the presence of a threat, while other PMT elements were coded alongside action recommendations. Framing categories, inspired by Entman (1993), were established through material review. Threats, action recommendations, and responsibilities were then coded. These categories were utilised in a previous study by Meissner et al. (2024) on cybersecurity reporting in German media, facilitating theoretical comparison. Additionally, categories covering comprehensibility, prominence, target group, and unambiguity were included.

5. Findings of the analysis

a) Quantitative findings

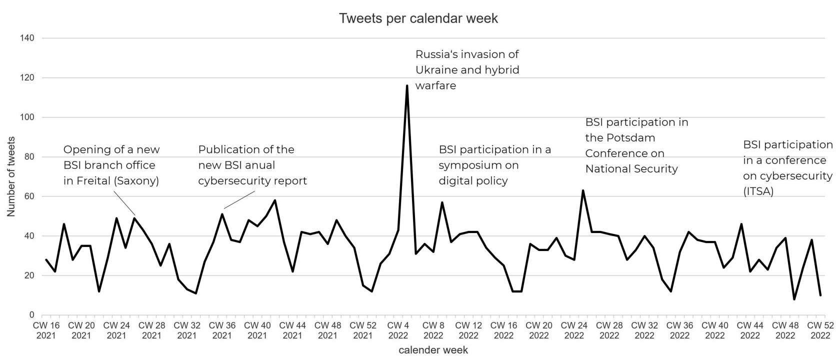

Figure 2. Time course of the number of BSI tweets per week, own illustration

First, the activity of the BSI account was examined over time (Figure 2). On average, the BSI publishes around 35 tweets per week, including individual tweets within threads. Overall, the weekly number of tweets fluctuates to a limited extent. A single clear outlier can be seen in February 2022, which is due to the Russian invasion of Ukraine and the associated hybrid warfare. Smaller peaks can all be attributed to the BSI's organisational communication. This indicates that, at least in quantitative terms, the BSI's communication is less geared towards topical media issues and instead focuses primarily on self-referential messaging.

Table 1. BSI hashtags without #deutschlanddigitalsicherbsi, own illustration

Used hashtags (#)	Amount
#deutschlanddigitalsicherbsi	1203
#bsikongress2022	92
#ransomware	46
#teambsi	41
#bsilagebericht2021	36
#cybersecurity	32
#ki	32
#einfachabsichern	31
#cybersicherheit	30
#coronawarnapp	29
…	

Table 2. Mentions without @BSI Bund and @ArneSchoenbohm, own illustration

Used mentions (@)	Amount
@cyberallianz	128
@bmi_bund	63
@certbund	58
@nancyfaeser	29
@cio_bund	28
@zivilehelden	27
@coronawarnapp	26
@bsi	23
@bka	22
@lilithwittmann	20
…	

The analysis of the most frequent hashtags proved to be less informative in terms of the cybersecurity threats covered. By far, the most frequently used hashtag was '#deutschlanddigitalsicherbsi', followed by further BSI-owned hashtags such as '#bsikongress2022' or '#teambsi'. In addition, there are generic hashtags such as '#cybersecurity' or '#itgrundschutz'; only a few specific topics such as '#ransomware' or '#coronawarnapp' are reflected here (Table 1).

In contrast to non-governmental accounts like Chaos Computer Club (a group of ethical hackers) or netzpolitik.org (a journalistic website on digital policy), analysed for comparison, the BSI account appears not to engage with government surveillance, as evident in hashtag analysis. This omission could be attributed to its status as a state authority under the Ministry of the Interior. The analysis of the most frequent @-mentions shows that almost exclusively official accounts of ministries and politicians are addressed (the accounts of the BSI itself and former President Arne Schönbohm were excluded here because they are replies to their own tweets or routine mutual references). A frequently mentioned account is @cyberallianz, which is a BSI-led initiative designed to communicate and promote dialogue with the corporate sector.

Taken together, both the hashtag analysis and the analysis of @-mentions indicate a self-referential communication approach regarding topics and accounts of state institutions and representatives, while topical issues that could be more appealing to the public are addressed only to a limited degree.

b) Qualitative findings

In a second step, 34 tweets (the first ten are documented in the appendix below) were filtered out of the total data set (N = 3058) that had at least 100 likes, which we pragmatically considered as a threshold for a minimum amount of public attention. We decided to focus only on likes because, in terms of numbers, they were the most important engagement metric and were also highly correlated with retweets and replies. These tweets were, in a first step, categorised on an inductive basis. Accordingly, the tweets can be roughly divided into three main categories:

1. Tweets that focus on a specific warning about a security threat or another current topic, such as Kaspersky antivirus (Figure 10), Log4j (Figure 6; Figure 7), Ukraine-related phishing waves (Figure 9; Fig 11), a Covid-19 warning app (Figure 3), Wi-Fi vulnerability (Figure 4);
2. Tweets that focus on humorous content with a (sometimes informative) IT reference (Figure 8; Figure 12);
3. Tweets that provide information on cybersecurity topics without any recognisable topical reference, like the explanation of Trojans (Figure 5).

In this context, it has to be noted that none of the self-referential postings (see above) is among the most popular tweets.

In a second step, the tweets were analysed qualitatively based on the categories specified in section 4 (**PMT, framing approach**). A key result is that when it comes to specific threats or security measures, the BSI avoids addressing those affected directly. Quotes such as 'companies and organizations should ...' (Figure 6) or 'be vigilant ...' (Figure 9) illustrate that the perceived personal vulnerability remains unclear. Therefore, there is only a very abstract perception that a threat is relevant to the individual or the organisation. However, other elements of the PMT are also neglected, which could represent an obstacle to protective behaviour in cybersecurity. Furthermore, the severity of the threat is rarely stated or remains unclear. Only in one tweet was the threat associated with a 'warning level red' (Figure 6), which indicates a high threat. Although suitable security measures were mentioned occasionally within the tweets, these also remained relatively unclear with instructions like 'replace products' (Figure 10) or 'be vigilant and stay informed' (Figure 9). Only one tweet on phishing in the context of the Russian invasion of Ukraine clearly stated, 'In such cases, don't transfer anything and find out about reputable aid organizations' (Figure 11).

In some cases, the measures were only to be found in a linked press release (Figure 6; Figure 7). However, no information was provided on the effectiveness of the security measures described. The self-efficacy related to the measures (meaning the degree to which they can be implemented effectively by ordinary humans) was also not presented directly. Moreover, the countermeasures that were presented suggest a high level of effort. Vague language such as 'substitute products' (Figure 10), for example, implies that the user first has to make an effort to find out what can be used to replace an existing product.

With reference to framing theory (Entman, 1993), a problem definition was generally given, but only in the further course of the Twitter thread (Figure 11) or in a linked press release (Figure 10). Although the recommended measures were mentioned, they were mainly vague (Figure 4; Figure 10). The analysis also revealed that the comprehensibility of most recommendations for action was only medium to low. In some cases, instructions were only to be found in the press release (Figure 11). Moreover, the target group remained unclear in several cases (Figure 6, Figure 10).

Finally, the qualitative analysis revealed two tweets with the potential for serious misunderstandings or promoting unsecure behaviour. For instance, the mention of replacing 'such products' without guidance on alternatives or the importance of keeping antivirus protection until a replacement is available (Figure 10) may lead users to deactivate Kaspersky antivirus without an immediate replacement, increasing their vulnerability. Another potential misunderstanding concerns the case of WLAN vulnerability (Figure 4). After a prominently placed 'Attention', it was stated that security researchers had found a vulnerability that could affect all WLAN devices. Before a link to a press release followed, the tweet read: 'Encryption doesn't matter', indicating that multiple WLAN standards were impacted. This statement can be misunderstood in the sense that encryption generally offers no protection, reinforcing an existing misconception among many humans (Abu-Salma et al., 2017).

6. Conclusions

The results show that there is room to improve the communication of the BSI on Twitter (now X), both in terms of topicality (to increase outreach) and message design (to support protective behaviour).

Regarding the quantitative results, BSI's communication on Twitter is – to a large degree – self-referential as it revolves around state institutions and representatives and could benefit from focusing more on current topics discussed in journalistic and/or social media to achieve greater relevance among the general public. It became apparent that certain topics, such as state surveillance, are avoided as the BSI is an authority subordinate to the Ministry of Interior.

Looking at the qualitative results, there is usually a threat (problem definition), and in some cases, countermeasures (recommended actions) are mentioned, although the latter often remain vague (including the target group of the recommendation). However, the central elements of the PMT were addressed only to a limited degree. Neither severity, vulnerability, efficacy, nor self-efficacy were clearly presented. Instead, the communication was often unclear and vague.

Against this backdrop, we recommend including a clear mention of target groups and their respective vulnerability in the first sentence. Doing so makes it easier for recipients to understand instantly whether the information given is relevant to them. Also, the severity of the threats should be presented in a more understandable way. Special attention should be paid to self-efficacy in communication as various studies have highlighted the importance of giving an explanation of how individuals can implement secure behaviour (Abroms & Maibach, 2008; Anker et al., 2016; Feltz & Öncü, 2014). Finally, experts should examine tweets for possible misunderstandings before publication to prevent incorrect conclusions and unsecure behaviour. Future work planned for this study will include an assessment of the tweets by cybersecurity awareness experts, who will assess both the correctness and suitability of the tweets. The results will provide further implications for BSI's social media communication.

7. Limitations and outlook

This study has several limitations. First, the study so far focuses only on Twitter (X). Secondly, the qualitative analysis is based on a relatively limited number of (the most popular) tweets that have been analysed in depth. Thirdly, even though other studies have already shown a positive effect of social media on behaviour (Ghahramani et al., 2022; Laranjo et al., 2015; Scholtz et al., 2016), the actual effect of cybersecurity-related messages on target groups – beyond predictive statements based on theory – still needs to be investigated. In essence, it is reasonable to assume that the federal German cybersecurity authority wields influence. Despite Twitter's relatively limited reach among citizens, it is plausible that political and journalistic stakeholders will carry BSI's messages to other channels. Therefore, further research will be needed to address these desiderata.

References

Abroms, L. C., & Maibach, E. W. (2008). The effectiveness of mass communication to change public behavior. *Annual Review of Public Health*, *29*(1), 219–234. https://doi.org/10.1146/annurev.publhealth.29.020907.090824

Abu-Salma, R., Sasse, M. A., Bonneau, J., Danilova, A., Naiakshina, A., & Smith, M. (2017). *Obstacles to the adoption of secure communication tools*. 2017 IEEE Symposium on Security and Privacy (SP), 137–153. https://doi.org/10.1109/SP.2017.65

Alagheband, M. R., Mashatan, A., & Zihayat, M. (2020). Time-based gap analysis of cybersecurity trends in academic and digital media. *ACM Transactions on Management Information Systems*, *11*(4), 1–20. https://doi.org/10.1145/3389684

Anker, A. E., Feeley, T. H., McCracken, B., & Lagoe, C. A. (2016). Measuring the effectiveness of mass-mediated health campaigns through meta-analysis. *Journal of Health Communication*, *21*(4), 439–456. https://doi.org/10.1080/10810730.2015.1095820

Bada, M., Sasse, A. M., & Nurse, J. R. C. (2019). Cyber security awareness campaigns: Why do they fail to change behaviour? https://doi.org/10.48550/ARXIV.1901.02672

Boholm, M. (2021). Twenty-five years of cyber threats in the news: A study of Swedish newspaper coverage (1995–2019). *Journal of Cybersecurity*, *7*(1). https://doi.org/10.1093/cybsec/tyab016

Boss, S., Galletta, D., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly*, *39*(4), 837–864. https://doi.org/10.25300/MISQ/2015/39.4.5

Branley-Bell, D., Coventry, L., Dixon, M., Joinson, A., & Briggs, P. (2022). Exploring age and gender differences in ICT cybersecurity behaviour. *Human Behavior and Emerging Technologies*, 1–10. https://doi.org/10.1155/2022/2693080

Buse, C., & Meissner, F. (2019). Much ado about hacking? How news media in Germany, the United Kingdom, and the United States report cyber threats. In N. N. (Ed.), GigaNet Annual Symposium. https://www.giga-net.org/2019symposiumPapers/26_Buse_Meissner_Much-Ado-About-Hacking.pdf

Buse, C., & Meissner, F. (2023, 20.05). 'So bleiben Sie sicher im Cyberspace ...'. Die Darstellung von Cybersicherheit in deutschen Online-Medien. 68. Jahrestagung der Deutschen Gesellschaft für Publizistikund Kommunikationswissenschaft (DGPuK), Bremen, Deutschland.

Crossler, R., & Bélanger, F. (2014). An extended perspective on individual security behaviors. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *45*(4), 51–71. https://doi.org/10.1145/2691517.2691521

Das, S., Lo, J., Dabbish, L., & Hong, J. I. (2018). Breaking! A typology of security and privacy news and how it's shared. In R. L. Mandryk, M. Hancock, M. Perry, & A. L. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3173574.3173575

Dechand, S., Naiakshina, A., Danilova, A., & Smith, M. (2019). In encryption we don't trust: The effect of end-to-end encryption to the masses on user perception. *2019 IEEE European Symposium on Security and Privacy (EuroS&P)* (pp. 401–415). https://doi.org/10.1109/EuroSP.2019.00037

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58.

Feltz, D. L., & Öncü, E. (2014). Self-confidence and self-efficacy. In A. G. Papaioannou & D. Hackfort (Eds.), *Routledge companion to sport and exercise psychology: Global perspectives and fundamental concepts* (pp. 417–429). Taylor & Francis. https://books.google.de/books?id=_zYsAwAAQBAJ

Florencio, D., Herley, C., & Shostack, A. (2014). FUD: A plea for intolerance. *Communications of the ACM*, *57*(6), 31–33. https://doi.org/10.1145/2602323

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, *30*(2), 407–429. https://doi.org/10.1111/j.1559-1816.2000.tb02323.x

Fulton, K. R., Gelles, R., McKay, A., Roberts, R., Abdi, Y., & Mazurek, M. L. (2019). *The effect of entertainment media on mental models of computer security*. USENIX Symposium on Usable Privacy and Security.

Ghahramani, A., De Courten, M., & Prokofieva, M. (2022). The potential of social media in health promotion beyond creating awareness: An integrative review. *BMC Public Health*, *22*(1), 2402. https://doi.org/10.1186/s12889-022-14885-0

Griesbacher, E.-M., & Griesbacher, M. (2020). Cybersecurity im medialen Diskurs. HMD Praxis *Der Wirtschaftsinformatik*, *57*(3), 584–596. https://doi.org/10.1365/s40702-020-00618-7

Herbert, F., Becker, S., Schaewitz, L., Hielscher, J., Kowalewski, M., Sasse, M. A., Acar, Y., & Dürmuth, M. (2022). A World Full of Privacy and Security (Mis)conceptions? Findings of a Representative Survey in 12 Countries. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2212.10382

Laranjo, L., Arguel, A., Neves, A. L., Gallagher, A. M., Kaplan, R., Mortimer, N., Mendes, G. A., & Lau, A. Y. S. (2015). The influence of social networking sites on health behavior change: A systematic review and meta-analysis. *Journal of the American Medical Informatics Association*, *22*(1), 243–256. https://doi.org/10.1136/amiajnl-2014-002841

Meissner, F., Buse, C., & Wilke, A. J. (2024, August 1). Mediated perspectives on cyber risk: A content analysis of news reporting about cyber threats and safety measures. *International Crisis and Risk Communication Conference Proceedings*. International Crisis and Risk Communication Conference. https://doi.org/10.69931/YQNA8053

Menges, U., Hielscher, J., Kocksch, L., Kluge, A., & Sasse, M. A. (2023). Caring not scaring – An evaluation of a workshop to train apprentices as security champions. *Proceedings of the 2023 European Symposium on Usable Security* (pp. 237–252). https://doi.org/10.1145/3617072.3617099

Mwagwabi, F., & Hee Jiow, J. (2021). Compliance with security guidelines in teenagers: The conflicting role of peer influence and personal norms. *Australasian Journal of Information Systems*, *25*. https://doi.org/10.3127/ajis.v25i0.2953

Reeder, R. W., Ion, I., & Consolvo, S. (2017). 152 simple steps to stay safe online: Security advice for non-tech-savvy users. *IEEE Security & Privacy*, *15*(5), 55–64. https://doi.org/10.1109/MSP.2017.3681050

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology: A sourcebook* (pp. 153–176). Guildford Press.

Sasse, M. A., Hielscher, J., Friedauer, J., Menges, U., & Peiffer, M. (2022). 'Warum IT-Sicherheit in Organisationen einen Neustart braucht', in Cyber-Sicherheit ist Chefinnenund Chefsache!, online, Feb. 2022: https://www.researchgate.net/publication/358277373_Warum_IT-Sicherheit_in_Organisationen_einen_Neustart_braucht

Sasse, M. A., Hielscher, J., Friedauer, J., & Buckmann, A. (2023). Rebooting IT security awareness – How organisations can encourage and sustain secure behaviours. In S. Katsikas, F. Cuppens, C. Kalloniatis, J. Mylopoulos, F. Pallas, J. Pohle, M. A. Sasse, H. Abie, S. Ranise, L. Verderame, E. Cambiaso, J. Maestre Vidal, M. A. Sotelo Monge, M. Albanese, B. Katt, S. Pirbhulal, & A. Shukla (Eds.), *Computer security*. ESORICS 2022 International Workshops (pp. 248–265). Springer International Publishing.

Scholtz, B., Burger, C., & Zita, M. (2016). A social media environmental awareness campaign to promote sustainable practices in educational environments. In J. Marx Gomez, M. Sonnenschein, U. Vogel, A. Winter, B. Rapp, & N. Giesen (Eds.), *Advances and new trends in environmental and energy informatics* (pp. 355–369). Springer International Publishing. https://doi.org/10.1007/978-3-319-23455-7_19

Vogler, D., & Meissner, F. (2020). How users tweet about a cyber attack: An explorative study using machine learning and social network analysis. *Journal of Digital Media & Policy*, *11*(2), 195–214. https://doi.org/10.1386/jdmp_00016_1

Appendix: List of analysed tweets


BSI
@BSI_Bund

Dies ist ein Thread zu den zahlreichen Anfragen rund um die #LucaApp.
▶️Wir schätzen das Angriffs-Szenario einer Code-Injection über das Luca-System abhängig von der konkreten Einsatzumgebung als plausibel ein. (1/5)

10:31 vorm. · 28. Mai 2021

229 Reposts **54** Zitate **449** „Gefällt mir“-Angaben **16** Lesezeichen

Figure A1. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021a. ‘This is a thread about the numerous requests around the #LucaApp. ▶ We estimate the attack scenario of a code injection via the Luca system to be plausible depending on the specific deployment environment. (1/5)’. Twitter, 28 May 2021. https://twitter.com/BSI_Bund/status/1398195272400920578


BSI
@BSI_Bund

📢 Achtung ❗ Sicherheitsforschende haben herstellerübergreifende WLAN-Schwachstellen veröffentlicht, die fast alle WLAN-Geräte betreffen können. Verschlüsselungstechnik spielt keine Rolle. Mehr: ➡️ bsi.bund.de/SharedDocs/Cyb... #FragAttacks #DeutschlandDigitalSicherBSI

Neue WLAN-Schwachstellen
Bewertung des BSI
- aktuell ist nur von lokaler Ausnutzung auszugehen, d.h. Angreifer muss in Reichweite des WLAN-Geräts sein
- aber die mögliche Betroffenheit vieler bzw. aller WLAN-Geräte erhöht das Risiko
- Cyber-Kriminelle könnten auch an Infos gelangen, die weitere Attacken im jeweiligen Netz bzw. ggf. auch aus der Ferne ermöglichen
- Tipp: Hersteller-Webseiten checken und bereitgestellte Patches zeitnah installieren

Bundesamt für Sicherheit in der Informationstechnik

11:20 vorm. · 12. Mai 2021

195 Reposts **13** Zitate **207** „Gefällt mir“-Angaben **13** Lesezeichen

Figure A2. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021b. ‘📢 Attention ❗ Security researchers have published manufacturer-independent Wi-Fi vulnerabilities that can affect almost all Wi-Fi devices. Encryption technology does not play a role. Read more: […]’. Twitter, 12 May 2021. https://twitter.com/BSI_Bund/status/1392409331212210179

Figure A3. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021c. 'Disguised malware can infect your devices unnoticed and cause further damage. More information here: 👉 https://bsi.bund.de/DE/Themen/Verbr...'. Twitter, 10 June 2021. https://twitter.com/BSI_Bund/status/1402947681354387459

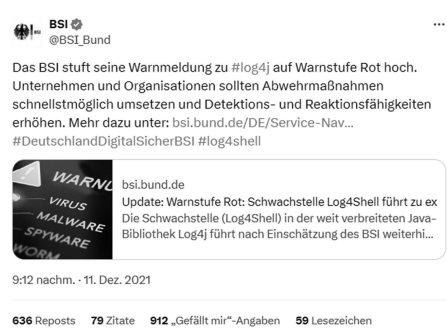

Figure A4. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021d. 'The BSI has upgraded its warning message for #log4j to warning level red. Companies and organisations should implement defensive measures as quickly as possible and increase their detection and response capabilities. Find out more at: [...]'. Twitter, 11 December 2021. https://twitter.com/BSI_Bund/status/1469761986313564167

Figure A5. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021e. 'Update on #log4j warning: compromises are currently being made by crypto miners & botnets. Other forms of attack are likely. Further help regarding detection and (incomplete) list of affected products: […]'. Twitter, 12 December 2021. https://twitter.com/BSI_Bund/status/1470036192112660485

Figure A6. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2021f. 'To everyone who is helping their loved ones set up and use digital devices over the festive period: Thank you 🙂 For your patience, but also for setting up and explaining the most important IT security settings! 😉 […]'. Twitter, 25 December 2021. https://twitter.com/bsi_bund/status/1474666271438479369

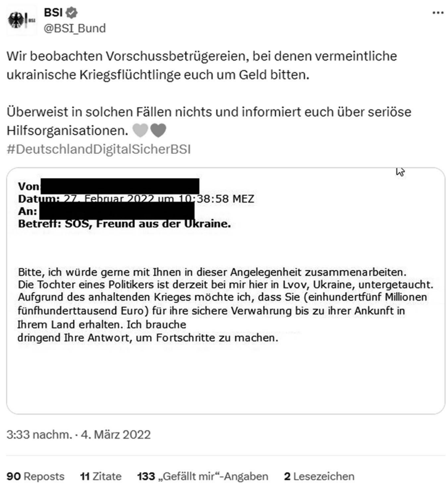

Figure A7. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2022a. ‘We are observing advance payment scams in which alleged Ukrainian war refugees ask you for money. Do not transfer anything in such cases and inform yourselves about serious help organisations. […]’. Twitter, 4 March 2022. https://twitter.com/BSI_Bund/status/1499754995545165829

Figure A8. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2022b. ‘In accordance with §7 of the BSI-Law, we warn against the use of virus protection software from the Russian manufacturer Kaspersky. We recommend replacing such applications with products of other manufacturers. To press release: […]’. Twitter, 15 March 2022. https://twitter.com/BSI_Bund/status/1503643699816845314


BSI
@BSI_Bund

Vorsicht vor neuem #Phishing-Versuch ❗
Aktuell nutzen Kriminelle "die durch den Ukraine-Krieg entstandene Energiekosten-Explosion" aus. Mails mit Betreffzeilen wie "Jetzt Energiepauschale sichern!/Wir überweisen die Energiepauschale/Bereit für Ihren Energiebonus?" sind Phishing.

Sparkasse

Liebe Kundin, lieber Kunde!

Um die Auswirkungen der gestiegenen Energiepreise für die Verbraucher abzumildern, wird im September ein Pauschalbetrag von 300 Euro an alle Erwerbstätigen ausbezahlt. Dies ist ein Beschluss der Bundesregierung und Inhalt des Entlastungspakets 2022, welches die durch den Ukraine-Krieg entstandene Energiekosten-Explosion etwas abfedern soll.

Wer erhält die Energiepauschale?

- **Steuerpflichtige** mit Einkünften aus Gewinneinkunftsarten (§ 13, § 15 oder § 18 des Einkommensteuergesetzes) und
- **Arbeitnehmerinnen und Arbeitnehmer**, die Arbeitslohn aus einem gegenwärtigen Dienstverhältnis beziehen und in die Steuerklassen I bis V eingereiht sind oder als **geringfügig Beschäftigte** pauschal besteuert werden.

Um Ihre Identät sowie den Anspruch auf eine Auszahlung feststellen zu können, benötigen wir eine Bestätigung Ihrer bereits angegeben Daten bei der Erstellung Ihres Girokontos in einer unserer Filialen.

Geben Sie noch heute Ihre aktuellen Daten auf unserer Homapege an und erhalten Sie innerhalb der nächsten vier Wochen Ihre Auszahlung der Energiepauschale. Dies können Sie ganz bequem von zu Hause aus erledigen, anbei finden Sie einen Direktlink zu den geforderten Angaben.

Vielen Dank für Ihre Zusammenarbeit!

Zur Homepage

Mit freundlichen Grüßen

Ihre Kundenberatung!

3:38 nachm. · 7. Sep. 2022

373 Reposts 24 Zitate 474 „Gefällt mir"-Angaben 11 Lesezeichen

Figure A9. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2022c. 'Beware of a new #phishing attempt ❗ Criminals are currently exploiting 'the explosion in energy costs caused by the war in Ukraine'. Emails with subject lines such as 'Secure your energy flat rate now!/We will transfer the energy flat rate/Ready for your energy bonus?" are phishing'. Twitter, 7 September 2022. https://twitter.com/BSI_Bund/status/1567507581496688641

Figure A10. Bundesamt für Sicherheit und Informationstechnik (@BSI_Bund). 2022d. 'We really couldn't resist this one for today's 'Day of Word Games'! 😅 And with that, we wish you all a great #weekend! 😍🤗💪 […]'. Twitter, 12 November 2022. https://twitter.com/BSI_Bund/status/1591375268710522880

The user experience of data donation: An experiment in making data governance tangible with design

Ame Elliott

Abstract: This chapter examines the intersection of design, data policy, and climate science to address complex societal challenges. Reporting results of a series of workshops with 75 participants on the theme 'The User Experience of Data Donation to Support the Clean Energy Transition', it describes research motivations, the workshop methodology, session outcomes, and challenges faced in cross-disciplinary research on the topic of data donation. Workshop participants had a mostly ambivalent reaction to the topic of data donation for clean energy. Rather than embracing the task as a wicked problem, participants mostly treated it as a 20th-century engineering challenge that they felt they lacked the professional knowledge to address. This chapter hypothesises possible explanations and alternatives, concluding with implications for strengthening cross-disciplinary collaboration to address complex challenges outside the scope of a single discipline. Changes in user design education to emphasise public interest technology could better equip future participants to work on topics such as user data donation.

Keywords: user experience design, participatory design, policy, energy transition

1. Introduction

At the time Horst Rittel co-authored *Dilemmas in a general theory of planning*, explaining the **role of wicked problems with no right or wrong answers in social policy**, he was working as a professor of the science of design (Rittel & Webber, 1973). Fifty years later, however, the design disciplines are not always present in discussions of social policy. This chapter reports on a series of four exploratory workshops that correct that imbalance and centre the discipline of user experience (UX) design in policy. The workshops bring a **design-centred approach to data governance** discussions by using a scenario that is a good fit with Rittel and Webber's definition of wicked problems: exploring trade-offs between communal energy efficiency and individual privacy. The workshops, titled 'The User Experience of Data Donation to Support the Clean Energy Transition', sit at the intersection of three distinct disciplines: user experience design, data policy, and climate science. The research is a preliminary attempt to break silos and strengthen cross-disciplinary collaboration to address an urgent social and environmental problem.

This chapter explains the research motivations and workshop methodology, shares the outcomes of the sessions, and identifies challenges for cross-disciplinary research on wicked problems.

2. Motivation and research questions

This section provides an overview of data donation in the context of clean energy.

a) How data donation can promote clean energy usage

During the winter of 2022–2023, energy-saving measures were an important topic of conversation in response to Russia's invasion of Ukraine and an urgent need to use fewer fossil fuels and shift energy policy. Amidst concern that electricity might be rationed, there was increased interest in alternatives to supplement or replace the existing grid. Coupled with technical advances producing new clean energy products suitable for urban infill in European contexts – for example, rooftop and balcony solar panels, companies began selling directly to consumers. These products are unfamiliar to many people and thus require new ways of thinking about energy use.

These new technologies enable new types of interactions, such as deciding to use solar energy immediately, store it in a battery, or sell it back to the grid provider. In some cases, people may need to make explicit decisions about how their energy systems perform, while in other cases, these decisions can be made automatically based on machine learning algorithms and behaviour modelling. Data about energy usage is an essential component in training systems to better manage power and use fewer fossil fuels. Because this technology is new to many people, and many customers are using green energy equipment for the first time, there is a need for knowledge-sharing and awareness-building to improve the efficiency of these supplemental energy supply systems. Peer-to-peer data-sharing networks could help people make better energy management decisions by learning how others use similar equipment. Even people not using clean energy products can directly contribute to improved efficiency by sharing how their energy use changes with circumstances.

b) Identifying a usage scenario

This explosion of new options for energy generation and storage and a renewed focus on energy conservation co-occurred with the easing of Covid-19 pandemic measures and the return to office work. Many companies had flexible policies about if or how frequently they expected employees to be in the office, and many people were absent due to their own illnesses or caring for family members. These two trends can be summarised from a facilities management and employer perspective as, 'Should we heat empty offices? Should we track where our employees are to avoid heating unused spaces?'

Although Internet of Things (IoT) devices such as smart sensors make it possible to track building occupancy, privacy concerns and worker protections limit their use. In this exploration, building occupancy and energy usage would be shared not with a private company selling devices, such as Google Nest, but with a cooperative with a data governance structure designed to protect mutual and planetary well-being. Initial explorations of workshop activities considered data governance mechanisms for workplace data; however, after a discussion with the building director of a 16,000m^2 office facility in Berlin, the focus shifted away from workplaces and towards private homes. The reasons to shift away from workplace data were 1) the difficulty of establishing permissions and accountability when including additional stakeholders, e.g. employers, employees of various companies, building owners, and facility managers; 2) concerns that participants would self-censor or decline to participate if they felt uncomfortable speaking about an issue related to their employer or workplace; 3) recognition that workers' councils and trade unions are already involved in this area, and some individuals are in 'wait and see' mode, expecting organisational leadership to act on their behalf, whether to protect the climate or their rights.

In the interest of simplicity for an initial feasibility study, the workshops were organised in the context of citizens taking action to manage data about energy usage in their own homes. A **data governance platform** is multi-sided, and potential stakeholders could include:

- Residents of apartment buildings
- Apartment owners
- Owners' associations
- Building managers
- Hardware companies that install, develop, and potentially own energy generation and storage equipment
- Grid providers of electricity
- Data brokers or associations that hold the data.

Given this complexity, the additional complexity of a workplace environment was undesirable.

c) Goals

The initial hypotheses for this research were that 1) data donation is helpful for managing the energy transition, 2) careful user experience design would explain the benefits of data donation and motivate people to do so, and 3) the benefits of data donation are not primarily financial (energy cost savings), but psycho-social (pride in taking action for the energy transition, feeling a sense of belonging to a community).

The workshops were short and intended to be initial introductions to build awareness of the topic of data donation and to bring together new groups of people with the goal of establishing cross-disciplinary collaborations. The expected outcomes of the workshops were identifying interested partners for more in-depth workshops working directly with the public. Beyond finding interested collaborators, secondary goals included gaining an understanding of the current state of professional discourse around data donation and capturing open questions that could be addressed in future rounds of research. **'Professional discourse'** includes at least three distinct specialisations: user experience designers/software developers, policymakers, and scientists. A tertiary goal was co-developing initial design recommendations for user interfaces for data donation.

3. Workshop methodology

Four multidisciplinary workshops in Helsinki, Lucerne, Berlin, and Paris were conducted between June and November 2023 on the topic of 'The User Experience of Data Donation to Support the Clean Energy Transition'. In total, 75 people with a range of professional backgrounds, including user experience designers, service designers, data scientists, climate scientists, and policy experts, participated. The workshops were invited sessions as part of professional conferences/meetings (Helsinki and Berlin) and university symposia (Lucerne and Paris). Participation was voluntary. Attendees at the workshops chose to attend the session based on personal interest and were not recruited or compensated. There was a very broad range of familiarity with the topic, and the heterogenous backgrounds were considered an advantage for forming new collaborations rather than a limitation to overcome by restricting participation to experts. Self-reported lack of expertise did inform participants' experience of the workshops.

a) Structure and activities

The sessions lasted 60 minutes (Lucerne and Paris) or 90 minutes (Helsinki and Berlin) and began with a presentation introducing the concept of smart microgrids (solar, wind, and battery storage) for domestic use, describing use cases as discussed in section 2b. Next came a basic overview of data donation and data governance principles, followed by a sample user interface for a hypothetical energy management application. There was then a pause for questions and facilitated discussion. The Paris session ended here, but the Helsinki, Lucerne, and Berlin sessions continued into a participatory design activity with attendees prototyping user interfaces for data donation on paper using worksheets, paper, and pens. These methods are consistent with other approaches in the smart city space (Bødker & Zander, 2015; Yasuoka, 2023).

Recognising the short duration of the workshops, the activities were designed as **awareness-building advertisements** to promote interest in further collaborations and in-depth discussion in other contexts. Participants were not trained in data governance and then assigned a task to demonstrate their mastery – rather, the discussion and questions were key.

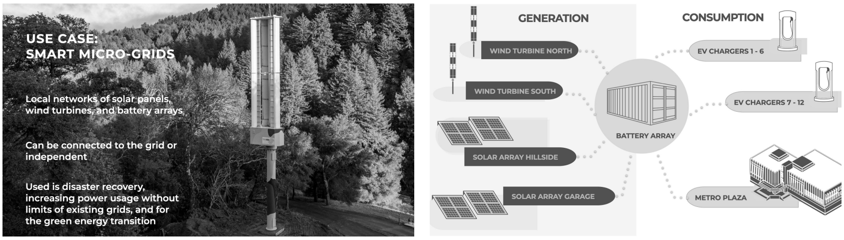

Figure 1. Slides from an introductory presentation explaining smart microgrids for using wind, solar, and battery storage for energy independence.

b) Explanatory material

The introductory presentation explained the **concept of smart microgrids** to help people move towards energy independence and away from fossil fuels. It was clearly explained this project was a general civic technology project to benefit the public sector, not a commercial technology development project, explained as follows: 'Imagine an app for people living in multi-family apartment complexes with their own solar/wind generation and battery storage. This app lets people understand their system's performance and the performance of other similar installations'.

Sharing with **'other similar installations'** was highlighted because peer-to-peer support for other people learning to manage new equipment is a relevant use case for decentralised data governance that opens the door to discussion of training for machine learning systems. Two potential use cases included 1) disaster recovery, when the existing grid was not available as a source of power, so backup power from batteries would be an emergency stopgap measure after a disaster such as a fire, flood, or unexplained outage, and 2) a voluntary shift away from the existing grid to support a clean energy transition and use power from renewable sources. After covering the basic technology, the presentation shifted to an introduction to data donation, followed by a discussion.

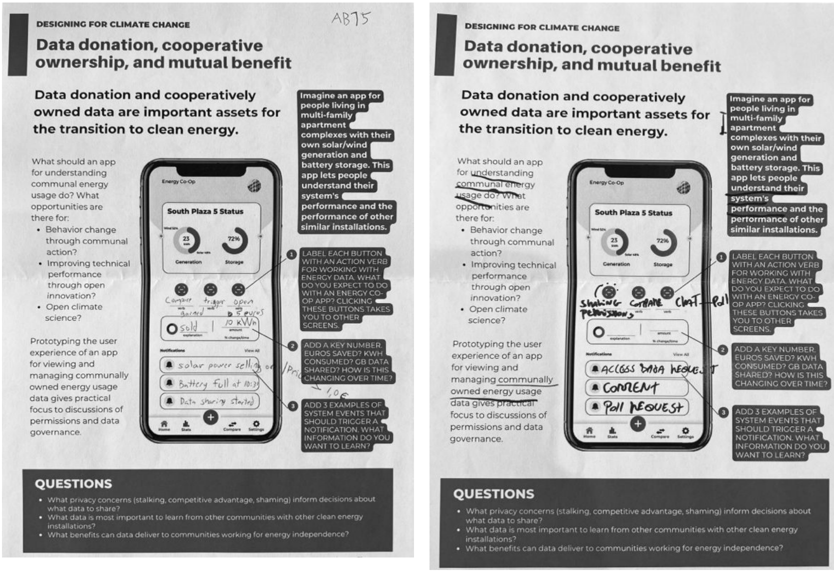

Figure 2. Worksheets used in the workshops

The following discussion questions were shared with participants to kick-start the discussion:

Data donation and cooperatively owned data are important assets for the transition to clean energy.

- What should an app for understanding communal energy usage do?
- What opportunities are there for:
  - Behavioural changes through communal action?
  - Improving technical performance through open innovation?
  - Open climate science?
- What privacy concerns (stalking, competitive advantage, shaming) inform decisions about what data to share?
- What data is most important to learn from other communities with other clean energy installations?
- What benefits can data deliver to communities working for energy independence?

These questions were presented to encourage discussion about the advantages, limitations, and harms of data donation and were not asked or answered in a structured order.

c) Paper prototyping

Two of the workshops consisted primarily of user experience designers, an audience familiar with sketching and critiquing user interfaces. One of the workshops was primarily data governance and policy experts, although some designers were present. The fourth workshop primarily comprised data scientists and climate scientists and did not do the prototyping activity.

Paper prototyping is a well-established part of design education and professional practice (Snyder, 2003), and its use by non-designers in participatory settings has a long history, as discussed, for example, in a number of literature reviews (e.g. Hollomann et al., 2018; Maartmann-Moe & Joshi, 2022). Paper prototyping was selected as a method to focus the abstract topic of data governance into something tangible and specific. The visual vocabulary of smartphone applications was familiar to all the participants, and everyone understood the task. Participants grasped that this was an exploratory exercise created for research purposes and that there was no existing software or proposal to develop actual software.

It was, however, necessary to clarify that the app depicted was a **research prototype** that did not exist and that the workshop was not providing free labour or market research for a for-profit clean tech start-up. There was initial scepticism and mistrust about who benefitted financially from such an app and how developing this type of app would be compensated. People worked independently on their own prototyping worksheets, using additional paper as necessary, occasionally sharing their thinking with a group and asking for input on their ideas.

4. Outcomes

During the workshops, handwritten notes were taken of the discussion. An analysis of the notes indicates three main categories: questions requiring further clarification (the largest groups), expression of individual opinions and preferences (fewer), and recommendations (minimal). The worksheets and sketches were qualitatively assessed in keeping with norms for participatory design projects (Bødker et al., 2022; Bratteteig & Wagner, 2016).

The primary goal of the workshops was building awareness about the topic of data donation and identifying potential partners for further research. Data donation for climate change did not spark ongoing interest among workshop participants, and there was ambivalence to ongoing participation in research. Contrary to hopes, none of the participants were part of an existing project working in this area seeking additional partners, such as an open climate science project looking for UX guidance or climate activist groups interested in in-situ prototyping. The secondary goal of capturing the state of professional discourse revealed deep silos between disciplines, and unclear mechanisms for collaboration, while the tertiary goal of developing initial design recommendations for user interfaces for data donation was too ambitious for an initial series of workshops of such short duration.

a) Hypothesis

An evaluation of the initial hypotheses follows.

aa) Hypothesis: Data donation is helpful for managing the energy transition.

This was only weakly supported. There was broad agreement that data donation is useful, an unsurprising finding for a group of participants who chose to participate in data donation workshops. However, the dominant sense was that it has unrealised potential and lacks a compelling use case in this domain.

bb) Hypothesis: Careful user experience design would explain the benefits of data donation and motivate people to do so.

This hypothesis was disproven. The value proposition for donating data to fight climate change was not clear, and user experience design was an insufficient motivator.

cc) Hypothesis: **The benefits of data donation are not primarily financial** (energy cost savings) **but psycho-social** (pride in taking action for the energy transition and feeling a sense of belonging to a community)**.**

This hypothesis was strongly disproven. The benefits of data donation were seen as almost exclusively financial (energy cost savings). In second place behind cost savings was enthusiasm for data governance and policy generally, without any special interest in the climate and energy transition aspects. That is, data donation is an intrinsically interesting concept to some. The ambivalence was unexpected, particularly as data donation for energy issues has been reported elsewhere (Baumann & Laube, 2022; Wenninger et al., 2021). Psycho-social benefits (pride in taking action for the energy transition, feeling a sense of belonging to a community) were a distant third place. This absence of positive contributions to the community was particularly unexpected and stands in contrast to patient motivations for donating medical data (Krutzinna & Floridi, 2019).

b) Potential explanations for the ambivalent response

There are many possible explanations for workshop participants' ambivalence toward the topic of data donation for the energy transition. Potential factors to change include:

- **Different participants:** Perhaps the participants were not sufficiently interested in the topic. The workshop composed primarily of data scientists and climate scientists did not take part in the paper prototyping exercise because it was outside the event structure. Maybe having participants with this type of domain expertise involved would have led to a different result. Additionally, recruiting workshop participants who already identify as motivated by climate change issues and are involved in community activism may have had a better response than conference participants. Potentially recruiting existing groups of people who are already connected – for example, residents of the same apartment building – might have garnered a warmer response.
- **Different facilitator or activities:** It is possible that a different facilitator would have obtained different outcomes. However, the author is an experienced facilitator with more than 20 years of experience leading paper prototyping exercises with different groups, including civic tech and open science projects (Elliott, 2005). Recommended changes to the activities include providing more detailed instructions and a written glossary of terms. Discussion of data donation is a very abstract topic, and paper prototyping was not successful in making it concrete and actionable in the time allowed. Alternative activities could include much longer workshop sessions with opportunities to explore more subtopics in detail or clearly marking some subtopics as out of scope for the exercise. For example, an economist participating repeatedly suggested that there could be a market for donated data, but this was never illustrated with any concrete user interface suggestion, such as sketching what a data market might look like in a phone app. No participant with UX experience built on this idea, presumably because they had not

encountered a data marketplace before. Potentially closing off topics such as marketplaces might have narrowed focus and avoided repetitive discussion during limited workshop time to attain more tangible outcomes.

- **Different topic:** Focusing on an area other than home energy usage may have confirmed the hypotheses, for example, by returning to the original concept of looking at workplace energy consumption and involving building managers. Alternately, a different example scenario, such as data donation on public transport, might have been a better choice, such as Switzerland's PosMo (Positive Mobility) data donation effort (Baumann & Laube, 2022).

c) Insights from the workshops

The absence of a clear use case explaining the benefits of data donation for the home energy transition dampened participant enthusiasm, and there was little appetite for additional collaboration. A general sense of climate emergency, geopolitical vulnerabilities due to fossil fuel purchases, or worries about domestic security in a climate disaster like a flood were not sufficient to translate into willingness to donate home energy use data.

There are some additional insights that contextualise the outcomes:

- *There is deep mistrust of smart devices in the home, and monitoring energy usage is perceived as invasive.* Questions about exactly how a hypothetical system would work and exactly what data would be shared were difficult to answer and did not lead to the kinds of generative prototyping of ideas that a data governance expert might have expected, such as differential privacy and synthetic data. Rather, it led to a suspicion that the premise of data donation was only marginally useful enough to warrant collecting such data. Concerns about corporate surveillance outweighed curiosity about other forms of communal data governance.
- *Workshop participants had positive opinions of their local power companies and expected that they would produce an app like this 'if it were necessary'.* There was a general passivity and expectation that qualified experts, particularly power companies, would develop a data donation framework and an accompanying app or potentially add features to existing apps. The app displayed below was shared by a participant who was a customer of Nordic Green Energy, and they pointed out that it handled 'everything', including car-charging management. They expected any additional information related to energy to be part of this singular app

and were sceptical that there would be another place for interacting with energy data.

Figure 3. Screenshots from the Nordic Green Energy app taken from the Google Play store.[1]

- *The level of digitalisation by power companies varied widely.* The exact mechanisms for selecting an energy provider, recording energy usage, and estimating/paying for energy use depend on regional norms. Some Scandinavian participants were already using power management apps on their phones, which they shared and discussed, and that session had little sense that this was a new topic. Some participants outside Scandinavia considered basic digitalisation a higher priority than exploring donation data, which they referred to as an 'edge case' that they did not expect a power company communicating by post to address any time soon.
- *Electricity delivery is highly localised, and everybody thinks their community is too unique for a general-purpose solution.* Workshop participants in all four locations objected to some aspect of the task, saying, 'In [Switzerland, Finland, Paris, the former East Berlin] we have a unique situation and need a unique solution'. These comments dealt with cultural identity and norms, such as familiarity with cold, habits of dressing indoors, relationships with neighbours, and standards of entertaining

1 https://play.google.com/store/apps/details?id=fi.nordicgreen.app

guests at home. For example, Scandinavian participants were proud of their high-quality insulation and cosy sweaters. Parisian participants were proud of their robust social lives, which kept them away from chilly apartments with historical character.

5. Conclusion

Although the workshops did not yield the expected outcomes, there are relevant implications for future projects, though there will not be a second round of workshops with the public due to limited support for future collaborations from participants. First, working in a cross-disciplinary manner is challenging. These workshops at the intersection of user experience design, data governance, and climate required a broad range of skills. Many participants felt that they lacked the skills to be effective because of missing domain knowledge, potentially self-reporting that they were the wrong participants, as suggested in section 4a. The downside of such a sentiment is a feeling of helplessness and passivity while waiting for other, presumably better-qualified people to intervene. But taken positively, these missing skills are a call to action.

a) Crisis of professional knowledge in handling wicked problems

In a world of increasing complexity and interrelated systemic crises, traditional models of professionalism are no longer adequate. Re-examining both norms for professional training and ways of collaborating is essential to creating cross-disciplinary teams that work smoothly. Revisiting Rittel and Webber (1973) on the topic of professional knowledge and interconnected systems is instructive:

> Based in modern science, each of the professions has been conceived as the medium through which the knowledge of science is applied. In effect, each profession has been seen as a subset of engineering. … Because it was fairly easy to get consensus on the nature of problems during the early industrial period, the task could be assigned to the technically skilled, who in turn could be trusted to accomplish the simplified end-in-view. Or, in the more work-a-day setting, we could rely upon the efficiency expert to diagnose a problem and then solve it, while simultaneously reducing the resource inputs into whatever it was we were doing. We

> have come to think about the planning task in very different ways in recent years. We have been learning to ask whether what we are doing is the right thing to do. That is to say, **we have been learning to ask questions about the outputs of actions** and to **pose problem statements** in valuative frameworks. We have been learning to see social processes as the links tying open systems into large and interconnected networks of systems, such that outputs from one become inputs to others. In that structural framework it has become less apparent where problem centres lie, and less apparent where and how we should intervene even if we do happen to know what aims we seek.

Data donation is a complex topic, weaving together multiple socio-technical systems in ways that transcend efficiency. Workshop participants struggled to identify and articulate the value of data donation in terms other than efficiency and cost savings, not embracing the nature of wicked problems and treating the task as a 20th-century engineering problem. Future efforts should begin by providing a better theoretical grounding in the nature of the problem as opposed to jumping immediately to attempted solutions. Beginning by providing a more defined point of view could also lead to more impactful workshop outcomes.

b) Helping design professions serve the public interest

About half of the workshop participants had formal training in design, primarily user experience design and service design. Design education does provide training in addressing ambiguous, systemic problems (Vorvoreanu et al., 2017), but despite this training, many designers found it difficult to engage with the abstract concepts of data donation and the energy transition. Showing money saving by sketching a $ in an app is easy and familiar. Demonstrating the benefits of data donation in an app is unfamiliar and difficult. Designers have the skills to address wicked problems, but the socio-technical context of policy differs from commercial product development. To effectively contribute to cross-disciplinary policy challenges, designers should build and maintain professional practices for working in the public interest. The language of 'users' and 'consumers' of products and services is firmly situated within a market-based framework for decision-making, so it is logical that designers found cost savings the easiest impact of data donation to quantify. An opportunity for future work is to more explicitly prepare the design professions to work on different types of policy problems without financial exchange as the central idea.

Adequately addressing complex systemic problems requires a cross-disciplinary approach, and more demands are being placed on professionals in every domain to stretch outside their comfort zones and work on wicked problems. The methods of user experience design have a role to play, but additional research is needed to strengthen the discipline and support work in non-commercial contexts.

Acknowledgements

The author wishes to thank the workshop participants, without whom this chapter would not be possible. Thank you to Ladies That UX Berlin, MyData 2023, Hochschule Luzern, and Université Paris Cité for space and ideas.

References

Baumann, D., & Laube, A. (2022, July 6). *Posmo: Ethical revolution in the mobility data market*. SocietyByte Wissenschaftsmagazin der Berner Fachhochschule. https://www.societybyte.swiss/en/2022/07/06/data-and-ethics/.

Bødker, S., & Zander, P. (2015, June 27–30). *Participation in design between the public sector and local communities*. Proceedings of the 7th International Conference on Communities and Technologies, New York, NY, United States. https://doi.org/10.1145/2768545.2768546

Bødker, S., Dindler, C., Iversen, O. S., & Smith, R. C. (2022). What are the results of participatory design? In *Participatory design: Synthesis lectures on human-centered informatics* (pp. 95–102). Springer Nature. https://doi.org/10.1007/978-3-031-02235-7_9

Bratteteig, T., & Wagner, I. (2016, August 15–19). *What is a participatory design result?* PDC 2016: Proceedings of the 14th Participatory Design Conference. New York, NY, United States.

Elliott, A. (2005). Physical artifacts for promoting bilingual collaborative design. *EPIC 2005: Ethnographic Praxis in Industry*. Redmond, WA: American Anthropological Association.

Holloman, A., Hernandez-Cuevas, B. Y., Thomas, S., Swain, K., & Crawford, C. S. (2018, June 3–5). A systematic literature review of paper prototyping. In *Woodstock '18: ACM Symposium on Neural Gaze Detection*. New York, NY, United States.

Krutzinna, J., & Floridi, L. (Eds.). (2019). *The ethics of medical data donation*. Springer.

Maartmann-Moe, H., & Joshi, S. G. (2022, August 19–September 1). Prototyping and prototype use in participatory design with older adults: A systematic literature review. In V. Vlachokyriakos, J. Yee, E. Grönvall, R. Noronha, A. Botero, C. Del Gaudio, Y. Akama, R. Clarke, & J. Vines (Eds.), *PDC'22: Proceedings of the Participatory Design Conference 2022 volume 1* (pp. 169–182). ACM, New York, NY, United States. https://doi.org/10.1145/3536169.3537791

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169. https://doi.org/10.1007/BF01405730

Synder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann Publishers.

Vorvoreanu, M., Gray, C. M., Parsons, P., & Rasche, N. (2017, May 06–11). Advancing UX education: A model for integrated studio pedagogy. In G. Mark & S. Fussell (Eds.), *CHI'17*: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1441–1446). ACM, New York, NY, United States. http://dx.doi.org/10.1145/3025453.3025726

Wenninger, M., Maier, A., & Schmidt, J. (2021, July 15). DEDDIAG: A domestic electricity demand dataset of individual appliances in Germany. *Sci Data* 8, 176. https://doi.org/10.1038/s41597-021-00963-2

Yasuoka, M. (2023). How citizens participation begin and continue in developing smart cities. In N. A. Streitz & S. Konomi (Eds.), *Distributed, ambient and pervasive interactions. HCII 2023. Lecture notes in computer science* (pp. 182–192). Springer. https://doi.org/10.1007/978-3-031-34609-5_13

Digital technology for people: Learning from 'non-Western' and Indigenous cultures

Marc Steen

Abstract: This essay explores alternative ways to manage societal and environmental issues. Several assumptions of the European Enlightenment – such as the perspectives that people are isolated individuals and nature is an object – have led to many of our current problems. Often, digital technology is designed and deployed within these assumptions. It can be useful to explore alternative ways to design and use digital technology by looking at non-Western, Indigenous cultures. Four examples are discussed briefly: Ubuntu philosophy and the example of a language app for African languages; Māori data sovereignty; the building of a Lakota sweat lodge to inform building computer hardware; and the role of LAN Houses and mobile phones in the empowerment of people in a favela in Brazil. The essay closes with suggestions for learning from these examples.

Keywords: AI, indigenous, empowerment, mutual learning

1. Introduction

Our societies and daily lives are shaped to a great extent by digital technology. We used to be happily surprised with the new and shiny, . But not anymore. Weare increasingly and justifiably concerned about the impacts, harms, and risks of social media, algorithms, and Artificial Intelligence (AI) on society, on our daily lives, and on democracy, inequality, and our environment (Bender et al., 2021; Benjamin, 2019; Buolamwini, 2023; Crawford, 2021; Eubanks, 2017; Noble, 2018; O'Neil, 2016). Crucially, we need to understand that technology is not inherently good or bad, nor is it neutral. Rather, people, as well as the corporations and the states in which they collaborate (Runciman, 2023), design and deploy technology to serve their particular interests. Key assumptions of the European Enlightenment, *that people are isolated individuals, that nature is an object to be exploited*, are very present, in our societies and daily lives. We tend to focus on means, on technology, but are largely at a loss when it comes to articulating and pursuing ends (MacIntyre, 2007).

Below, several examples are introduced for doing things differently: how to empower citizens to design and deploy digital technology that promotes values like justice, freedom, conviviality, and democracy. The examples originate from so-called 'non- Western' or Indigenous cultures or marginalised communities. This approach is useful for exploring solutions to problems that the European Enlightenment, with its assumptions of individuality, domination, and exploitation, has created (Steen, 2022). Please note that this approach does not imply that non-Western or Indigenous cultures are better than Western cultures, or vice versa. Please also note that this gesture – *to look to other cultures in search of useful things* – carries the risk of (neo)colonialism and harm. My current gesture is meant to be respectful; I intend to learn from diverse cultures. Moreover, I understand that there is diversity of opinions within a culture; "A single 'Indigenous perspective' does not exist" (Lewis, 2020, p. 4).

The goal of this essay is to explore alternative ways to design and deploy digital technologies, thus attempting to more effectively deal with societal and environmental issues. It is meant to help raise awareness and to inform, inspire, and empower people to act more collectively, working as **active citizens,** rather than as individual and passive consumers or subjects at the receiving end of what corporations or states design and deploy.

This exercise has roots in Aristotelean virtue ethics, which is concerned with enabling people to cultivate relevant virtues to live well together. The premise is that people *can* use digital technologies as tools to cultivate specific virtues (Vallor, 2016). Rather than what often happens, the other way around. Currently, corporations, with business models for grabbing and monetising people's attention, offer social media apps that corrode people's virtues like self-control, honesty, or civility by luring them into their platforms and incentivising them to spread fake news and engage in mud-throwing and polarisation (Vallor 2016: pp. 159-187).

The examples below are from different continents: Ubuntu and the example of a language app for African languages; Māori data sovereignty from New Zealand; the building of a Lakota sweat lodge to inform building computer hardware; and the empowering use of digital technology in a favela in Brazil.[1] Please note that the discussions below are merely vignettes or illustrations. For a fuller appreciation of people’s perspectives and experiences in these vignettes, more extensive research would be required.

2. Ubuntu, a language app, and relational ethics

Several cultures in sub-Saharan Africa adhere to variations of the philosophy of Ubuntu; Ubuntu recognises the humanity of a person through that person’s relationships with other people. It is often summarised as follows: *I am because we are*. Ubuntu has been a key tenet in the work of the Truth and Reconciliation Commission in South Africa in the mid-1990s and in Bishop Desmond Tutu’s leadership.

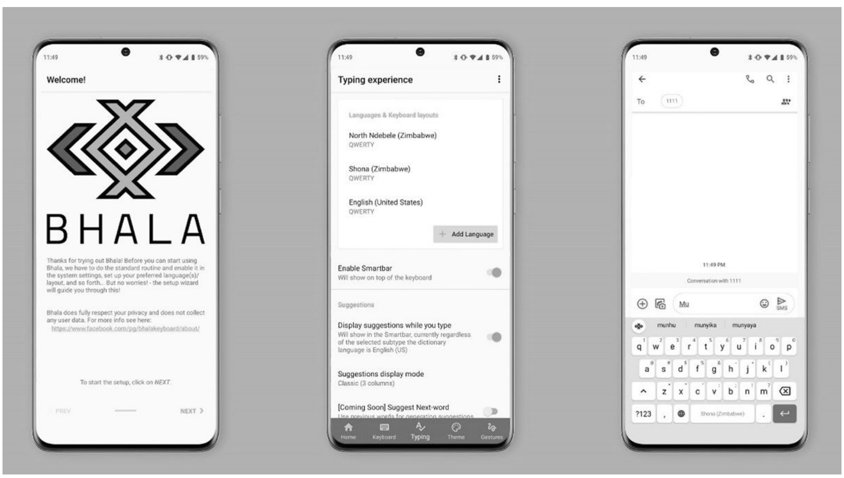

Figure 1: Bhala keyboard app (https://mamgobozidesign.com/bhala-rebranding), designer: Osmond Tshuma (Mam’Gobozi Design Factory), 2021; used with permission

Sábëlo Mhlambi, a Fellow at Harvard’s Berkman Klein Center for Internet & Society and a Technology and Human Rights Fellow at the Carr Center

1 This approach, to present examples, was inspired by Jer Thorp’s ‘*Living in data*’ (2021); it has many illustrations, to tell stories not only verbally, but also visually. Similarly, I also included illustrations. The Māori example (below) also appears in his book.

for Human Rights Policy has pioneered the application of Ubuntu philosophy to the design and application of AI systems. He articulated the following critiques: people from marginalised communities are often excluded from the design process; there are biases in the collection of data and in the selection of features; the people involved in design often view technology as neutral and fail to recognise the many and diverse relationships that exist in society; and systems can lead to harmful commodification and centralisation of data and resources. Drawing from Ubuntu, Mhlambi (2020) argues that "[t]echnology should be created with a normative goal to eradicate inequality through the participation of the most disenfranchised", and proposed that systems can be "used for public good and made available to the public in ways that protect privacy and promote the wellbeing of society" and that "[g]reater funding and access to technical skillsets must be made available to the most disenfranchised" (p. 25). Moreover, Mhlambi (2020) founded *Bhala*, an AI startup that aims to democratise the advances of AI to millions of Africans.[2] The company offers a free *keyboard and stickers app* for African languages like Ndebele, Shona, Swati, Swahili, Xhosa, and Zulu, with a spell-checker, auto-complete function, and tools to classify, generate, and correct texts in these languages (see Figure 1). This is meant to remedy the default bias, persistent in digital technology towards English, and thereby promote more linguistically and culturally appropriate communication via digital technology.

More broadly, a similar approach has been developed under the header of *relational ethics* (Birhane, 2021; Birhane & Cummins, 2019; Mhlambi, 2020; Mhlambi & Tiribelli, 2023), as an effort to repair and supplement the default assumption that people are isolated individuals. Relational ethics can be understood to include ethics of care and feminist ethics (Held, 2006). This approach is concerned primarily with, for example, the well-being of people who are affected by a specific system's deployment and the empowerment of the people affected by decisions regarding design and deployment. Their participation in these processes and criticism of systems – notably, the ways in which their deployment can exacerbate injustices or inequalities – is imperative in shifting power balances.

2 https://sabelo.mhlambi.com/ and https://www.itweb.co.za/article/bhala-app-lets-locals-create-online-content-in-their-own-lingo/raYAyqodRDnvJ38N. Please note that this app is still in beta and is only available for Android at the time of this publication.

3. Māori data sovereignty and self-determination

This story starts with the Treaty of Waitangi (see Figure 2) (Thorp 2021: pp. 219-235). In 1840, representatives of the British Crown and Māori chiefs from New Zealand's North Island signed a treaty with three articles. In Article 1, the Māori people agreed to cede their governance rights to the Crown. Article 2 establishes that the Māori will retain full chieftainship (*rangatiratanga*) over their lands, villages, and all their treasures (*taonga*). In Article 3, the Māori people obtain full rights and protections as British subjects. We will focus here on the word "treasures". This includes not only material objects but also culturally valuable resources and immaterial objects, ideas, and techniques. Over several years, the scope of taonga has been debated.

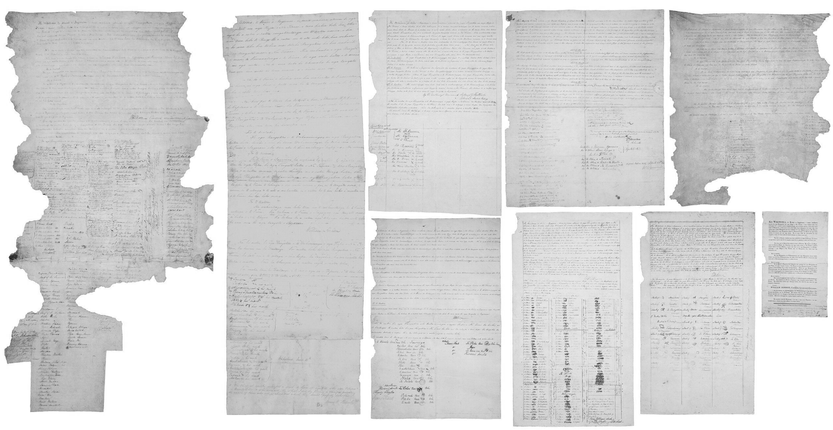

Figure 2: Treaty of Waitangi; https://en.wikipedia.org/wiki/File:Treaty_of_Waitangi_all_documents.jpg

This is not unusual with treaties; before and after their signing, parties debate the precise wordings and their meanings, especially if a treaty is written in different languages, as is the case with the Treaty of Waitangi, which was written in both English and Māori. A key question is: *What belongs to taonga?* The answer determines over which taonga the Māori will retain full chieftainship. In the course of the 1980s and 1990s, a series of rulings determined that radio frequencies, spiritual places, and knowledge about flora and fauna belonged to taonga. Similarly, the Māori Data Sovereignty

Network (*Te Mana Raraunga*) advocates for Māori rights regarding all sorts of data that pertain to their way of life, their environment, and their culture.[3] This would mean that these data are and remain Māori's taonga even if these data are collected by an organisation in the UK and stored in a cloud service of a corporation in the US).

For data sovereignty, the United Nations' Declaration On The Rights Of Indigenous Peoples of 2007[4] is also relevant. It prioritises and emphasises rights for self-determination, autonomy, and participation and includes rights that pertain to cultural heritage and knowledge (e.g., public health data and data about animals and plants). For former colonies, like New Zealand, this is crucial. Data sovereignty is an effort to try to repair some of the damages of colonisation, similar to giving back ownership over objects that were stolen during colonisation. Moreover, data sovereignty is meant to combat common practices in which states or companies from abroad collect data on Indigenous people and their environment, owning, analysing, and utilising this data for their ownbenefit and causing harm to the people to whom these data pertain. We can learn from this example concerning the technological sovereignty of the EU[5] vis-à-vis the US (where corporations have much power) and China (where the state has much power) to promote freedom, equality, democracy, and participation.

4. Building a Lakota sweat lodge and building computer hardware

First a bit of context. In 2019, a group of diverse, mainly Indigenous, people – from Canada, Australia, New Zealand, and the US – came together "over 20 months, across 20 time zones, during two workshops" (Lewis, 2020, p.4) to explore various approaches to AI. Although their approaches are very different, they share several tenets, such as Locality, i.e. the requirement that AI systems are designed in partnership with Indigenous communities, and Relationality and reciprocity, i.e. the requirement that AI systems help to appreciate how humans and non-humans are interdependent (Lewis,

3 https://www.temanararaunga.maori.nz/
4 https://www.un.org/development/desa/indigenouspeoples/declaration-on-%20the-rights-of-indigenous-peoples.html
5 https://digital-strategy.ec.europa.eu/en/news/digital-sovereignty-european-chips-act-enters-force

2020, p. 20-21).[6] In the workshops, one subgroup drew parallels between the process of building a Lakota sweat lodge and building computer hardware, proposing that both can be done "in a Good Way ... because AI cannot be made ethically until its physical components are made ethically" (Lewis, 2020, p. 76). For the Lakota, "[t]he sweat lodge is a place where knowledge is generated about the world" (Lewis, 2020, p. 76) just as a computer is a place where knowledge is stored and processed (see Figure 3).

Figure 3: Building a sweat lodge; (https://commons.wikimedia.org/wiki/File:Building_the_Turtle_Shell_(2028489212).jpg)

When building a sweat lodge, it is critical to start with identifying a specific need and involving relevant stakeholders: "individuals and community members, known and unknown, seen and unseen, including: Stone Spirits, Plant Peoples, Animal Peoples" (Lewis, 2020, p. 77). Similarly, building

6 Several more tenets are mentioned in the *Indigenous Protocol* (Lewis 2020, pp. 20-22): 'Responsibility, relevance and accountability; Develop governance guidelines from Indigenous protocols; Recognize the cultural nature of all computational technology; Apply ethical design to the extended stack; Respect and support data sovereignty.'

a computer device would need to start with identifying a specific need and involving, for example, the communities of the places from which raw materials originate. Furthermore, each object that goes into building a sweat lodge needs to be compensated for in some way, "offering something valuable in exchange" (Lewis, 2020, p. 78). Analogically, regarding computer hardware, the people involved in mining the materials are entitled to fair compensation and safe working conditions. After mining, the earth would need to be repaired to a healthy state. Finally, the people who build a sweat lodge are required to repurpose, return, or transform the materials they have used. For computer hardware, this would entail requirements to repurpose or otherwise take care of the materials after the computer hardware's life cycle.

One can draw parallels with concerns for the materials, labour, and energy that go into building and using AI systems and the costs and risks to people and the environment (Bender et al., 2021; Crawford, 2021). This problem is particularly hard to solve. Very often, especially for organisations with few resources, it is convenient to procure affordable hardware and cheap cloud services from big tech (e.g., an Android phone and Gmail) rather than more 'responsible options' (e.g., a Fairphone and Protonmail). While some people are able to build their own hardware with, for instance, an Arduino or Raspberry Pi, or their own software models (e.g., Hugging Face), this requires advanced expertise and skills. This need motivate public agencies or grassroots organizations to provide training options to develop such technical expertise and skills.

5. LAN Houses and mobile phones in a favela

This example considers how disadvantaged ("oppressed") people in a favela in Brazil can utilise technology for their empowerment (see Figure 4). This is based on extensive ethnographic studies by David Nemer (2022). He grew up close to the favelas, but never physically went there. This is typical; favela residents often stay in the favelas, and non-residents rarely go into the favelas. Nemer builds on the ideas of Paulo Freire and discusses how people can use digital technology to fight for freedom.

The favelas of Brazil are infamous for their gangs, trafficking, and violence. This is, however, not the whole story. The favela residents use technology in creative ways to build community, foster safety, and improve their lives. In the favelas, there are Telecenters and LAN Houses. These

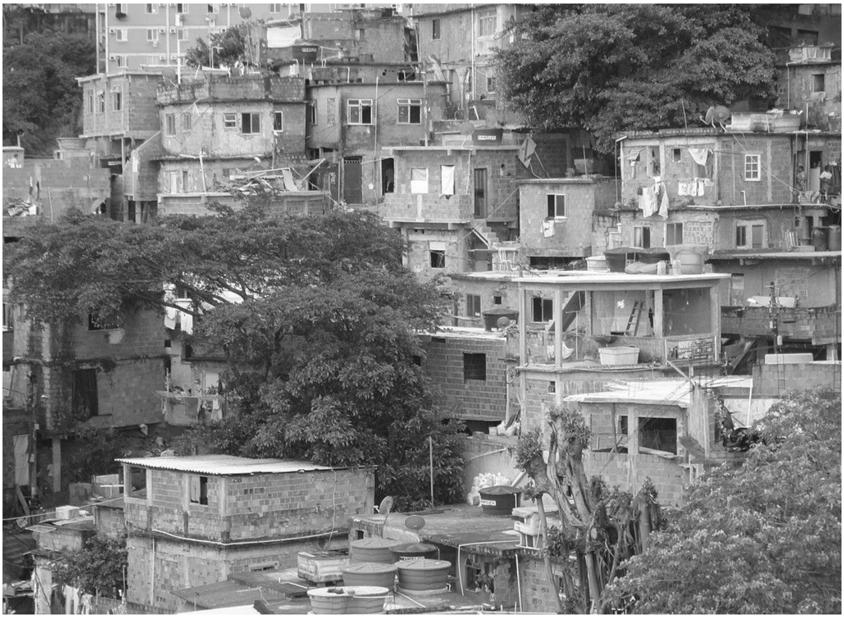

Figure 4: Example of a favela in Rio de Janeiro; (https://commons.wikimedia.org/wiki/File:Favela_cantagalo.JPG)

function not only as places where people can use computers and go online but also as shelters when a gunfight breaks out on the street. They also function as places where children can hang out safely, use Facebook, and play games. They are educational hubs where people can learn computer skills, write their CVs, and apply for jobs online (Nemer, 2020, p. 55-80). Moreover, some Telecenters and LAN Houses enable the favela residents to access the internet on their mobile phones. Typically, the mobile operators' service within the favelas is poor, and subscriptions to internet providers are often too expensive for favela residents. To combat this issue, Cyber LAN House Gustavo "used fifteen Linksys routers placed inside plastic boxes on the light poles and five hundred meters of Ethernet cable to connect his LAN House" (Nemer, 2020, p. 43). With this hack, he was able to provide affordable internet access to the favela residents.

Furthermore, this LAN House now had strong Wi-Fi, which changed how favela residents can use their mobile phones or "*xinglings*". Typically, they would switch off the mobile data (too expensive) and use Wi-Fi instead to go online (Nemer, 2020, p. 47). Now, groups of, for instance, three or four friends, could use their mobile phones in ways that promote community. One friend brings a mobile phone. Another brings a charger. Another friend brings a USB cable to connect to the LAN House computer. They can then hang out and socialise, sharing the phone and the computer. This looks distinct from a group of four people looking at their separate mobile phones, as is typical in many other places. This example can inspire novel ways to empower disadvantaged communities, not by inventing products or services for them, but by enabling them to find ways to use digital technology that supports them in their specific circumstances.

5. Discussion and conclusion

Now, what can we learn from these examples? How can they help us to explore ways to enable people to design and use digital technology differently and, more specifically, to find ways to live well together?

As noted already, there is a risk of looking at other cultures to find useful things, to grab these and to use these to do more of the same, and *not* learn anything new (Steen, 2012). This we need to prevent. Instead, to do justice to the people we want to learn from, we can organise communication, collaboration, and mutual learning (Steen, 2022). Meanwhile, to facilitate such communication and collaboration, below are several tentative conclusions, mainly meant as suggestions.

Please note that some findings may seem obvious on a theoretical level. On a practical level, however, it can be very challenging to design and deploy technology in ways that facilitate freedom, equality, and conviviality, especially given the dominant and default emphasis on designing and using technology to increase control over people, make short-term, financial profits based on exploitation, and increase efficiency. The move from individuality to conviviality can be especially challenging indeed.

From Ubuntu and relational ethics, we can learn to include a broad diversity of people – not just those who write and read English – in the design and deployment of technology. Such perspectives can help to draw attention to the deleterious impact of power differences or imbalances. Often, corporations and states have too much power. From the example of Māori data sovereignty and self-determination, we can learn about ownership and control. Using legal documents and legal reasoning, people can obtain ownership and control over data about their own lives. This would be a welcome alternative to the default practices of corporations and states that collect and utilise data to further their objectives. The example of building a Lakota sweat lodge draws attention to the materials, labour, and energy that go into the creation and deployment of digital technology – state-of-the-art AI systems in particular. The resources involved typically are outside of most people's awareness. The extraction of materials and the consumption of energy happen overseas: in mines, often in conflict areas, and in unhealthy sweatshops and unfair supply chains. Any initiative aimed at managing materials, labour, and energy with greater care and responsibility is encouraged. Finally, the examples of favela residents using digital technology for their empowerment remind us of the need to make technology flexible, so that people can adopt and modify it to fit their own ways of living and goals. Furthermore, it can remind us to give ordinary people, as prospective users, a say in the design process, so they can contribute meaningfully and transcend their role of mere 'users' ((Schuler & Namioka, 1993, Oudshoorn & Pinch, 2003). Moreover, we need to involve more stakeholders, to speak on behalf of people affected by the technology's deployment, and act to protect and conserve the environments from which materials and energy are sourced.

Acknowledgements

The author would like to thank the Weizenbaum Institute for the Visiting Research Fellowship (Open Fellowship Program) and to acknowledge the inspiring conversations with various people at the Weizenbaum Institute which helped to write this article , and two anonymous reviewers for their valuable comments.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns, 2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Birhane, A., & Cummins, F. (2019). *Algorithmic Injustices: Towards a Relational Ethics*. https://arxiv.org/abs/1912.07376

Buolamwini, J. (2023). *Unmasking AI: My mission to protect what is human in a world of machines*. Penguin Random House.

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Eubanks, V. (2018). *Automating Inequality: How high-tech tools profiles, police, and punish the poor*. St. Martin's Press.

Held, V. (2006). *The Ethics of Care: Personal, Political, and Global.* Oxford University Press.

Lewis, J.E. (2020). *Indigenous protocol and artificial intelligence position paper.* The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research. https://doi.org/10.11573/spectrum.library.concordia.ca.00986506

Mhlambi, S. (2020). From rationality to relationality: Ubuntu as an ethical and human rights framework for Artificial Intelligence governance. *Carr Center Discussion Paper Series, 9*.

Mhlambi, S., & Tiribelli, S. (2023). Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms. *Topoi, 42*(3), 867-880. https://doi.org/10.1007/s11245-022-09874-2

Nemer, D. (2022). *Technology of the oppressed: Inequity and the digital mundane in favelas of Brazil.* MIT Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Penguin.

Oudshoorn, N., & Pinch, T. (2003). *How users matter: The co-construction of users and technology*. MIT Press.

Runciman, D. (2023). *The handover: How we gave control of our lives to corporations, states and AIs*. Profile Books.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices.* Lawrence Erlbaum Associates.

Steen, M. (2012). Human-centered design as a fragile encounter. *Design Issues, 28*(1), 72-80.

Steen, M. (2022). Learning From Indigenous Cultures. *IEEE Technology and Society Magazine, 41*(4), 39-43. https://doi.org/10.1109/MTS.2022.3215875

Thorp, J. (2021). *Living in data: A citizen's guide to a better information future*. MCD.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

From theory to practice: An empirical study on communities of practice at the intersection of science, civil society, and the arts

Ines Weigand

Abstract: The understanding of science and the knowledge landscape has undergone a continuous paradigm shift from science as an exclusive authority and institutionally protected space for the production and validation of knowledge with a high degree of autonomy, to more open and social models in which institutionalised science engages with non-scientific actors in a collaborative, reflective and transparent manner. This requires the development of interdisciplinary and participatory research and learning practices that integrate multiple perspectives and epistemologies, as advocated by the Open Science movement. Alongside these theoretical debates and reflections, communities and collectives that explore and practice new ways of learning and understanding are emerging internationally. With do-it-yourself and do-it-together biology, as well as bio-hacking, bio-making, and bio-art, there is an emerging landscape of life science initiatives with parallels to the Open Source movement, which features free software principles, hacking practices, and the potential to democratise and open up academic knowledge. This paper presents initial findings from interviews with 10 Communities of Practice in this field, analysing their knowledge practices and cultures. It concludes with an outlook theorising how academia can participate in these knowledge ecosystems, focusing primarily on the role of participatory and practice-based design research.

Keywords: open science, citizen science, communities of practice, open knowledge, science-art relations, situated knowledge, situated learning

1. Introduction

a) Complex problems and the need for new ways of knowledge production

Complex problems such as climate change and biodiversity loss are among the most pressing challenges of our time. These issues are characterised by an unmanageable number of interacting factors that make objective and definitive answers as impossible as analysing the problems from a single perspective or discipline (Rittel & Webber, 1973). Because individual disciplines produce fragmentary views of reality, interdisciplinary and transdisciplinary approaches are required, integrating scientific disciplines with knowledge produced by non-scientific actors. This insight, and the consequent **need for transdisciplinary knowledge production and learning practices**, is now widely accepted in the scientific community, as reflected in the range of publications, discussions, and research on the subject, such as those compiled by Thorsten Philipp and Tobias Schmohl (2023) in *Handbook Transdisciplinary Learning.*

The **critique** of how **science produces** and **constitutes knowledge**, and thus a conception of reality, has not only been triggered by the recent realisation of anthropocentric climate change. This critical school of thought challenges the ontological separation of nature and culture as conceived by the academic disciplines of the natural and cultural sciences, calling for more holistic concepts that take into account the interaction between these spheres. But the critique goes back even further to the twentieth century, when the **exclusive authority** and **institutionally protected** way in which science **produces, legitimises, and validates knowledge** was criticised and the pure **objectivity** of science was questioned. Instead, the practice-philosophical turn, including feminist philosophers such as Donna Haraway, emphasised the role of **socio-cultural influences** and **practice** in producing knowledge. This included an understanding of "science and theory not as the opposite of practice, but as doing science and uncovering its power relations and hidden mechanisms" (Langemeyer & Zimpelmann, 2023, p. 248). Further concepts, such as that of **situated knowledge**, as described by Donna Haraway in 1988, emphasise that knowledge becomes effective in situated local contexts; it is always socially mediated and should be critically explored in relation to the context in which it is produced (Haraway, 1988).

Thus, the traditional ways of producing knowledge through academic disciplines, their respective epistemologies and exclusively within academic institutions have been increasingly challenged and concepts such as Michael Gibbons' **Mode 2 of knowledge production** (1994) have gained popularity. This mode describes an alternative concept of knowledge production that goes beyond the traditional academic framework and recognises different forms of knowledge. In contrast to the traditional Mode 1, which is hierarchical, disciplinary, homogeneous and strongly academic, with a clear separation between scientific professionals and social actors, Mode 2 is **transdisciplinary, heterogeneous,** and **anti-hierarchical**. It takes place in an **applied context** and involves **non-academic actors** in the knowledge production process (Gibbons et al., 1994) with a **rhizomatic** and **network-like structure** (Nowotny, 1999). In this context, science "does not stand outside society and distribute its gifts of knowledge and wisdom" (Gibbons et al. 1994, p. 22) nor can it be seen as an autonomous space separate from the rest of society. Rather, these spheres are interdependent; non-scientific modalities of knowledge production are combined into scientific knowledge as a result of transdisciplinary cooperation.

The concern to make the institutionalised boundaries of science more permeable has been put into practice by the **Open Science movement**, which aims to make science more **open, transdisciplinary and application-oriented** and to promote collaboration and dialogue with knowledge practices outside of science, as described in Mode 2 of knowledge production. While there is no precise definition of the term that reflects the diversity of activities and perspectives of Open Science – it is more of an umbrella term that encompasses "a multitude of assumptions about the future of knowledge creation and dissemination" (Fecher & Friesike, 2014, p. 17) – there are some characteristic commonalities of Open Science collected and structured in a 2014 literature review authored by Fecher and Friesike: first, concerns about the exclusive infrastructure of science and efforts to open up and share technological resources (e.g. collaboration platforms and tools); second, a commitment to expanding access to knowledge creation beyond professionals (e.g. citizen science); and third, considerations of alternative impact assessments (e.g. peer review); and finally, advocacy for expanded access to knowledge itself (e.g. open access, intellectual property rights) and for collaborative research (e.g. open data).

This brief overview highlights a significant shift in how we understand the knowledge landscape and the role of science. Traditionally, science and academia operated as exclusive authorities with a high degree of autonomy, serving as institutionally protected spaces for the production and validation of knowledge. However, this paradigm has evolved towards more open and collaborative models, where science actively engages with non-scientific actors and incorporates their knowledge. This new approach emphasises collaboration, reflection, and transparency in scientific practices, requiring transdisciplinary and participatory forms of research and learning practices that integrate multiple perspectives and epistemologies.

b) Practicing the debate: new ways of learning and understanding

Alongside these ongoing theoretical debates, reflections and analyses, communities and collectives are emerging around the world to explore and practice new ways of learning, understanding and being to address life science-related anthropogenic challenges. With **do-it-yourself** and **do-it-together biology, bio-making** and **bio-hacking**, but also through the integration of artistic and design approaches, there is an emerging landscape of life science-related initiatives with parallels to the **Open Source movement**, which features free software principles and hacking practices as well as the potential to **democratise** and **open up** life sciences and biotechnology (Delfanti, 2013). The Community Biotechnology Initiative considers its common purpose as follows:

> Fundamentally transform life sciences & democratize biotechnology to inspire creativity and
> improve lives by organizing life science change-makers and bioenthusiasts to build an inclusive global
> network, cultivate an accessible commons of knowledge and resources, launch community labs and
> projects, and enable local educators. (The Community Biotechnology Initiative, 2018, Statement of Shared Purpose 3.0, para. 2)

While the Community Biotechnology Initiative (2018) has already defined itself as a movement by identifying its shared purpose and values, further initiatives and collectives can be found in the broader field of creative biology, biotechnology, and ecology that build open learning environments that blur the boundaries between disciplines and actors. This includes initiatives such as the Floating ~~University~~, a NatureCulture learning site in the heart of Berlin that revolves around a water retention basin, or DIY Hack the Panke, a collective that explores Berlin's Panke River using transdisciplinary methods. New languages of practice are being tested here, a "practice that negates the institutional way of doing things, challenges the need for unified transparencies of meaning, elaborates on processes of making and thinking together and frees spaces in the city to do so collectively" (Karjevsky, 2019, abstract). These initiatives can be classified as **Communities of Practice (COP)**, defined by Etienne and Beverly Wenger-Trayer (2015) as a group of people "who share a concern or passion for something they do and learn how to do it better by interacting regularly" (p.2). In this way, these communities situate knowledge production and learning in the context of **social relations**, highlighting the **importance of communities** in the production and negotiation of knowledge. In doing so, they translate the principles of **collaboration** and **openness** into the **production and distribution of knowledge, infrastructure**, and **resources** that transcend traditional academic boundaries.

This paper is dedicated to such COP and draws on insights from interviews that were conducted with 10 COP over nine months within the Research Group Design, Diversity and New Commons at the Weizenbaum Institute and the University of the Arts in Berlin. This research group explores and develops **new forms of transdisciplinary and interdisciplinary research** that integrate different actors and knowledge cultures into the research process through **design methods**. As described, transdisciplinary and participatory forms of research and learning practices that integrate multiple perspectives and knowledge cultures are seen as essential to support new approaches to tackling societal and environmental problems.

What kind of learning and research practices and environments already exist beyond academia? What characterises their practice, and how can academia fruitfully interact with them? The paper aims to contribute to this discussion by exploring the **ecosystems** that these COP build and analysing their **knowledge cultures, (infra)structures** and **practices**. It also aims to raise awareness within academia of existing practices and communities that are outside academia and therefore may not be seen.

2. Method

To explore these questions, 10 interviews were conducted with COP. The focus was on the motivation, agenda and history behind the projects, the understanding and role of communities and networks, and practices and methods that proved valuable or challenging. The interviews lasted between 45 and 60 minutes, took place with one or two people at a time, were recorded and transcribed Clean Verbatim, and were finally coded thematically using qualitative content analysis according to Mayring (2022). The interviewees were the initiators or directors in eight cases, and members of the projects in the remaining two. This was because the groups were often self-organised in organically growing networks without strong hierarchies; there was not always one person who could be identified as the director or head of the group.

All the interviewed COP are challenging academic structures of knowledge production and dissemination by pursuing an activist agenda by opening up and democratising processes of knowledge creation and dissemination, involving multiple actors, perspectives and epistemologies in a low-threshold collaborative learning process. All communities were dedicated to the life sciences (i.e. they worked with biological or ecological knowledge, issues and materials, and related anthropocentric challenges). Representatives of the following initiatives were interviewed:

1) Open Wetlab of the Waag Futurelab: A laboratory for bio-design, bio-art and do-it-together biology in Amsterdam, the Netherlands, that explores biotechnologies and their impact on society and ecology.
2) Top e. V.: An association around an open lab in Berlin that reimagines the laboratory as a shared space and aims to demystify biology and radically lower the barrier to participation in modern bioscience, enabling everyone to innovate and explore through do-it-yourself biology.
3) WoeLab: A free laboratory for social and technological innovation in Lomé, Togo, addressing various urban issues such as waste management, as well as food and energy resources, using a collective approach.
4) Art Laboratory Berlin: A research platform in Berlin that bridges the gap between science, art, design and the public, offering various formats of citizen science.
5) Floating e. V.: A self-organised space and group that revolves around a water retention basin in Berlin, the Floating ~~University~~, where practitioners from diverse backgrounds meet to collaborate, co-create and imaginatively work towards the future.
6) Mediamatic: An art centre in Amsterdam, the Netherlands, that focuses on nature, biotechnology and art+science in an international network and organises lectures, workshops, and art projects for the general public.
7) Symbiotic Lab: A collective in Berlin that works through design and art to make ecological issues and concerns accessible, focusing on aesthetic and participatory approaches.
8) OpenLabBrussels: A lab at the Erasmus Brussels University of Applied Sciences and Arts in Belgium that regularly opens its doors to allow anyone interested in biomedical sciences to set up research projects, making biotechnology as accessible as possible to a diverse audience.

9) DIY Hack the Panke: An art-science collective that uses transdisciplinary methods to explore Berlin's Panke River with the public, co-organised with Art Laboratory Berlin.
10) MY-CO-X: A Berlin-based interdisciplinary art-science collective of artists, architects and fungal biotechnologists that enables an artistic-scientific discussion about the future social significance of fungi.

The initiatives were linked in various ways, creating a network of collaboration across national borders and institutions. For example, the citizen science project Mind the Fungi, a cooperation between the Institute of Biotechnology at the TU Berlin and the Art Laboratory Berlin between 2018 and 2020, extended scientific research with artistic and design-based research and worked on the development of new ideas and technologies for fungal and lichen-based materials (Art Laboratory Berlin, 2020). They involved various actors from the COP, such as Regine Rapp and Christian de Lutz (Directors of Art Laboratory Berlin), Vera Meyer (co-initiator of MY-CO-X, in the project in her role as Head of Applied and Molecular Microbiology at the TU Berlin), Fara Peluso (member of DIY Hack the Panke), and Alessandro Volpato (member of Top e. V.). Given this network configuration, the selection process of the communities was often guided by their mutual collaboration. Often, one interview led to another by following up on projects they had undertaken together.

Although the communities differed significantly in some respects, such as the extent to which their practice was informed by critical theory, how closely they were linked to academic structures or how strongly they pursued the idea of a laboratory, the interviews revealed some underlying common principles of their work. To identify them, the content was coded following inductive category formation based on Mayring's (2022) qualitative content analysis. The interpretative categories that emerged are outlined in the next chapter. The relationship with academia is described first followed by a more detailed discussion of the principles of their work. Finally in the conclusion, the findings are placed in a theoretical context, and an outlook is given on how these communities and academia might work together and what role participatory and practice-based design research can play in this.

3. Empirical study

a) Relationship with academia and its institutions

The relationship between the initiatives studied and the institutionalised academic world proved to be complex and ambivalent, both in terms of the focus on academic knowledge itself and the interaction with the structures of the academic system. While academic knowledge is an important basis for many of the projects, and many of the interviewees had an academic background or even worked in academia, all projects were characterised by a desire to break out of academic structures.

Frustration was expressed by the COP regarding the funding system behind academic knowledge, which directs and disciplines research and decides what counts as research or science in general and what does not: "The struggle is related to the pressure to produce results that appeal to funders and institutions [...] It's the way of what kind of output you consider valuable to be output in general" (T. Peeters, initiator of OpenLab-Brussels, personal communication, November 20, 2023). There was also criticism of the linear, categorising and structured division of work and methods required in funding project proposals, which contrasted with the open-ended and process-oriented work of the communities. Role conflicts were described in relation to both funding and the academic structures themselves, the categorisation of which “scientists” and “non-scientists” had previously failed to capture: "What defines the citizen scientist? If an artist leads a workshop through the forest to get materials for the scientists [...] does that count too?" (C. de Lutz, Co-Director of Art Laboratory Berlin, personal communication, March 11, 2024). Other organisations recognised the need for academia to differentiate itself from other methods and actors behind the term citizen science, problematising the authority to define citizen science: "Science comes from academia, but from our point of view it's do-it-yourself science, the initiative comes from yourself, so we don't have the problem of defining citizen scientists or citizen science – academia has the problem of defining it" (A. Volpato, member of Top e. V., personal communication, July 10, 2023).

Another issue related to the relationship with institutions was also the dependence on academic institutions for equipment and technology that was donated or borrowed from institutions as it exceeded the financial means of the projects and initiatives.

b) Crossing borders: transdisciplinarity and art-science collaborations

All initiatives highlighted transdisciplinary collaborations, especially those involving artistic and design approaches and methods – that is, collaborations that go beyond the use of artistic approaches to science communication (e.g. information visualisation) but are conceived as an epistemological practice. Different aspects were identified as benefits. In one respect, the ability of the arts to critically reflect and question knowledge introduces a more social and personal view of the conditions and consequences of scientific knowledge and shifts the boundaries of science to include political and activist perspectives: "Artists look at working with living matter in a fundamentally different way, their interest often goes to the ethical, whereas in science, because there are high expectations of science, it has to move forward" (L. Evers, Head of Open Wetlab and Make programme at Waag Futurelab, personal communication, November 24, 2023). The same applies to an experiential and sensual approach to issues. An experience-based and self-reflexive approach cannot be achieved with purely scientific procedures, the latter blocking access to certain types of knowledge. Artistic and design approaches have been able to make knowledge tangible and experienceable. For example, the temperature-insulating properties of fungal materials could be experienced directly through designed artefacts, which could only later be proven by scientific methods. In addition, the freedom of artistic approaches to take unconventional paths and the permission to have a personal research position to work on and explore issues was valued: "Artists are allowed to have a personal research position and way of working, whereas science is bound to more methodological ways of producing knowledge" (L. Evers, Head of Open Wetlab and Make programme at Waag Futurelab, personal communication, November 24, 2023).

Another obstacle to cooperation between the arts and sciences could be observed in differing ideas about working standards. Conflicting ideas about methods, work processes and the measurability of results can be debilitating as these factors have an impact on the funding of projects. To obtain project funding, working methods, work and project plans, and measures and milestones often have to be described and defined in detail in advance in applications, which contradicts the open, experimental and critical approach of artistic work.

The lack of institutional interfaces between science and art was cited as a further obstacle. This deficit requires a particularly high level of personal motivation and effort on the part of those involved as well as an active search for suitable actors to collaborate with: "They were completely off the radar of the scientific community because art and design publish in their journals or in their media and exhibitions" (V. Meyer, co-initiator of MY-CO-X, personal communication, November 10, 2023).

Other challenges mentioned were related to semantic issues, such as finding a common meaning of vocabulary to avoid misunderstandings and promote effective communication.

c) Co-creation, community learning and networks

The communities and networks that expressed the underlying principle of co-creation and engagement with others were seen as essential to the initiatives. Most of the initiatives were based on emerging or existing communities. These groups formed organically over time, usually based on individual projects and collaborations that led to further plans and projects, often resulting in the formation of whole ecosystems of actors. Informal bottom-up networking was successful, while top-down approaches to 'force' a community or network often failed. It was notable that the language used to describe these processes of networking and self-organisation repeatedly drew on biological and ecological terminology, reflecting the organic dynamics of the structures, agendas, and concerns that the communities were dealing with.

Diverse and non-hierarchical collectives, where the group rather than the individual comes to the fore and takes the lead, were described as enriching:

> To work in a collective where you hand over responsibility [...] suddenly everyone brings their own ideas or everyone is always thinking along with you, which suddenly opens up possibilities that weren't even on your radar before. If you only do it from one discipline, you miss things. (V. Meyer, co-initiator of MY-CO-X, personal communication, November 10, 2023)

It was also seen as valuable if the community participants themselves came from different backgrounds and were willing to combine and share these experiences. These backgrounds need not be disciplinary and could represent other dimensions of diversity, such as different age groups. Some of the initiatives, such as the SymbioticLab or DIY Hack the Panke, also extended the traditional community or actor concept to non-human actors and explored approaches to giving them a voice in their processes, for instance integrating the Spree or Panke rivers in Berlin as actors.

Within the communities, the common agenda and ways of working together were negotiated in ongoing processes of dialogue. Again, the fluid and organic nature of constant change and negotiation within the communities – as well as the testing of new constitutions – was emphasised, including spatial aspects such as the water retention basin in the Floating ~~University~~, the Spree for the SymbioticLab or the Panke River in the DIY Hack the Panke project. The constant transformation of the group and the network was concomitant with a transformation of the space and its actors.

d) Openness and accessibility: do-it-yourself and do-it-together

As the academic world is often criticised for its inaccessibility, the issue of access and openness plays an important role in the initiatives. The COP attempted to create openness by making their structures, processes and places transparent and freely accessible (i.e. welcoming to participate in the projects and join the communities), thus following a do-it-together approach. In addition, most of the COP had a strong focus on participatory activities, such as workshops, meet-ups, exhibitions, and lectures, to encourage public involvement in project discourse. Some projects, such as the Waag Futurelab, offer course formats, such as the BioHack Academy, where individuals can learn how to design, grow, and extract biomaterials using only Open Source hardware.

The do-it-yourself approach was also emphasised; many initiatives provided low-threshold access to scientific topics, contributing to the demystification of scientific knowledge:

> That's why I started this, to make biology accessible to more people. I never did a PCR [Polymerase Chain Reaction] when I was at school, and they still hardly do it. I think that's crazy. I think they should be able to do it when they're 16. (T. Peeters, initiator of OpenLabBrussels, personal communication, November 20, 2023)

The do-it-yourself approach also played a role in keeping costs down and strengthening self-empowerment and independent as well as self-directed learning: "I think that's the value of do-it-yourself biology [...] it shows you the possibilities to get involved and also if you can open something up and build it yourself, you own it" (T. Peeters, initiator of OpenLabBrussels, personal communication, November 20, 2023).

e) Situating the practice: site-specific and practice-based work

The COP were characterised by their practice-based, material-oriented, and site-specific work. On the one hand, site-specific and space-specific work expressed itself in the evaluation of the site as a place of interest, in which conditions, situatedness, and materials were salient. Common space was also seen as important; in common places, different activities can take place side-by-side, and actors could connect with and be inspired by one other: "Co-working space where people share their knowledge, this is very important to also have co-working spaces to share knowledge, not only experience but also different kind of knowledge" (M. Gapševičius, member of Top e. V., personal communication, July 10, 2023).

It was mentioned that the location should be neutral and not be primarily populated by one discipline or one institution:

> We know that universities have to follow certain structures, but what if members of a team decide to work in less closed environments? Then I think the opportunities are interesting and innovative, therefore it is possible. Of course, protocols and rules shouldn't be excluded, especially in a scientific environment and when working for the safety of everyone, but it is also worth considering those invisible structures that people create among themselves. (F. Peluso, member of DIY Hack the Panke, personal communication, November 13, 2023)

Practice played an important role for the communities as they sought to bridge theoretical discourses with material realities, emphasising the importance of working with the materials themselves to bring theories to life:

> They [the philosophers] do it in the theoretical but they haven't come back to the practical [...] in the case of Haraway I think it's really strong because I don't think she builds enough on her biological background. So sometimes when she writes about it, she's talking about something that doesn't quite make sense on the scientific side. And so, there's this interesting split. They're talking about science, but from a meta-level, and they're not really doing it anymore. [...] And the artists we've worked with, on the one hand, they're really interested in the posthumanism and new materialism [...] but they also go into the labs [...] The work they

make has to function. (C. de Lutz, Co-Director of Art Laboratory Berlin, personal communication, March 11, 2024)

Practical experience can therefore be seen as essential for community knowledge building.

4. Discussion and outlook

As described previously, in light of the complexities of our time, there is a growing demand for recognition of and collaboration with **different forms of knowledge** and **knowledge production**. This paper began by exploring reflections and concepts that challenge and extend traditional notions of scientific knowledge production, starting with the practice-philosophical turn that emphasises the roles of **socio-cultural influences practice** in producing knowledge. Concepts such **as Mode 2 of knowledge production** and the **Open Science movement** were additionally introduced. These concepts critically reflect how science produces knowledge and thus a construction of reality. They aim to open the exclusive, disciplinary, and linear structures of scientific knowledge production by seeking more **trans-disciplinary, open, anti-hierarchical, social,** and **heterogeneous modes** of **knowledge production, research** and **learning environments**. A central principle is the appreciation of other forms of knowledge beyond the academic system and bringing academic knowledge more into dialogue with non-academic actors and social contexts.

The interviews with the 10 concrete COP showed that most of these aspects applied to the lived practice of these collectives. They were characterised by **heterogeneous groups of actors** with **different knowledge backgrounds** who came together to share their knowledge in **situational contexts** beyond academia and negotiated their agendas **autonomously** and **democratically.** In doing so, they put into practice **contemporary theories of knowledge production** and tested **new forms of methodologies** that go beyond traditional academic disciplines to include, among others, artistic and design approaches and methods. By involving places, collectives, and even non-human actors such as the Spree or the Panke rivers, the communities **locate, situate**, and **negotiate** their knowledge, combining **knowledge production** with **social** and **sustainability aspects**. Their **open** and **experimental** way of working, as well as the freedom to create their own working spaces, where no paradigm has been established, was described as enriching and important for their activities. Inclusivity and accessibility were created at different levels: the **openness** and **accessibility of the communities and sites** to interested parties, **do-it-yourself** and **do-it-together** approaches as well as **artistic methods**, and the provision and publication of knowledge under **open licenses** and **protocols**.

However, the very aspects that are described as positive also pose difficulties in terms of working with academia. **Funding structures** create challenges as they do not support the open, fluid, and hybrid ways in which the COP work, instead demanding a planned methodological approach, rigid roles, measurable and evaluable outcomes, and usability of results. This puts pressure on communities to institutionalise themselves in some way. Furthermore, the **inaccessibility of academic spaces,** such as laboratories, and the lack of **recognition as an independent form of knowledge production** beyond the function of transfer and dissemination were mentioned as challenges.

How can these barriers and obstacles be overcome now that they are conscious and articulated? How can collaboration between academia and these communities be mutually fruitful, and what is the responsibility of academia to ensure that this happens?

The network-like structures of the communities form an open knowledge ecosystem that can be understood as a "metabolism of knowledge" (Meyer & Rapp, 2020), reminiscent of the rhizome metaphor of Mode 2 of knowledge production. As described, such networks cannot simply be established and designed from the top down; they require sensitivity, time and forms of cultivation, as well as personal commitment to organic growth. Academia should see itself as part of this knowledge metabolism if it is to continue to meet the demand for openness and inclusiveness. The mutual encounter should take place in a respectful dialogue at eye level, by integrating **do-it-together practices** and by leaving space for an agenda that is not framed only by academia. This also requires the creation of **opportunities for funding** and **collaboration**, as well as **access to academic facilities** such as laboratories. Academia must recognise and acknowledge that this landscape is not neutral; rather, it is shaped by power dynamics that influence various positions within these structures. It must also clarify its role by **reflecting on the standpoint** from which it speaks.

Integrating **different perspectives** and **epistemologies**, emphasising the **role of practice,** and developing new types of **transdisciplinary knowledge networks** has always been part of **design research**, as it has always interacted with other disciplines and actors and mixed their methods and knowledge cultures (Mareis, 2014). As an "emerging discipline or trans-discipline" (Joost et al., 2016, p. 7), design can build bridges between knowledge cultures and epistemologies and democratise the process of knowledge production by involving multiple actors in research through do-it-together practices. These aspects can also be found in **space offers**, such as the **Berlin Open Lab,** initiated by the University of the Arts Berlin in cooperation with the Technical University Berlin, that serves as an **experimental space for transdisciplinary research** at the intersection of **technology, society and art**, located on the Campus Charlottenburg. With its project space **Critical Inquiry + Design,** the space is a "transdisciplinary and trans-university terrain that attempts to blur the boundaries of theory and practice, academia and activism, and science and civil society" (Christensen & Conradi, 2024). In doing so it explores new forms of knowledge production that share the means and values of the communities interviewed and aims to integrate a variety of perspectives beyond academia into research, strengthening the role of practice to open up new avenues of transdisciplinary understanding. Further research builds on these aspects by exploring how practice-based design research methods can integrate critical theory and scientific knowledge cultures, making them accessible through a personalised experiential learning approach.

References

Art Laboratory Berlin. (2020). *Mind the Fungi*. https://artlaboratory-berlin.org/de/forschung/mind-the-fungi/

Christensen, M., & Conradi F. (2024). *Design/Research*. https://conradichristensen.eu/

Community Biotechnology Initiative. (2018). *Statement of Shared Purpose 3.0.* Global Community Bio Summit 2018. https://www.biosummit.org/statement-of-shared-purpose

Delfanti, A. (2013). *Biohackers: The Politics of Open Science.* Pluto Press.

Fecher, B., & Friesike, S. (2013). *Open Science: One Term, Five Schools of Thought.* RatSWD Working Paper, No. 218. Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://hdl.handle.net/10419/75332

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.* Sage Publications.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies, 14*(3), 575–599. https://doi.org/10.2307/3178066

Joost, G., Bredies, K., Christensen, M., Conradi, F., & Unteidig, A. (Eds.). (2016). *Design as research: Positions, arguments, perspectives.* Birkhäuser.

Karjevsky, G. (Ed.) (2019). *VOILÀ, IT'S NOT FLOATING AND IT'S NOT A UNIVERSITY!: A conversation between members of a non-institution*. Floating ~~University~~. https://floating-berlin.org/voila-its-not-floating-and-its-not-a-university/

Langemeyer, I., & Zimpelmann, E. (2023). Mode 2. In T. Philipp & T. Schmohl (Hrsg.), *Handbook Transdisciplinary Learning* (pp. 245-257). Transcript Verlag.

Mareis, C. (2014). *Theorien des Designs zur Einführung*. Junius Verlag GmbH.

Mayring, P. (2022). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (13. Neuausgabe). Beltz.

Meyer, V., & Rapp, R. (Eds.) (2020). *Mind the Fungi*. Universitätsverlag der TU Berlin. https://doi.org/10.14279/depositonce-10350

Nowotny, H. (1999). *Es ist so. Es könnte auch anders sein*. Suhrkamp.

Philipp, T., & Schmohl, T. (Eds.). (2023). *Handbook Transdisciplinary Learning*. Transcript.

Rittel, H.W.J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences, 4*(2), 155-169. https://doi.org/10.1007/BF01405730

Wenger-Trayner, B., & Wenger-Trayner, E. (2015, June). *Communities of practice: a brief overview of the concept and its uses.* https://wenger-trayner.com/introduction-to-communities-of-practice/

After the hype: the uncertain future of Smart Cities

Mennatullah Hendawy, Jana Baum, Anouk Cenan, Niklas Frechen, Anne Goldmann, Pauline Heger, and Christoph Bieber

Abstract: The promising vision of Smart Cities has been the driver for many projects, plans, and funding schemes across the globe within the last decade. While making cities 'smarter' has always involved some form of digital innovation in urban spaces, the term has remained open to interpretation. Having spent several years researching Smart City plans and their actual (non-)implementation, we propose that the concept of Smart Cities does not seem useful in practice, as comprehensive urban administration systems are not being established. We present observations to illustrate that we may rather be entering a 'post-Smart City era' as the crucial aspects of both 'smart' and 'city' have become blurred. Finally, we provide some perspectives as to what this new era might entail.

Keywords: Smart City, Smart Region, urban innovation, perspective, autonomy, centralisation

1. Introduction

In recent years, the concept of **Smart Cities** has permeated discourses about urban development, driven by the promise of integrating digital technologies to enhance urban management and quality of life. However, the practical utility of this concept remains under scrutiny. This article critiques the Smart City paradigm by examining its implementation and effectiveness, proposing that we may be entering a **post-Smart City era**. The necessity of rethinking the notion of a Smart City arises from the observation that the term has been widely misused, often serving as a veneer for projects that focus on technological advancement and enable greater surveillance and control rather than genuine social improvement. Our critique highlights a research gap in the existing literature on Smart Cities, which has predominantly adopted a realist epistemology focused on technological solutions without adequately addressing the **socio-political dimensions of urban digital transformation**.

Two years of dedicated collaborative research on Smart Cities have significantly enriched the foundation of this contribution, providing insights and empirical data that inform our current analysis. Our interdisciplinary collaboration and prior experience have allowed us to identify persistent trends and evolving challenges within the Smart City paradigm, offering new perspectives on what seems to work and what does not. Our ongoing investigations have revealed not only the technological potential of Smart Cities but also their socio-political complexities, guiding us to question and refine the assumptions underlying their development. Engaging in a continuous dialogue with Smart City practitioners, we began collecting data in the form of field trips and expert interviews, which helped us reconsider and expand our methodological approach. As cooperation and co-creation are already common practices of Smart City professionals, we tried to include elements of action research and living lab research into our portfolio. This background has been instrumental in shaping our approach to evaluating the impact of Smart Cities, ensuring that our analysis is grounded in a nuanced understanding of both their achievements and shortcomings. Through this informed lens, we aim to present a balanced and critically engaged discussion that contributes to the discourse on urban digital transformation.

This paper provides a critical examination of the Smart City concept, questioning its practical utility and suggesting that we may be entering a post-Smart City era. The primary research question we seek to address is: What are the practical realities and future trajectories of Smart Cities, and why might it be necessary to reconsider the conceptual framework underlying the Smart City paradigm? Ontologically, this research challenges the traditional notion of a ‘city’ as a bounded urban space by considering broader regional dynamics and decentralised urban-rural linkages. Epistemologically, it adopts a critical stance that questions the underlying assumptions of the Smart City narrative, emphasising the need to integrate social sciences to understand the complex reality of urban digital ecosystems. This perspective has the potential to influence future research and practices in urban planning and technological policymaking, marking a critical step towards reimagining our approach to Smart Cities. We conclude our reflections by designing three speculative scenarios that might shape the future beyond the Smart City discourse: (a) the evolution of Smart Cities into Smart Regions, (b) the transformation into AI Cities, and (c) the potential decline and obsolescence of Smart Cities.

2. Smart Cities: a history

Since the 1990s, the term Smart City has become relatively popular when discussing the relationship and dynamics between technology and urbanism (Sharifi et al., 2021). In various scientific and political contexts, Smart Cities were established as a concept describing how cities may use digital technologies to become smarter, better, and more functional organisms, for example, in terms of economy, mobility, or renewable energy. These sectors of digitalisation acted as a technological backdrop for specific narratives that supported the emergence of initiatives for city development, gratefully adapted by local administration officials as well as international companies and service providers who have driven the process over the last two decades (Bieber & Bihr, 2016).

In their overview, Rapp and Moebert (2023) outline four phases of Smart City development, drawing on technological trends and research literature. A more precise specification of the distinctive phases lists 'telematics' as the signature keyword of the 1980s describing the first steps of networked computing, followed by 'early digitalisation' efforts during the 1990s and 2000s with growing online communication on a global scale. A considerable leap forward was constituted by the rise of highly visible digital infrastructures, such as sensors and cameras, which contributed to the traffic and mobility narrative of Smart City development in the 2010s. Another boost resulted from broad discussions concerning the Sustainable Development Goals (SDGs), particularly discourses on climate and urbanisation, and closely connected to various energy narratives. An underlying current since the early 2000s has been the ubiquitous datafication of urban spaces (Bieber & Bihr, 2016; Hashem et al., 2016), driven by the proliferation of multiple devices and processes that produce, trace, and store huge troves of data, thus creating an Internet of Things (IoT). This development was most recently followed by the rise of urban artificial intelligence (AI) triggered by the advent of large language models and 'transformer'-architectures (Frechen et al., 2023; Weber & Ziemer, 2023).

Recently, there has been a shift towards alternative perspectives of the Smart City, with scholars from interdisciplinary backgrounds advocating for smart system interventions that are more attuned to users' needs, preferences, and comprehension, reflecting 'bottom-up' expressions of smartness (Hendawy & Da Silva, 2023). Simultaneously, concerns about issues in Smart Cities like sustainability, decision-making, security, privacy, and data operations (Sharifi et al., 2021) have been raised since the mid-2010s. This orientation toward the socio-political dimensions of urban planning and Smart City development has shifted away from more descriptive, technology-based approaches, gradually paving the way for a different set of studies focusing on specific stakeholders, their positions, arguments, controversies, and conflicts.

3. Smart Cities: ideal concept vs reality

Although 'Smart City' has become an often-heard phrase, defining this term remains somewhat difficult. The most basic concept of a Smart City could be derived from the two parts of its name: 'Smart' implies a comprehensive, intelligent, and efficient implementation of digital and technological innovations to improve the city in one way or another, be it from a citizen's or administration's point of view, or both. 'City', on the other hand, evidently refers to the idea that such implementations take place in a potentially large, closely interconnected urban space. Over the past few years, we and many other scholars have extensively discussed Smart Cities from various perspectives and across disciplines, sometimes also in a genuinely interdisciplinary setting (Augusto, 2021); we now posit that both of the aforementioned basic elements ('Smart' and 'City') are challenged in practice – continuously and without a clear path for a solution.

Regarding smartness, we earlier highlighted a fundamental flaw that is found in both the scholarly discourse and the practical implementation of Smart Cities – namely, the tendency for discussions to occur within distinct disciplinary or infrastructural silos (Frechen et al., 2023, Kitchin, 2015). This challenge is also associated with big data in Smart Cities, which tends to remain a sector-centred 'mess' (Frechen et al., 2023, p. 48). Distinct Smart City narratives and corresponding sectors (Aoun, 2013) have resulted in a loosely coupled array of silos, each containing substantial amounts of data and insufficiently connected to each other (Weber & Ziemer, 2023). We also found that notions of Smart Cities are frequently put into practice through a series of fragmented initiatives rather than cohesive, integrated strategies aimed at effectively managing both the city and its infrastructure (see also Attaran et al., 2022). For instance, many cities have initially expressed a commitment to solutions in the smart mobility sector (e.g., HafenCity/Hamburg, Kalasatama/Helsinki), while other cities have focused on smart climate considerations (e.g., Bo01/Malmö, Masdar City/Abu Dhabi). As a result, the prevailing trend in smart solutions continues to focus on one sector-specific narrative without adequately considering the interconnectedness, or lack thereof, among various other sectors. This fragmented approach in dealing with Smart Cities lacks the essential holistic and cooperative perspective necessary for the effective design and implementation of Smart Cities in the real world (Weber & Ziemer, 2023). Therefore, we have advocated for a departure from mono-disciplinary approaches and for the dismantling of silos at the urban data level, favouring instead the cultivation of innovative urban development through interdisciplinary collaboration (Frechen et al., 2023). Considering this scattered and chaotic landscape in terms of planning and data management, practical applications of the Smart City concept seem to be less smart (and less comprehensive) than the term may imply (see also Gonella, 2019).

Parallel to this, we have also noted the transition from city-centric thinking to a more comprehensive regional perspective (Goldmann & Baum, 2023). In adopting technologies developed by powerful platform and technology providers or partner cities, networks of Smart Cities take shape, easily transgressing traditional city limits. For instance, one resulting phenomenon we observe is smart urban-rural areas, where a larger Smart City spills over into the surrounding area and existing ideas and concepts are extended beyond the actual city limits into the surrounding rural areas (Matern et al., 2020). In contrast to these 'extensions of the city', we also see developments on a hyperlocal scale, where thoroughly planned and technology-heavy 'smart districts' within existing city structures are already in an experimental mode, while the overall effects for the 'host cities' remain unclear (Radulova-Stahmer, 2023, p. 177). This phenomenon is also illustrated by loosely organised smart energy communities that aim to establish self-sufficient units within small portions of the city, trying to break away from traditional infrastructures and administrative jurisdiction (Meleti & Delitheou, 2020). Thus, the traditional concept of a Smart City as one specific urban space seems to have been overtaken by a more complex intertwined phenomenon, which we further discuss in the following section.

In sum, on the one hand, we see *non-smart cities* (sector-specific data mess, infrastructure silos) and, on the other hand, *smart non-cities* (regions, urban-rural areas). In other words, we find a difference between the Smart City as a concept and the phenomenon in reality. This difference is reflected in research about digitalisation, smartification, and urbanism. Here, we aim to push the conversations around Smart Cities toward a more interconnected and holistic understanding of the existing sector-specific governance structures.

4. (Technical) decentralisation vs (administrative) centralisation?

Based on the issues presented above, there is a need for a more comprehensive dialogue that delves into the interactions between the fragmented sectors of Smart Cities and evaluates their impact on urban development. Furthermore, we argue that there has been a struggle between centralised city administrations and the increasing demand for decentralisation, supported by the advanced usage of devices and data in various urban contexts. This tension between centralisation and decentralisation can be interpreted as a reflection of the question of what drives Smart City development: administration, citizens, or technology?

As discussions about centralisation and decentralisation in the context of Smart Cities are often linked to questions about sector-specific challenges, this also applies to questions about the centre and the periphery. While the term Smart City emphasises the importance of a single city and its administration, there is another trend focusing more on Smart Territories or Smart Regions, thus signalling a shift from centralisation towards decentralisation (Matern et al., 2020). The basic idea is that cities are no longer seen as isolated entities (Amin & Thrift, 2002); instead, they are now considered within larger geographical contexts (Matern et al., 2020, p. 2061).

Even if this may sound quite simple, the implementation is not yet convincing. For example, Greco and Cresta (2017) look at the development of Smart City projects in Italy, criticising that many initiatives only focus on specific projects and not on the development of Smart Regions. In contrast, the authors would like to see a process that '[extends] the benefits and new opportunities for development [...] to the peri-urban and infra-urban territories that are physically and functionally in the area of gravitation of the main urban pole' (Greco & Cresta, 2017, p. 283). This would not only allow the regions around a city to benefit from digital progress but would also create a complex system in which the territorial advantages of small cities are utilised to benefit the region as a whole. Matern et al. (2020) emphasise that most of the work on Smart Regions only concentrates on the economic perspective for the purpose of growth and innovation. Instead, they propose a more integrated understanding of Smart Regions as 'urban-rural areas that are spatially reframed by digital technologies' (Matern et al., 2020, p. 2064; see also Radulova-Stahmer, 2023).

While political initiatives, such as those seen in Europe, have pushed more towards a Smart Region approach (European Commission, 2024; FMFA, 2024), other initiatives have focused on singular, robust Smart City centres, such as the New Administrative Capital in Egypt (Hendawy et al., in press) or Masdar City in the United Arab Emirates. In summary, the developments of centralisation and decentralisation in the Smart City sector are occurring simultaneously at a global level.

In addition to this more spatial notion of 'centre' and 'periphery', the terms 'centralisation' and 'decentralisation' are not only intricately connected concepts but are also consistently characterised by inherent conflict due to their contrasting approach to authority and decision-making (Page & Goldsmith, 1985). Centralisation refers to the concentration of decision-making and management functions within a singular central authority or system, thereby consolidating control and oversight (Çiner, 2022). Pooling resources and data in a centralised manner empowers cities to exploit economies of scale, diminish redundant efforts, and bolster their ability to adapt to shifting challenges and emergencies. Within urban management, centralisation entails consolidating diverse functions and operations, such as traffic regulation, waste management, and emergency services, into an integrated platform or structure. Alternatively, decentralisation entails shifting authority, duties, and decision-making from a central body to lower tiers of governance. This empowers local stakeholders to craft decisions that suit their specific needs, priorities, and circumstances, allowing diverse communities, entities, organisations, or sectors to tackle their localised issues.

It is often assumed that in democratic societies, administrative structures and decision-making processes tend to be decentralised, with power distributed across multiple levels of government and agencies. In contrast, administrative structures in autocratic regimes may exhibit centralised control, with decision-making authority concentrated within a small group or individual leader. As we observe the fragmented Smart City developments, the decentralisation of infrastructure can be perceived as being heavily tied to citizens wanting more autonomy from the original city. However, the question remains: would decentralisation via digital technologies in Smart Cities denote more autonomy for citizens? Therefore, to think about a feasible and realistic implementation of Smart Cities, one must balance efficiency and autonomy in the face of decentralisation trends. Originally, the concept of digital technologies (and the internet) was often intended to promote decentralisation and autonomy (see Benkler, 2006; Hafner & Lyon, 1998; Shirky, 2008; Turner, 2006). Yet, it has been demonstrated that they can be either misused to exert greater control or can contribute to increased control through the dominance of large digital monopolies (Lovink, 2022; Morozov, 2011; Zuboff, 2018). While this stark contrast between diverging theoretical perspectives might fuel future discussions about the level of (de-)centralisation in Smart Cities, it also bears the potential to destroy the conceptual framework.

5. The benefit of the Smart City hype

Although we provocatively question Smart City as a useful conceptual envelope, we must note that while the concept may not be entirely realistic, it definitely can serve as a catalyst for transformative activities within urban landscapes. As mentioned earlier, we must differentiate between Smart City as a concept and Smart City as a phenomenon. Despite the potential challenges associated with actually implementing the building blocks of Smart Cities, the popularity of the term has brought several benefits. For instance, we have seen that the label of a Smart City and the vision to create them has not only incentivised municipalities to take action but has also sparked a redirection of initiatives and activities in the vast field of urban innovation. As a conceptual trend, Smart Cities have raised awareness about the importance of leveraging technology to improve urban living conditions (Bieber & Bihr, 2016). This raised level of awareness has encouraged innovation and investments in related technologies and infrastructure. Additionally, the Smart City hype has fostered collaboration among various stakeholders, including government agencies, businesses, and citizens, leading to the development of partnerships and initiatives aimed at addressing urban challenges. Furthermore, the attention garnered by Smart Cities has attracted talent and expertise to the field, driving research and development efforts aimed at finding solutions to complex urban issues. Overall, while the practicality of the Smart City concept may be questioned, the hype surrounding it has spurred positive advancements and discussions in urban development. From a scientific point of view, the many – sometimes productive, sometimes erratic – activities in Smart City-oriented research have led to manifold perspectives on a diverse field. Although there will be no coherent scientific community structured by well-built and specialised academic institutions any time soon, many contacts and discussions across a broad set of disciplines have originated during this process. In this regard, the complexity of the topic and the dynamics of technological progress and innovations contribute to a slow but steady alignment across disciplinary boundaries.

6. (How) will the concept of Smart Cities survive?

Currently, we are at a point where the Smart City concept is fading into various other perspectives. Within this process, the future of the Smart City remains uncertain, reminding us of the importance of the central factors: the development of technology, the political response, and the role of the public. But still, one could ask for alternative scenarios. Instead of a conclusion, we will sketch out three possible futures of digitally enhanced cities:

1) Smart Cities could turn into Smart Regions: Smart City cooperations could turn out to be effective as Smart Regions that overcome the traditional city. Such regions are 'spatially reframed by digital technologies and the respective social practices in a variety of fields (citizenship, governance, economy, environment, mobility, infrastructure) on a discursive, implemental and regulative level' (Matern et al., 2020, p. 2064). It is difficult to predict how this development would come into effect as various technological, environmental, demographical, and political factors blend into each other. To a certain extent, this process would echo the rise of ever-growing 'Megacities' as a popular model of global civilisation. On the other hand, the process might question the duality between 'centre' and 'periphery' that is often translated into the rivalry between 'city' and 'region'. A possible convergence is an intriguing outlook from both an empirical and conceptual perspective.
2) Smart Cities could develop into AI Cities: The discussion and narratives on Smart Cities might be absorbed by or integrated into the next wave of digitisation, that is, AI. The amount of digital data in urban environments collected from various IoT devices, as well as from sensors and cameras, is continuously increasing. Thus, AI tools could routinely be used to process and analyse these enormous troves of information and might also contribute to overcoming the segregated data silos derived from traditional Smart City-development projects. Narratives of 'Urban AI' would then function as a key ingredient in new

paths of urban development, offering AI-driven features and functions such as e-mobility, smart grid, urban climate monitoring, engineering, and more. As Popelka et al. (2023) note, the specific nature of AI Cities as hybrid settings characterises this field of research:

> Digital urban AI systems have intrinsic linkages with physical urban systems, be it through the implementation of autonomous robotics that operate in urban space or through the infrastructure of the urban sensor systems that feed data to many urban AI platforms or by virtue of the fact that the results of urban AI platforms play out in urban space, affecting residents, as well as the built and natural environment. (p. 14)

This could keep the current debates and projects of Smart Cities alive, just with a new spin.

3) The Death of Smart Cities: After commercial investments have ceased and funding periods have ended (and failed to be renewed), these once popular initiatives could slowly dry out and die. Staff would be laid off or transferred to other units, software would not be updated, data silos would become outdated, and these spaces would turn into 'digital lost places'. As most of the research focusing on Smart Cities projects a certain yet unspecified future, there is not much literature about a possible decline or termination of such projects. Nevertheless, a more cultural studies-based perspective could help explore this possible development. Mattern (2021) specifies a broad landscape of 'urban media' in which Smart Cities can be integrated at a certain stage of development. Additionally, a 'failed' Smart City could be understood as a new form of 'digital ruin', a term applied by Garcia and Miller (2021, p. 163) to online environments like 'Second Life' or the digital landscapes of computer games.

All three scenarios suggest that researching the evolution and potential futures of the Smart City concept, particularly in the context of transitioning to a post-Smart City era, will require a multifaceted methodological approach. At this point, while assuming possible futures of the Smart City, we recommend using scenario analysis as an instrument to further explore the transformation into Smart Regions, the development of AI Cities, and the potential decline of Smart Cities. Constructing detailed narratives about the future based on current trends, emerging technologies, and potential policy decisions – for example, by organising a Delphi study with a diverse set of experts – would help understand different trajectories based on technological, economic, and socio-political variables.

In this article, we have cautiously followed the shifting landscape of Smart Cities. As a closing note, we would like to highlight the need to move beyond siloed discussions and fragmented Smart City solutions (replicating siloed urban administration structures). Instead, a more interconnected, interdisciplinary, and wider-scale approach to urban development should be embraced – while accepting that Smart Cities will always be a work in progress.

References

Amin, A., & Thrift, N. (2002). *Cities: Reimagining the urban*. Polity.

Aoun, C. (2013). The smart city cornerstone: Urban efficiency. *Schneider Electric White Paper*, *1*, 1–13.

Attaran, H., Kheibari, N., & Bahrepour, D. (2022). Toward integrated smart city: A new model for implementation and design challenges. *GeoJournal*, *87*(S4), 511–526. https://doi.org/10.1007/s10708-021-10560-w

Augusto, J. C. (Ed.). (2021). *Handbook of Smart Cities*. Springer International Publishing. https://doi.org/10.1007/978-3-030-69698-6

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Bieber, C., & Bihr, P. (2016). *Digitalisierung und die Smart City. Ressource und Barriere transformativer Urbanisierung*. WBGU.

Çiner, C. U. (2022). Centralization and Decentralization. In A. Farazmand (Ed.), *Global Encyclopedia of Public Administration, Public Policy, and Governance* (pp. 1444–1449). Springer International Publishing. https://doi.org/10.1007/978-3-030-66252-3_2451

European Commission. (2024). *Smart Regions*. https://ec.europa.eu/regional_policy/policy/communication/ partnership-with-media/smart-regions_en

FMFA. (2024). *Digital vernetzt auf dem Land*. Federal Ministry of Food and Agriculture, Germany. Digital Vernetzt Auf Dem Land. https://www.bmel.de/DE/themen/laendliche-regionen/digitales/smarte-landregionen/smarte-landregionen_node.html

Frechen, N., Heger, P., Bieber, C. & Hendawy, M. (2023). "Making Sense of the Big Data Mess. Why Interdisciplinarity Matters in Smart Cities." *Easy Social Sciences* 69, 45-52. https://doi.org/10.15464/EASY.2023.12.

Garcia, G. C., & Miller, V. (2021). Digital Ruins: Virtual Worlds as Landscapes of Disconnection. In G. C. Garcia & V. Miller, *Disentangling* (pp. 163–188). Oxford University Press. https://doi.org/10.1093/oso/ 9780197571873.003.0008

Goldmann, A., & J. Baum. (2023). "Smart Cities in Metropolitan Regions – Analyzing Strategies and Funding Schemes in the Ruhr Area, Germany." 6. ICPP, Panel "Science, Technology and Innovation Policies in Metropolitan Regions", Toronto, June.

Gonella, F. (2019). The Smart Narrative of a Smart City. *Frontiers in Sustainable Cities*, *1*, 9. https://doi.org/10.3389/frsc.2019.00009

Greco, I., & Cresta, A. (2017). From SMART Cities to SMART City-Regions: Reflections and Proposals. In O. Gervasi, B. Murgante, S. Misra, G. Borruso, C. M. Torre, A. M. A. C. Rocha, D. Taniar, B. O. Apduhan, E. Stankova, & A. Cuzzocrea (Eds.), *Computational Science and Its Applications – ICCSA 2017* (Vol. 10406, pp. 282–295). Springer International Publishing. https://doi.org/10.1007/978-3-319-62398-6_20

Hafner, K., & Lyon, M. (1998). *Where wizards stay up late: The origins of the Internet.* Simon &Schuster.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748–758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Hendawy, M., & Da Silva, I. F. K. (2023). Hybrid Smartness: Seeking a Balance Between Top-Down and Bottom-Up Smart City Approaches. In R. Goodspeed, R. Sengupta, M. Kyttä, & C. Pettit (Eds.), *Intelligence for Future Cities. CUPUM 2023* (pp. 9–27). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-31746-0_2

Hendawy, M., Riad, R., & Elgredly, S. H. (in press). Visual politics of a mediatized urban age: Tracing what the press news makes visible about urban planning—Case of the new administrative capital city in Egypt. *Nature Cities.*

Kitchin, R. (2015). Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society*, *8*(1), 131–136. https://doi.org/10.1093/cjres/rsu027

Lovink, G. (2022). *In der Plattformfalle: Plädoyer zur Rückeroberung des Internets* (P. Ilyes & J. S. Theodor, Trans.). transcript.

Matern, A., Binder, J., & Noack, A. (2020). Smart regions: Insights from hybridization and peripheralization research. *European Planning Studies*, *28*(10), 2060–2077. https://doi.org/10.1080/09654313.2019. 1703910

Mattern, S. (2021). *A City Is Not a Computer: Other Urban Intelligences* (Vol. 2). Princeton University Press. https://doi.org/10.2307/j.ctv1h9dgtj

Meleti, V., & Delitheou, V. (2020). Smart Cities and the Challenge of Cities' Energy Autonomy. In J. C. Augusto (Ed.), *Handbook of Smart Cities* (pp. 1–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-15145-4_50-1

Morozov, E. (2011). *The Net delusion: The dark side of internet freedom.* PublicAffairs.

Page, E., & Goldsmith, M. J. (1985). Centralisation and Decentralisation: A Framework for Comparative Analysis. *Environment and Planning C: Government and Policy*, *3*(2), 175–185. https://doi.org/10.1068/c030175

Popelka, S., Narvaez Zertuche, L., & Beroche, H. (2023). *Urban AI Guide*. Urban AI. https://doi.org/10.5281/ZENODO.7708833

Radulova-Stahmer, R. (2023). *Stadtraum im digitalen Wandel: Räumliche Auswirkungen digitaler Technologien auf Umwelt und Mobilität*. De Gruyter. https://doi.org/10.1515/9783868597875

Rapp, H. P., & Moebert, J. (2023). *Smart Cities: Investing in the digital future.* Deutsche Bank Research.

Sharifi, A., Allam, Z., Feizizadeh, B., & Ghamari, H. (2021). Three Decades of Research on Smart Cities: Mapping Knowledge Structure and Trends. *Sustainability*, *13*(13), 7140. https://doi.org/10.3390/su13137140

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. Penguin Press.

Turner, F. (2006). *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism*. University of Chicago Press.

Weber, V., & Ziemer, G. (2023). *Die digitale Stadt: Kuratierte Daten für urbane Kollaborationen*. Transcript.

Zuboff, S. (2018). *The age of surveillance capitalism: The fight for a human future at the new frontier of power* (First edition). PublicAffairs.

Can civic data be counterdata and open data? Exploring the limits of data, contestation and governance

Yuya Shibuya, Seyi Olojo, Andrea Hamm, Radhika Krishnan, and Teresa Cerratto Pargman

Abstract: The increasing surveillance by big tech companies or/and governments has raised concerns about the democratic and participatory structure of the datafied society. Meanwhile, over the course of the past decade, various bottom-up civic tech and digital civic initiatives have emerged to tackle pressing local issues, such as air pollution and disaster response, often via technology-mediated data collection, curation, analysis, design and visualisations, thus promoting democratic participation. In this article, we discuss how these data are understood in diverse contexts beyond the realm of civic tech and digital civics. In doing so, we explore the potential and limits of civic data by exploring the intersections of and differences between civic data and adjacent data-related concepts often used by civic tech communities themselves: counterdata and open data. Through our discursive exploration of these three data concepts, we conclude that understanding is limited when it comes to determining which data are 'civic', and that discussion of questions related to power structures, diversity and inclusion and infrastructuring of civic data has been minimal.

Keywords: civic data, civic tech, counterdata, open data, participation

1. Introduction

Civic tech initiatives (Harrell, 2020; Schrock, 2019) often use the Internet of Things (IoT), online survey tools and others to engage citizens to collect data with the aim of improving community services, civic engagement/participation and citizens' quality of life. Such initiatives have made various contributions to generating open data (Dunn, 2016; Shibuya et al., 2022), community building (Cerratto Pargman et al., 2018; Le Dantec & DiSalvo, 2013) and outreach (Le Dantec et al., 2011; Wehn & Evers, 2014). Such '**civic data**' – both captured and owned by the citizens for the citizens –

(Hamm et al., 2021; Maskell et al., 2018; Shibuya et al., 2022) empower people with the knowledge and resources to take action on pressing local issues. Compared with data that are passively collected by big tech and governments, civic data involve relatively small amounts of data that are inextricably linked to a place, a group of people or a purpose/issue (Maskell et al., 2018; Taylor et al., 2018). Civic data is connected to what Wells (2015) called 'civic information', which is conceptualised as 'the continuous flow of facts, opinions, and ideas that help citizens understand matters of potentially public concern and identify opportunities for action' (p. 7). Civic data is understood as not just a collection of data but rather a process of formulating what is important, for whom and why, and thus developing strong bases for action (Maskell et al., 2018; Williams, 2020). In civic tech activities, the adjacent data-related concepts, counterdata and open data, are often used to describe their data activities. **Counterdata** can be defined as the production of data that correct institutional data sources that often misrepresent communities. Such counterdata enable former data subjects to regain political power (Burrell et al., 2024; Dunn, 2016; Meng & DiSalvo, 2018). **Open data** refers to data that are complete, primary, timely, accessible, machine-processable, universally available, non-proprietary and license-free (Dawes, 2010, p. 379). It should be freely sharable and reusable (Gao & Janssen, 2022, p. 2). According to existing definitions, availability and access, reusability, redistribution and universal participation in production and distribution are the cornerstones for open data.

This article explores the intersections and differences between civic data, counterdata and open data to provide future research topics on the subject of citizen participation in the datafied society. In doing so, we ask whether these three data concepts are interchangeable (i.e. 'Can A be B?'). This format allows us to explore and stretch the boundaries of each concept and gain a deeper understanding of their limitations and potentials. Through this exploratory study, we identify a need to continuously investigate the question of which data can be considered to be 'civic' and to examine issues regarding power structures, diversity and inclusion and how to infrastructure civic data. Addressing these issues will help strengthen citizen participation in the datafied society.

2. Can counterdata be civic data?

Much of the literature documenting counterdata production initiatives reports on the efforts of activists who use data as a means of bringing attention to social issues that have been mishandled by legislative bodies. This includes the production of data to contest harmful policy decisions (Meng & DiSalvo, 2018) and to create awareness around deeply stigmatising issues, such as gender-based violence (D'Ignazio et al., 2022). The collection and sourcing of data serve as the first – and oftentimes the most important – means of effectively garnering the attention of the government when the necessary recognition was initially lacking. Therefore, the nature of counterdata production does not fit neatly into 'civic' data initiatives. Issues related to whose advocacy is readily received by governing bodies become strikingly apparent when reflecting on the conditions that lead to the production of counterdata. For example, during the COVID-19 pandemic, many communities of colour in the United States were unable to track their local mortality rates from the virus, as official statistics, collected at the city and state level, failed to provide comprehensive data on this. As many members of Black communities in the United States are already seen as second-class citizens, the invisibilisation of their experiences within official statistics reflects the lack of regard that state-funded institutions have for them as compared with other racial groups. This example asks a central question for civic data initiatives: what does participation look like for communities whose relationships with governing bodies are politically tenuous?

3. Can civic data be counterdata?

Various local civic tech initiatives can collect a wealth of data, but such data can only be beneficial if they are actually used (Alvarado Garcia et al., 2017; Kim et al., 2011). It is crucial for the data to be collected, shared and interpreted in a way that enables people to take action related to their concerns and issues. These actions may involve countering existing power structures and institutions, just as counterdata do. For example, in Stuttgart, Germany, the data collection activities of the local civic tech initiative Luftdaten (now renamed to Sensor.Community) began by monitoring air quality because the municipality had installed only two air quality data collection points in the entire city. The city's reluctance to collect air pollution information was linked to the fact that the automotive industry is the city's mainstay. Thus, to overcome the lack of air pollution information, the civic tech initiative, on its own, began installing IoT sensor devices to monitor air pollution levels. This bottom-up data collection and Luftdaten's visualisation of the data on their website became the basis for counter-narratives related to the city's reluctance to address air pollution.

However, in most cases, civic data may not be used as counterdata. Rather, many civic data initiatives work in collaboration with established entities, including local governments. For example, the core members of Safecast, the civic tech initiative that collected radiation data after the Fukushima nuclear disaster, were approached by local politicians and governments who were also eager to find missing radiation information in the communities (Hamm et al., 2021). Because the civic tech initiative and public sectors shared both a common concern and the objective of filling in the missing data for the cities, they collaborated to collect more data and disseminate the radiation information to communities. Similarly, the relationships of other civic tech initiatives with the public sector are collaborative in nature (Harrell, 2020; Le Dantec & DiSalvo, 2013; Shibuya et al., 2021). These examples prompt underexplored questions, such as how these civic data production initiatives contest and collaborate with previously underrepresented community members, such as noncitizens, indigenous populations and marginalised communities.

4. Can civic data be open?

Transparency is highly critical in civic data; thus it may seem that civic data may be intuitively understood as open data, yet this has not been always the case. For example, publishing data can be complex in cases where a need exists for independent verification if that data and information has been produced and maintained by powerful actors, such as the state or private corporations. The processes of institutional capture and institutional bypass lead to the unfettered exercise of executive power, even in functional democracies such as India in the Global South (Anderson et al., 2020; Chatterjee et al., 2019; Hansen et al., 2020; Jaffrelot, 2021; Komireddy, 2019). In addition, in some cases, opening up data comes with risks of being attacked by those who are unhappy about the publication of data. In particular, civic tech projects in non-democratic countries or underrepresented communities may face such challenges. Furthermore, some instances have difficulty in anonymizing data due to the size, quality or nature of the dataset, or people are simply unwilling to share the data. Depending on the local context and a project's focus, participants engaged in data collection can represent a narrower demographic. For example, people considered cyclist data collected in a small town in the UK to be non-representative, because the cyclist population in the town was limited and less diverse (Maskell et al., 2018). It should also be noted that infrastructure for open data is resource-consuming. There is a lack of sustainable infrastructure for creating and maintaining the data structure so that anyone can access and use it. How open civic data can exist is a complex social and technical matter that brings us to question general, technical and universal understandings of open data that too often forget about the social risks and implications of making data transparent/open.

5. Can open data be civic data?

The skewness implicit in the production of data and knowledge represents another issue (Zuiderwijk et al., 2014). For instance, in the context of the Global South, skewing can impair the production of reliable and widely recognised civic data. Therefore, the question that must be asked is not just whether civic data should be open, but also whether open data can and should be civic. In a situation where the capacities to produce civic data are mediated through socially constructed obstacles, it behoves us to recognise and acknowledge this skewing. It is also incumbent upon us to recognise the new ways civic data plays out. The production, distribution and consumption of 'civic' data in ways that could feed into and strengthen undemocratic online spaces force us to rethink simplistic relationships between civic and open data. The opening of civic data, even if it circumvents the problems highlighted earlier, must be analysed in this context. Moreover, this opening must be rooted in questions of power: what do different articulations and frameworks of data tell us about the exercise of power and sociopolitical dynamics underlying the production of data? If data are 'open', what are the conditions that could also allow it to be classified as 'civic'?

If the production of and access to civic data are seen as enablers of democratic processes and as integral components of the process of allowing hitherto marginalised voices into the decision-making sphere, then civic data will often not co-exist as open data. There are structural, technological, ethical and legal constraints that must be addressed to make the use of open data more effective (Bigagli & Nativi, 2017; Wessels et al., 2014, 2017). Open data can often act in ways that further strengthen underlying hierarchical social structures and historical inequities (Bigagli & Nativi, 2017; Wessels et al., 2014, 2017). Because open data mandates that data should be universally available to use, reuse and redistribute, this opens up space for often problematic majoritarian narratives to propagate if the data that are being used and distributed allow for the spread of widely accepted biases and prejudices (Chêne & Vrushi, 2020; Kirdemir, 2020). The means of producing and maintaining open data can allow for the undermining of myriad voices in the margin; this can produce and reproduce hegemonic narratives and, in so doing, be counterproductive to the deeper democratization of society.

6. Discussion

a) Collaboration between diverse stakeholders, including big tech companies

Despite the rising availability of easy-to-implement technological solutions to collect, analyse and visualize data, most communities may struggle to contribute to civic data. Many local communities lack the skills and resources to work with/on data. Sometimes, they need to use solutions available from big tech companies. In such cases, for example, a data partnership or agreement between said community and the big corporation is crucial. The question that remains is how local communities negotiate with big tech/government. These questions should be carefully handled because of the political hierarchy that is implied within their relations. Because of the issue of access to viable datasets, the need to work together with big tech will also be continuous. If communities decide not to use big tech solutions, civic tech initiatives themselves need to establish and maintain alternative schemes to support their collection, analysis, distribution and maintenance of data.

Data partnerships involving the state must be equally cognisant of the sociopolitical dynamics involved and the varying nature of citizen–state relationships in different contexts across the globe. In the context of the lack of functional democratic systems and protective checks and balances, data partnerships that involve the state can come with several (possibly unintended) consequences. In the deeply networked and technology-driven contemporary era, the state is clearly capable of generating, analysing and utilizing large amounts of data. However, it is equally true that civil society is increasingly discovering the potential it has to enter into this domain of information gathering and analysis. As we see the systematic weakening of legislation intended to allow for more transparent oversight of the state (e.g. the Right to Information [RTI] in India), the role of independent data gathering will assume even more significance. Although the importance of broader participation in the policy-making process has been widely acknowledged (Chatterji et al., 2019), it is also important to ensure more collaboration and more widespread availability of information. The importance of numbers and data is undeniable; good data can empower local communities and improve the process of governance (Rukmini, 2021). Data, however, hardly comprise a neutral entity and they are often the subject of intense contestations, as social and economic forces attempt to exert their hegemony (Rukmini, 2021). The terms and conditions of collaborations between local communities and big tech/government may play a crucial role in dictating the outcomes. Co-production and co-management of data regimes would require the presence of active protections for non-state actors and transparent, horizontal decision-making structures of management.

b) Maintaining ownership and understanding the context

A recent trend of passive data collection (e.g. purchasing GPS data, rather than installing data collection devices with local community participation) may run the risk of diminishing people's sense of ownership over these data (Shibuya et al., 2021). Although passive data collection offers its own set of advantages, including efficiency and scalability, it may inadvertently weaken individuals' participation levels and undermine transparency in data collection processes. Without sufficient transparency in collection, publication and maintenance procedures, the resultant data may lack the essential qualities necessary for catalysing civic action.

Moreover, it is essential to recognise that civic data collection extends beyond the mere aggregation of raw data points; it encompasses the generation of context-rich data sets (Maskell et al., 2018; Taylor et al., 2015). Such contextual information enriches the understanding and usability of the data, empowering stakeholders to derive meaningful insights and drive informed decision-making processes. Therefore, initiatives aimed at fostering civic engagement through data collection must prioritise transparency, active participation and the integration of contextual elements to maximise the utility and impact of the generated datasets.

7. Conclusion

The intersections of and differences between civic data, counterdata and open data present a complex landscape wherein various concepts and initiatives coalesce, often with overlapping objectives, yet distinct methodologies and outcomes. Although each concept embodies distinct principles and approaches, their similarities underscore the complex dynamics inherent in data-driven governance and civic engagement. Civic data have played critical roles in various communities, yet the dynamics surrounding data ownership, usage and transparency raise critical questions about the nature and potential of these data-driven approaches. For realizing the transformative potential of data in advancing democratic ideals and social justice agendas, transparency, inclusivity and contextual understanding remain imperative. Further research is necessary to understand the intersections and differences of civic data, counterdata, and open data to achieve wider societal impacts and maintain and strengthen participatory structures in the datafied society.

Acknowledgements

The work of Andrea Hamm was funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII131 (“Deutsches Internet-Institut").

References

Alvarado Garcia, A., Young, A. L., & Dombrowski, L. (2017). On making data actionable: How activists use imperfect data to foster social change for human rights violations in Mexico. *Proceedings of the ACM on Human-Computer Interaction, 1*(CSCW) (pp. 1–19). https://doi.org/10.1145/3134654

Anderson, E., & Longkumer, A. (Eds.). (2020). *Neo-Hindutva: Evolving forms, spaces, and expressions of Hindu nationalism*. Routledge.

Bigagli, L., & Nativi, S. (2017). Mobilising data: Environmental data, technical and governance issues. In B. Wessels, R. Finn, T. Sveinsdottir, & K. Wadhwa (Eds.), *Open data and the knowledge society* (pp. 121–140). Amsterdam University Press.

Burrell, J., Singh, R., & Davison, P. (2024). *Keywords of the datafied state*). http://dx.doi.org/10.2139/ssrn.4734250

Cerratto Pargman, T., Joshi, S., & Wehn, U. (2019). Experimenting with novel forms of computing: The case of the Swedish Citizen Observatory for Water Quality Conservation. *LIMITS '19: Proceedings of the Fifth Workshop on Computing within Limits)* (pp. 1–10). https://doi.org/10.1145/3338103.3338111

Chatterji, A., Hansen, T., & Jaffrelot, C. (2019). *Majoritarian state: How Hindu nationalism is changing India*. Oxford University Press.

Chêne, M., & Vrushi, J. (2020). *Getting ahead of the curve: Exploring post-Covid-19 trends and their impact on anticorruption, governance and development*. Transparency International.

Dawes, S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In M. A. Wimmer, J. L. Chappelet, M. Janssen, & H. J. Scholl (Eds.), *Electronic government. EGOV 2010. Lecture notes in computer science* (p. 6228). Springer.

Dunn, P. (2016). Open data: How access to public data is growing in Detroit. Model D. https://www.modeldmedia.com/features/open-data-090516.aspx

D'Ignazio, C., Cruxên, I., Suárez Val, H., Martinez Cuba, A., García-Montes, M., Fumega, S., Suresh, H., & So, W. (2022). Feminicide and counterdata production: Activist efforts to monitor and challenge gender-related violence. *Patterns, 3*(7), Article 100530. https://doi.org/10.1016/j.patter.2022.100530

Gao, Y., & Janssen, M. (2022). The open data canvas–Analyzing value creation from open data. *Digital Government: Research and Practice, 3*(1), 1–15.

Hamm, A., Shibuya, Y., Ullrich, S., & Cerratto Pargman, T. C. (2021). What makes civic tech initiatives to last over time? Dissecting two global cases. *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). https://doi.org/10.1145/3411764.3445667

Hansen, N. B., Klerks, G., Menendez Blanco, M., Maye, L., Strohmayer, A., De Waal, M., & Schouten, B. (2020). Making civic initiatives last: Ecosystems, technologies, approaches and challenges. *DIS '20: Proceedings of Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (pp. 433–436). http://dx.doi.org/10.1145/3393914.3395921

Harrell, C. (2020). *A civic technologist's practice guide*. Five Seven Five Books.

Jaffrelot, C., (2021). *Modi's India: Hindu nationalism and the rise of ethnic democracy*. Princeton University Press.

Kim, S., Robson, C., Zimmerman, T., Pierce, J., & Haber, E. M. (2011). Creek watch: Pairing usefulness and usability for successful citizen science. *CHI '15: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2125–2134). https://doi.org/10.1145/1978942.1979251

Kirdemir, B. (2020). *Exploring turkey's disinformation ecosystem*. Centre for Economics and Foreign Policy Studies.

Komireddy, K. S. (2019). *Malevolent republic: A short history of the new India.* Westland.

Le Dantec, C. A., & DiSalvo, C. (2013). Infrastructuring and the formation of publics in participatory design. *Social Studies of Science, 43*(2), 241–264. https://doi.org/10.1177/0306312712471581

Le Dantec, C. A., Farrell, R. G., Christensen, J. E., Bailey, M., Ellis, J. B., Kellogg, W. A., & Edwards, W. K. (2011). Publics in practice: Ubiquitous computing at a shelter for homeless mothers. *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 1687–1696). https://doi.org/10.1145/1978942.1979189

Maskell, T., Crivellaro, C., Anderson, R., Nappey, T., Araújo-Soares, V., & Montague, K. (2018). Spokespeople: Exploring routes to action through citizen-generated Data. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3173574.3173979

Meng, A., & DiSalvo, C. (2018). Grassroots resource mobilization through counter-data action. *Big Data & Society, 5*(2), Article 2053951718796862. https://doi.org/10.1177/2053951718796862

Rukmini, S. (2021). *Whole numbers and half-truths: What data can and cannot tell us about modern India*. Wetland Publications.

Schrock, A. R. (2019). What is civic tech? Defining a practice of technical pluralism. In P. Cardullo, C. Di Feliciantonio, & R. Kitchin (Eds.), *The right to the smart city* (pp. 125–133). Emerald. https://doi.org/10.1108/978-1-78769-139-120191009

Shibuya, Y., Hamm, A., & Raetzsch, C. (2021). From data to discourse: How communicating civic data can provide a participatory structure for sustainable cities and communities. *Proceedings of the 27nd ISDRS '21.* https://nbn-resolving.org/urn:nbn:de:0168-ssoar-74871-6

Shibuya, Y., Lai, C. M., Hamm, A., Takagi, S., & Sekimoto, Y. (2022). Do open data impact citizens' behavior? Assessing face mask panic buying behaviors during the COVID-19 pandemic. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-22471-y

Taylor, A S., Lindley, S., Regan, T., Sweeney, D., Vlachokyriakos, V. Grainger, L., & Lingel, J. (2015). Data-in-place: Thinking through the relations between data and community. *Proceedings of the 2015 CHI Conference* (pp. 2863–2872). https://doi.org/10.1145/2702123.2702558

Taylor, N., Clarke, L., & Gorkovenko, K. (2018). Community inventor days: Discussing data and creating speculative internet of things concepts. *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 747–758). https://doi.org/10.1145/3196709.3196772

Wehn, U., & Evers, J. (2014). Citizen observatories of water: Social innovation via eParticipation. *In Proceedings of the 2014 conference ICT for Sustainability* (pp. 1–10). Atlantis Press.

Wells, C. (2015). *The civic organization and the digital citizen: Communicating engagement in a networked age.* Oxford University Press.

Wessels, B., Linde, P., Mazzetti, P., Nativi, S., Riley, S., Smallwood, R., Taylor, M. J., Tsoukala, V., Wadhwa, K., & Wyatt, S. (2014). Issues in the development of open access to research data. *Prometheus: Critical Studies in Innovation, 32*(1), 49–66. http://dx.doi.org/10.1080/08109028.2014.956505

Wessels, B., Finn, R. L., Wadhwa, K., & Sveinsdottir, T. (2017). Navigating legal and ethical frameworks. In B. Wessels, R. L. Finn, K. Wadhwa, & T. Sveinsdottir (Eds.), *Open data and the knowledge society* (pp. 141–158). Amsterdam University Press.

Williams, S. (2020). *Data action: Using data for public good*. The MIT Press.

Zuiderwijk, A., & Janssen, M., (2014). The negative effects of open government data-investigating the dark side of open data. *Proceedings of the 15th annual international conference on digital government research* (pp. 147–152). https://doi.org/10.1145/2612733.2612761

Imaginaries and visions of futures

What are people's visions for the future when it comes to digital transformation and sustainability?

Josephine B. Schmitt and Samuel T. Simon

Abstract: Research exploring the intersection of digital transformation and sustainability reveals a growing recognition of the need to integrate digital technologies with environmental and social goals. However, it is essential to involve people, specifically their perspectives and experiences, in creating a shared vision of a future that fosters acceptance of decisions and inspires innovation. With this in mind, we pose the question: *What are people's visions of the future when it comes to digital transformation and sustainability?* To answer this question, we asked individuals to draw their visions for the future regarding the relationship between digital transformation and sustainability - for better or for worse. In this process, we built on various strands of research, such as those on mental models and imaginaries, and combined them with strategic foresight approaches. In addition to utopias and dystopias, the drawings conveyed concrete visions of how a sustainable and digitised life can succeed or fail. In doing so, these images invite reflection on the importance of integrating ethical considerations into the development and deployment of digital technologies.

Keywords: future foresight, digital transformation, sustainability, qualitative research, drawings

1. Introduction

Extreme heat, rain, flooding – the importance of sustainability in mitigating man-made climate change is clear. Simultaneously, everyday life is characterised by the universal use of digital applications. With every movie streamed, every smartwatch used, every file stored in Dropbox and the like, CO2 emissions increase (Statista, 2023). The tense relationship between the digital transformation of everyday life and its environmental impact emphasises the need to combine digital transformation and sustainability.[1]

Research exploring this intersection reveals a growing recognition of the need to integrate digital technologies with environmental and social goals. Studies such as those by Feroz et al. (2021), Robertsone and Lapina (2023), and Katsamakas (2022) highlight the importance of aligning digital transformation efforts with sustainability objectives to ensure economic growth, environmental stewardship, and social inclusivity, urging further research to deepen our understanding and guide actionable strategies for a sustainable digital future.

However, we must consider people, namely their unique perspectives and experiences, in these efforts; a future built on a shared vision encourages acceptance of decisions and allows for innovation. As such, we ask: *What are people's visions for the future when it comes to digital transformation and sustainability?* Studying people's visions of a future that merges digital transformation with sustainability is pivotal for several interconnected reasons. It shapes the trajectory of technological innovation towards environmentally conscious solutions, ensuring that advancements in digital technologies align with principles of sustainable development (Mulder, 2007). Research findings inform policy-making, guiding regulations that support sustainable practices and encourage the responsible use of technology (Cash et al., 2003). Understanding society's expectations for a digitally sustainable future influences consumer behaviour and drives demand for technologically advanced and environmentally friendly goods (Han, 2021; Yahya et al., 2016).

However, there is a dearth of research on such visions. Aiming to answer the question raised above, we invited people to physically draw their visions of the future of digital transformation and sustainability. With this approach, we are building on different schools of research, such as mental models (Rickheit & Sichelschmidt, 1999) and imaginaries (Taylor, 2003), and combining them with strategic foresight approaches (Cuhls, 2003). These approaches offer valuable insight into how individuals and societies conceptualise potential futures, providing a foundation for strategic planning and decision-making.

1 We define sustainability as the ability to meet the needs of the present without compromising the ability of future generations to meet their own needs. In this article we refer to environmental sustainability, which involves managing natural resources to prevent their depletion and degradation and to ensure the long-term health of the Earth's ecosystems (Brown et al., 1987).

2. Researching the digital future

a) Perceptions of the interrelatedness of digital transformation and sustainability

While there are myriad studies exploring perceptions of digital transformation and media change (e.g., P. Müller, 2016) and attitudes towards climate science communication (e.g., Sarathchandra & Haltinner, 2020) and climate change risk perception (e.g., Hoffmann et al., 2022), there is a lack of studies focusing on **future visions** of the **interrelationship** between digital transformation and sustainability. Previous research tends to focus on the organisational, technological, and policy aspects exploring the broader implications and strategies of integrating technologies with sustainability goals (e.g., Mondejar et al., 2021; Seele & Lock, 2017).

A systematic literature review by Gomez-Trujillo and Gonzalo-Perez (2021) provides a foundation for understanding the impact of digital transformation on sustainable development for companies and societies. The research indicates that businesses should enhance their digital capabilities and balance their economic, environmental, and social impacts to thrive in the digital era. The work of Feroz et al. (2021), contrarily, identifies disruptions caused by digital transformation on environmental sustainability. The authors present a framework outlining disruptions in four key areas: pollution control, waste management, sustainable production, and urban sustainability. However, both studies call for further research in this area, emphasising the importance of integrating sustainability and digital transformation.

Research using media coverage to explore the future scenarios of climate change offers a glimpse into the public's evolving views on environmental issues (Guenther et al., 2022; Guenther et al. 2023). Guenther et al. (2022) analysed all issues of four international news magazines ranging from the 1980s to 2019 which referenced climate change on the cover. They observed a transition from visions of apocalyptic climate futures to more varied and empowering reporting. The authors identified three multimodal frames: In the early stages, **global doom** narratives, illustrated by alarming, apocalyptic terms, and images, and **local tragedies**, pictures referring to the impact of climate change in certain regions, were prevalent. In recent years, a shift occurred towards a narrative placing a stronger emphasis on **sustainable futures**. Likewise, Guenther et al. (2023) discerned four narrative frameworks in their analysis of news texts from Germany, the USA, India, and

South Africa: solutions to climatic and social consequences, distant threats to humanity, economic opportunities, and distant threats to ecosystems. These frameworks outline various visions of climate futures, exhibiting minimal differences between countries in the Global North and Global South.

Transitioning from the exploration of future scenarios related to digital transformation and climate change, we encounter a critical gap in the literature regarding focused examinations of peoples' future visions of the relationship between digital transformation and environmental sustainability. We aim to fill this gap using a creative approach: letting people draw their visions. This gap further brings us to the pivotal role of foresight methods in understanding how individuals and societies envision the future.

b) Approaches to understanding future visions

Asking people about their perspectives on the future is a key component of strategic foresight, enabling organisations to navigate uncertainty and develop future-oriented strategies. To study the future, researchers and strategists employ a multifaceted set of methods, blending qualitative and quantitative techniques. These include scenario planning, which creates narratives of possible futures (Ratcliffe, 2002), the Delphi Method, leveraging expert consensus (Flostrand et al., 2020), trend analysis for forecasting based on current data (Birpınar et al., 2023), and horizon scanning to identify emerging trends and challenges (Cuhls, 2020).

Considering approaches such as imaginaries and mental models is crucially important as such methods meaningfully explore collective and individual visions of the future. Mental models are simplified ideas or concepts of individuals used to understand complex reality (Rickheit & Sichelschmidt, 1999). The models are based on experiences and assumptions influencing how people interpret situations and interact with the world (Jones et al., 2011). When mental models are shared within a group, commitment and performance are strengthened (Müller & Antoni, 2022). Socially shared visions are also referred to as imaginaries (Taylor, 2003). Imaginaries reflect and influence cultural, political, and economic values, guiding collective decision-making and behaviours. As dynamic frameworks, imaginaries evolve with societal changes, playing a critical role in social cohesion, innovation, and change (Dobbernack, 2010). While mental models are usually associated with individual cognitive understandings, imaginaries often focus on collective visions and societal constructs. The collective focus makes imaginaries particularly useful for understanding how broader cultural, social, and technological changes are envisioned and pursued. In this paper, we take particular interest in peoples' imaginaries of sociotechnical systems. According to Jasanoff (2015) sociotechnical imaginaries can be defined as "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology" (p. 6).

Our research intends to fill the existing gap by offering nuanced insights into the public's vision for a future where digital transformation relates to sustainable development, guiding inclusive and effective policies and practices.

3. Method

a) Procedure and sample

During a music festival in the summer of 2023, as part of the cultural support program, we asked people to illustrate their response to the following question: *What might a future look like that combines digital transformation and sustainability?* The study was organised as an open workshop, with the authors serving as facilitators. The workshop took place over three days during the festival from 2 p.m. to 5 p.m. It was located on the route to the main stage, and participants were free to join and leave at any time; there was no fixed schedule for participation. Festival attendees and performing artists were invited to participate either individually or in small groups as they passed by. They were informed about the project's objectives and data processing procedures, and their verbal consent was obtained before participation. Each participant was provided with a blank A3 sheet of paper that featured a questionnaire on the back and various drawing materials, such as ballpoint pens, watercolours, and glitter pens. While participants drew, the researchers maintained a respectful distance but remained nearby to answer any questions.

Studies that explore foresight as a dynamic method of planning and creating desirable futures highlight the distinct advantage of visual methods in articulating and sharing visions of the future (Cuhls, 2003). Through drawing, participants can externalise their mental models, uncovering common themes and unique perspectives regarding how digital advances and sustainability efforts may or may not converge. Drawing further encourages participants to understand and negotiate their images of the future. In this way, visual and emotional levels of meaning become visible and analysable (Freeman & Mathison, 2009). Through the lens of drawings, scientists can capture the rich tapestry of human thought and emotion that shapes collective visions of the future.

As recommended (Kearney & Hyle, 2004; Mitchell et al., 2011), participants were asked to describe their drawing to express what it is supposed to convey. This sets a context for the interpretation of each image. In total, 109 people between the ages of 6 and 42 (*M*=28.94, *SD*=5.27) participated in the study; 58 people identified as female, 42 as male, and 6 as non-binary. In total, 109 pictures built the basis for data analysis.

b) Analysis strategy

The analysis was carried out inductively, adapting the framework proposed by Freeman and Mathison (2009). First, we employed a *thematic analysis* to identify recurring themes, symbols, and motifs (Braun & Clarke, 2006). Thematic analysis involves coding and categorising the visual content into broader themes related to digital transformation and sustainability. We included the participants' descriptions of the pictures. In a constant comparative approach, we further determined the extent to which the drawings display common themes and cross-references. Drawings were clustered according to their themes. A cross-reference between clusters was considered to exist if similar motifs were taken up in different clusters or if motifs could be assigned to different themes. Second, we focused on image creation, conducting a *visual analysis* – following Rose (2001) – of the graphic elements of the drawings, looking for patterns that indicate common concerns, hopes, and expectations among participants. Third, another iteration of interpretation based on drawing characteristics and themes was conducted, exploring underlying patterns beyond superficial observations to better understand the connection between digital transformation and sustainability.

Six drawings were removed from the analysis as they did not relate to the task. In adherence to recommendations by Elo et al. (2014), one researcher was responsible for analysis and the other carefully followed up on the categorisation process. Divergent opinions were continuously discussed.

4. Results

The drawings range from abstract illustrations to figurative representations and mind maps. As themes, six bigger clusters (big dark blue) with subclusters (light blue) and five smaller clusters (small dark blue) emerge. Figure 1 below illustrates an overview.

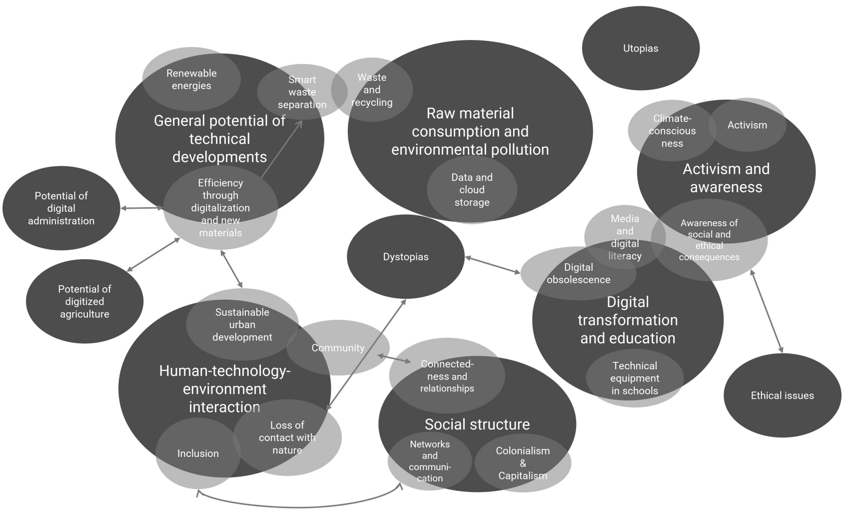

Figure 1: Thematic clusters and their cross-cuttings

Arrows and overlaps in the circles symbolise proximity and relationships between topics. Below, we take a closer look at the subclusters and illustrate the findings with selected images.

a) Characterisation of thematic clusters

Many participants address the *general potential of technical developments* for the future. They illustrate considerations of linking digital devices and the power of nature as an energy supplier: a car is powered by a bicycle (Figure 2) or a computer's plug ends in a socket in a flowerpot (Figure 3). Other images contain ideas on how to promote sustainable behaviour through digitalisation. Many images are highly schematic – some include excessive text –illustrating the complexity of different developmental steps and their impact (e.g., Figure 2).

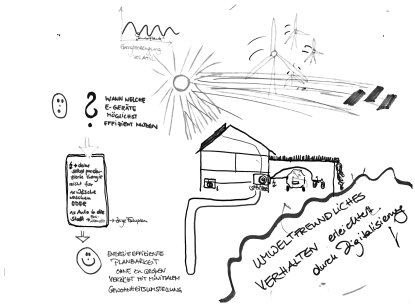

Figure 2: Digital transformation fostering environmentally friendly behaviour

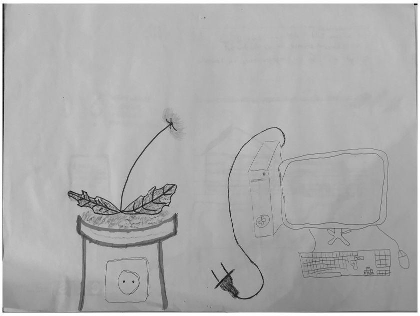

Figure 3: Renewable energies

This cluster connects to the topic of *waste and recycling,* a subcluster of the *raw material consumption and environmental pollution* cluster. We find images addressing waste generated by digitalisation (e.g., electrical waste), highlighting an often-overlooked consequence of increased reliance on digital devices. Furthermore, the images show a future in which digital technology, especially AI (Artificial Intelligence), makes waste separation and recycling more efficient and supports people's everyday lives (Figure 4, Figure 5). This aligns with current research and initiatives focusing on smart waste management systems, where AI and machine learning are applied to optimise recycling processes, reduce landfill waste, and support circular economy models (Fraunhofer IML, 2022). The visual simplicity of these images, through schematic representations and the occasional use of mind maps,[2] suggests a clarity of purpose and a straightforward approach to addressing complex issues. The restrained use of colour focuses the viewer's attention on key elements, underscoring the potential of digital innovation to contribute significantly to environmental sustainability.

2 Particularly, the mind map (Figure 5) also connects to the clusters *digital transformation and education* and *general potential of technical developments.*

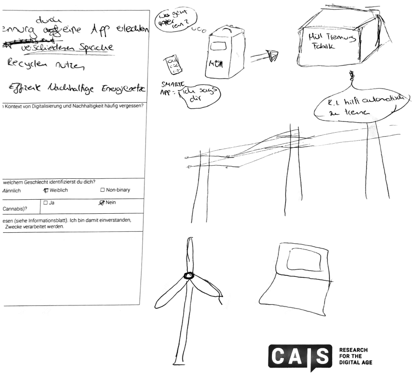

Figure 4: Waste separation

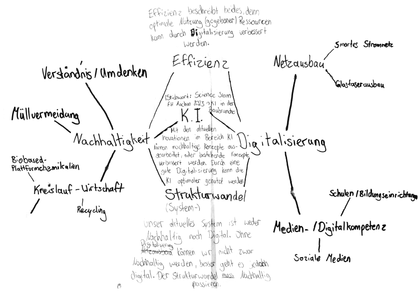

Figure 5: Mind map

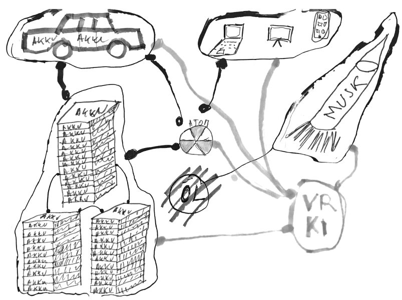

Figure 6: Server farms

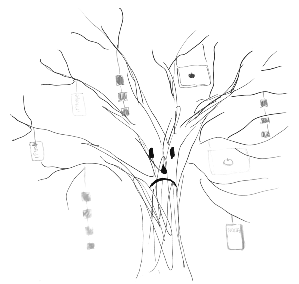

Figure 7: Sad tree

The cluster of *raw materials consumption and environmental pollution* contains another subcluster mainly consisting of images dealing with negative environmental consequences of server farms, data storage facilities, and excessive consumption of raw materials (e.g., Figure 6). The use of large storage units, laptops and smartphones in simple blue and grey tones suggests a cold, mechanised world dominated by technology, underscoring the pervasiveness of digital devices in modern life. A picture featuring a tree without leaves (Figure 7), bearing a sad face and burdened by the weight of hanging technical devices, serves as a powerful metaphor for environmental degradation caused by unchecked technological expansion. It reflects growing concerns about e-waste, the carbon footprint of digital technologies, and the loss of natural habitats due to mining the resources needed for electronic devices.

Three images can be assigned to the following topic: the *potential of digital administration*. The dominant colour in the images is the blue of a ballpoint pen. These images collectively capture the essence of digital administration's promise: eliminating unnecessary bureaucracy and modernising administrative functions. They resonate with current discussions on digital government initiatives aimed at enhancing public service delivery, increasing transparency and promoting sustainability through reduced paper use.

Images in the *potential of digitised agriculture* cluster reflect the transformative potential of digitised agriculture, illustrating how digital technologies may enhance agricultural efficiency and sustainability. They speak to the growing interest in smart irrigation systems to optimise water use, ensuring that crops receive the correct amount of water at the right time (Obaideen et al., 2022). In one image, a flower grows from a laptop with code on the screen representing the fusion of technology and nature, where digital advancements support agricultural growth (Figure 8). Another shows a field with a utility vehicle using AI to detect weeds and combat pests simultaneously (Figure 9). This reflects the ongoing discussions about precision agriculture, drones, and AI as means to achieve sustainable farming by reducing chemicals and improving crop yield (e.g., TUM Venture Labs, 2022).

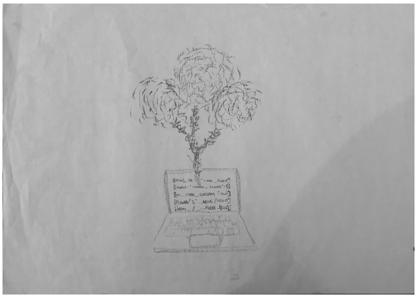

Figure 8: Flower growing

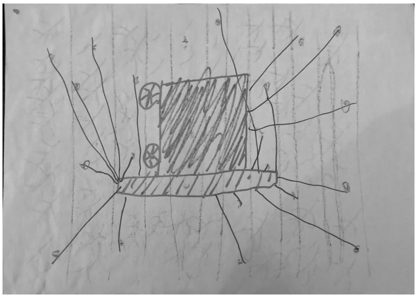

Figure 9: AI in agriculture

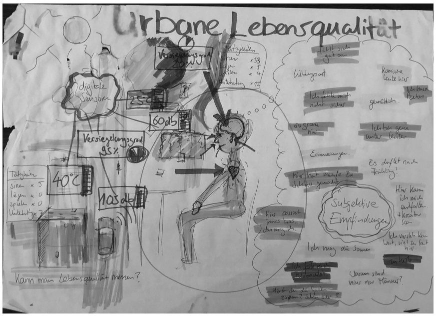

Figure 10: Quality of urban life

We can further connect the subcluster of *increased efficiency* to the subcluster of *sustainable urban development*. The presence of urban scenes intertwined with natural elements (e.g., trees, fruits) visually advocates for a future where digitalisation supports rather than detracts from ecological diversity and sustainability. The images in this subcluster do not require a great deal of text. Only one image (Figure 10) stands out in terms of combining visuals and text. Specifically, it places a person at the centre with different sensors measuring subjective sensations and quality of life (Figure 10). The complexity of the image underscores the multifaceted nature of sustainable urban development, emphasising the need for approaches balancing technological integration with personal and environmental prosperity.

This subcluster belongs to the larger theme of *human-technology-environment interaction* and includes the subcluster of *community*. Scenes depict houses closely entwined with plants. The descriptions of the pictures emphasise a desire for reorientating towards communal action in the future. The subcluster of *inclusion* represents a similar aspect, focusing on technological development that explicitly considers individuals' needs and breaks down barriers. Visually, the images convey these themes by featuring people connected by lines (Figure 12). Technological development is symbolised by laptops, smartphones or by the at sign (@) (Figure 11). Moreover, the authors included descriptive text on the picture to support their visions. Connecting these elements to current discourses reveals a growing desire and need for digital environments to be equitable, accessible, and empowering for all members of society.

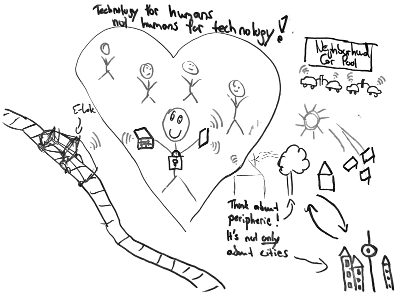

Figure 11: Humane technology

Figure 12: Everything is connected

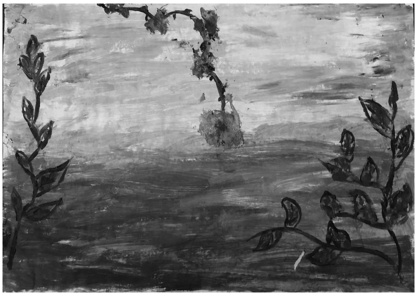

Figure 13: A world without digitalisation

However, many participants also foresee a *loss of contact with nature* and draw visions of a world *turning away from digitalisation* (e.g., Figure 13). The visual narratives, ranging from human-like figures actively rejecting digital devices to symbolic representations of disconnection underscore a yearning for an existence less digitally dominated. Such visions engage with philosophical and psychological debates about human fulfilment, the quality of social relationships, and mental health in an age where digital distractions are omnipresent (Nguyen et al., 2022). They further suggest a collective contemplation about what is lost in the race for technological advancement and, in addition, reimagine what constitutes a meaningful life. This subcluster is closely linked with dystopian perspectives.

One large cluster is titled *social structure*. Various subclusters belong to it. One of these is labelled *networking and communication* and includes three distinct images. Each of these images aims to illustrate the idea of increasingly using communication and information networks inspired by nature (e.g., mushrooms) as a model for digital communication and information processing, emphasising the importance of networked initiatives.

Together, these images underscore a movement towards leveraging digitalisation as a pivotal component of sustainable development. They encourage a rethinking of traditional structures through nature-inspired, collaborative solutions.

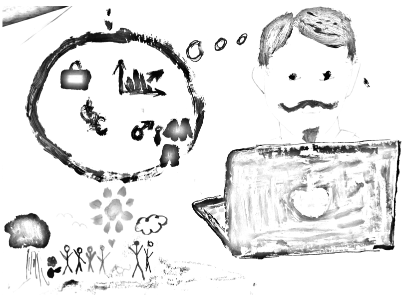

Figure 14: "Digitisation expert digitises what he thinks is right"

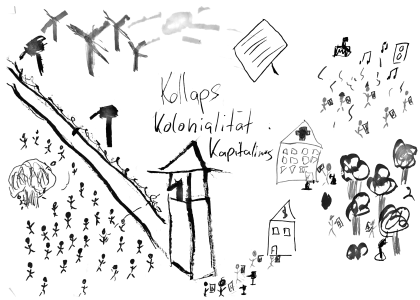

Figure 15: "Collapse, Colonialism, Capitalism"

Colonialism and capitalism form another subcluster. Images within this subcluster criticise current and future exploitation of nature and people, especially those who are marginalised due to social status, ethnic background or those suffering particularly from climate change. The images are restless. When people are depicted in these images, they are displayed in larger groups – usually as prisoners or as the exploited. The few who stand outside the groups are those possessing power and money (Figure 14, Figure 15). Power is symbolised by expensive consumer goods (e.g., Figure 14), money, and male gender. The images underscore the critique that digital transformation disproportionately benefits those with existing power and wealth, often at the expense of the marginalised and the environment. This resonates with the debate on ethical technology and responsible capitalism, calling for a re-evaluation of how digital technologies are developed and distributed (Ammanath, 2021; Financial Times, 2019).

The third subcluster, labelled *connectedness and relationships,* comprises 11 images. The images depict technical devices –usually smartphones – shaping our perception of the world and determining how we communicate and maintain relationships (e.g., Figure 16). While the image itself shows a negatively connotated approach, the explanatory texts advocate a positive perspective: harmonisation of people, nature, and technology. Two images argue in favour of keeping analogue connections with one another despite digital media. This is illustrated by direct connections between people and by abstract depiction of human-like forms and a rainbow (Figure 17), presumably referencing the rainbow flag as a symbol of diversity. This subcluster is closely linked to the subcluster of *community*.

Regarding existing technical, environmental, and social problems, various *dystopias* are drawn that depict unhealthy dependencies of people influenced by media platforms or digital technologies in general (e.g., Figure 18). One participant's last resort is the armed struggle against "the digitised" (Figure 19). These visualisations starkly contrast with optimistic visions of technology-enhanced futures, focusing on the potential for loss of control and freedom in highly digitised societies. The absence of nature-related colours and symbols, along with the predominance of garish or dark illustrations, further emphasises the disconnect from the natural world, suggesting a future where technology overshadows human agency and our connection to the environment.

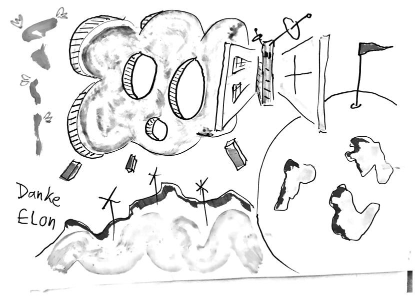

Figure 18: A world controlled by Elon Musk's satellites

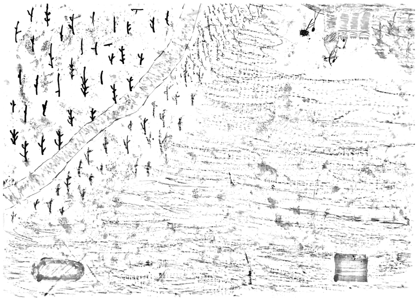

Figure 19: Fighting "the digitised"

The dystopias connect to the cluster of *digital transformation and education*. Here, we find topics like improving the *technical equipment of schools* (Figure 21) to facilitate knowledge acquisition and creativity, and a reference to *inclusion*. There are images bridging the gap between a childhood shaped by digital devices, a state of *digital obsolescence* and being trapped in social media filter bubbles, formulating an appeal for *media and digital literacy* to maintain contact with information about the environment (e.g., Figure 20). These themes align with current discourses on leveraging technology to democratise education, ensuring that all learners, regardless of their background, may access quality educational resources and opportunities. They underscore the importance of fostering media and digital literacy from a young age to empower individuals to critically engage with digital content, discern reliable information, maintain a healthy balance between digital and physical worlds, and understand how current actions impact future generations.

Figure 20: Duality: digital tools in education

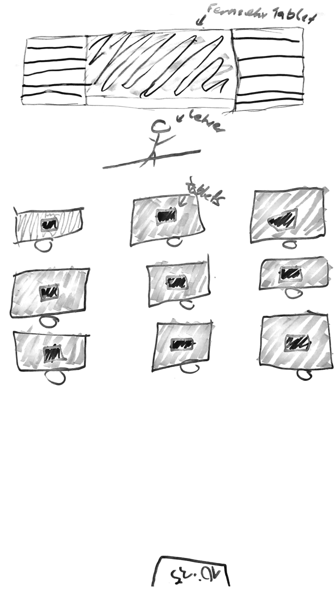

Figure 21: Classroom with tablets

This leads to the cluster of *Activism and awareness*. Images here address the need to recognise the *social and ethical consequences of digital transformation* and practice *climate consciousness*. Visually, images combine representations of technical devices and the earth emphasising the global importance of individual action. The juxtaposition of technical devices with images of the Earth serves as a potent reminder of the global stakes involved, suggesting that while digital transformation holds the power to connect and revolutionise, it also bears responsibility for the planet, requiring mindful stewardship from both individuals and collectives.

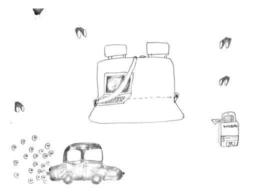

Figure 22: Danger of surveillance

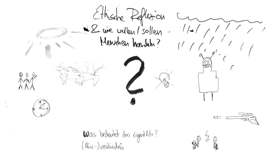

Figure 23: Advantages and disadvantages of digitisation

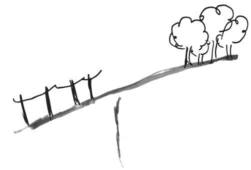

Figure 24: Where does digital transformation stop?

The cluster of *Ethical issues* raises overarching questions on digital transformation such as the danger of surveillance when digital devices are constant passengers (Figure 22). Another image explores tensions between the advantages and disadvantages of digital transformation (Figure 23). On the figurative "sunny side" of the aforementioned image, positive facets of digital transformation such as community (people holding hands), global networking (represented by connecting lines), and efficiency (symbolised by a clock) are depicted. The dark side of this image features symbols representing a lack of creativity and reflection and the takeover of power by artificial beings. In addition, two people separated by a lightning bolt represent isolation and conflict as consequences of a digitised world. The third picture (Figure 24) further explores this fragmentation. Through this illustration, the author asks, "How far does digital transformation go and where does it stop?" This is visually underscored by a seesaw: people's needs on one side considerably tip the scale to the environment's disadvantage. These pictures reflect the complexity of the current social discourse on digital transformation and its difficult entanglement with the environment. These images serve as a microcosm of larger ethical considerations that digital transformation engenders. They urge reflection on how to harness the benefits of digital technologies while mitigating adverse effects.

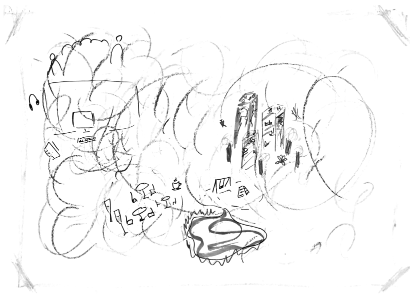

Figure 25: Humans, technology, nature: positively interdependent

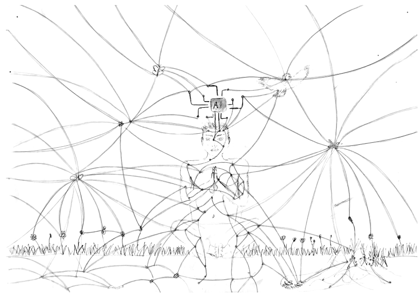

Figure 26: AI-Buddha

Both images belong to the cluster of *Utopias,* portraying the dynamic relationship between humans, digital progress, and nature in an abstract way. The utopian aspect of these depictions is the idea that in the future, humans, technological progress, and nature will influence each other positively with different needs resting in balance (Figure 25). For instance, consider the optimistic perspective that AI helps humans better understand themselves as part of the global fabric; this aligns with discussions in environmental science and technology ethics, where the potential of AI to drive sustainable practices and mitigate environmental issues is recognised. The depiction of a figure whose posture is reminiscent of Buddha and is captioned "AI" (Figure 26) might hint at the need for wisdom, mindfulness, and ethical considerations in our technological approaches. It reminds us that the path to sustainable futures requires – besides technological innovation – an ethically grounded understanding of our relationship with nature.

b) Visual interpretation

Several key patterns emerge from the images: Participants often use metaphors to link technology and nature. Consciously or unconsciously, this creates the impression that it is recognised as natural, that technological developments are integral in discussions on environment and nature. If so, this may be understood as an indicator that those depictions are close portrayals of the sociotechnical imagination of the study participants. Due to their narrative style, metaphors can serve as direct representations of imaginaries.

The images are characterised by clarity; each image focuses on either only one specific aspect or a low number of topics. These images collectively form a bigger picture. The balance between expressing distinct topics while embedding them in systemically rich contexts is achieved by using depictions of specific artefacts as symbolic and abstract representations of broader meaning. This is supported by the choice of colour, indicating either technology-associated aspects or nature-related tones for example.

One notable aspect throughout the images is the contrast of depictions of individuals or single artefacts in comparison to the illustration of collectives and networks. The first is found more often in contexts of negative meanings, whereas the latter predominately are used for representing positive aspects. This begs the assumption that the sociotechnical imaginaries represented in the drawings are a counterpoise to societal narratives of fragmentation and unhealthy hyper-individualisation.

Another dominant theme is the interplay of community and connectivity. Scenes depicting houses entwined with plants – or, as another example, people connected by lines, networks, or similar connectors – suggest a desire for communal action and technological development that prioritises human needs while leveraging a balance between physical and digital connections. Essential values for realising such a vision include social cohesion, knowledge dissemination, and technological progress that is inclusive and accessible to everyone. These are crucial for navigating the digital transformation in a way that ensures positive outcomes for current and future generations.

5. Discussion: How do the drawings relate to peoples' imaginaries?

The participants' sociotechnical imaginaries presented in the images are rich in their portrayal of how they envision the integration of technology with everyday life and the natural environment. They display a broad spectrum of expectations about technology's role in society, reflecting a diverse set of societal hopes, fears, and critical perspectives. The images allude to topics addressed in systematic literature analyses about the impact of digital transformation on sustainable development (Feroz et al., 2021; Gomez-Trujillo & Gonzalez-Perez, 2021). This shows that participants are acutely conscious of the connection between digital transformation and sustainability. These concrete visions demonstrate the creators' fluency in the topic of sustainability matters as well as the salience of these subjects in their everyday lives. It is important to note the festival at which this research was conducted has a very strong focus on sustainability; it can thus be assumed that festival visitors were highly motivated to engage with this topic.

Optimistic depictions suggest a future where technology facilitates sustainable relationships with the environment, exemplified by creative uses of renewable energy and digital advancements supporting resource efficiency. This reflects a sociotechnical imaginary that views technological progress as inherently capable of solving environmental and societal challenges. The optimistic view on the future, underscored by the strong symbolic nature of many images, aligns with the findings of Guenther et al. (2022), who observed modern media's tendency to portray future scenarios focusing on a sustainable future. The images also express concerns rooted in cautious or critical sociotechnical imaginaries. They highlight potential downsides to technology, such as environmental degradation from server farms and the perpetuation of socio-economic inequalities through digital means. These views are indicative of a growing awareness and scepticism about the unchecked expansion of digital technologies. The depiction of surveillance and privacy issues, along with dystopian visions of a world overwhelmed by digital control, visualises deep-seated fears about the loss of autonomy and privacy. These concerns are juxtaposed with images that advocate a balanced, ethical approach that harmonises human needs with environmental stewardship and technological advancement.

Given that the overarching themes which emerged from the illustrations are human-technology interactions and, more specifically, human-to-human interactions in the face of technological change, it must be emphasised that the participants expressed themselves in a way that suggests that there is untapped potential in linking the issues of digital transformation and sustainability. This does not automatically mean that this potential is not being exploited. Rather, it means that when this potential is exploited, it is invisible to those affected. Economic, scientific and political institutions must therefore work to make these efforts more visible so that they can be integrated into a widely accepted sociotechnical imaginary.

6. Conclusion

The participants' imaginaries reveal a complex tapestry of expectations about the future interactions between technology, society, and the environment. They underscore the dual potential of technology to foster sustainable and equitable futures or exacerbate existing challenges and inequities, urging a balanced approach to digital transformation. This synthesis of visual narratives and sociotechnical imaginaries emphasises the importance of collective reflection and ethical responsibility in navigating the path toward a technologically advanced society.

The pictures illustrated by participants are individual and social products. They reflect the personal interests and perspectives of their creators. They show how people explain their social existence. They underscore the need for a balanced approach to a digital transformation that both safeguards individual freedoms and promotes a harmonious relationship with the natural environment. Engaging in this discourse, future research and policymaking should aim to navigate the fine line between leveraging digital technologies for societal benefit and preventing dystopian outcomes. Future research may also employ quantitative approaches to broaden the understanding of societal perspectives on these topics. Comparative studies across different demographics and longitudinal studies could reveal diverse viewpoints and "help[s] to identify the content and contours of sociotechnical imaginaries" (Jasanoff, 2015, p. 35), enriching discussions on imaginaries integrating digital transformation with sustainability.

Through this study, we gained a better understanding of the aspects of digital transformation that people might see as risks to a sustainable future. By considering diverse perspectives, decision-makers can better develop resilient strategies in the face of uncertainty. Exploring the intersection of digitalisation and sustainability profoundly impacts digital transformation research by guiding research agendas and influencing policy development. Insights into societal expectations around this intersection further highlight a necessity for user-centred, environmentally friendly digital solutions, fostering interdisciplinary collaboration and emphasising ethical considerations.

References

Ammanath, B. (2021, November 9). Thinking Through the Ethics of New Tech…Before There's a Problem. *Harvard Business Review*. https://hbr.org/2021/11/thinking-through-the-ethics-of-new-techbefore-theres-a-problem

Birpınar, M. E., Kızılöz, B., & Şişman, E. (2023). Classic trend analysis methods' paradoxical results and innovative trend analysis methodology with percentile ranges. *Theoretical and Applied Climatology*, *153*(1), 1–18. https://doi.org/10.1007/s00704-023-04449-6

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, B. J., Hanson, M. E., Liverman, D. M., & Merideth, R. W. (1987). Global sustainability: Toward definition. *Environmental Management*, *11*(6), 713–719. https://doi.org/10.1007/BF01867238

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jäger, J., & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences*, *100*(14), 8086–8091. https://doi.org/10.1073/pnas.1231332100

Cuhls, K. E. (2003). From forecasting to foresight processes—New participative foresight activities in Germany. *Journal of Forecasting*, *22*(2–3), 93–111. https://doi.org/10.1002/for.848

Cuhls, K. E. (2020). Horizon Scanning in Foresight – Why Horizon Scanning is only a part of the game. *Futures & Foresight Science, 2*(1), e23. https://doi.org/10.1002/ffo2.23

Dobbernack, J. (2010). 'Things fall apart': Social imaginaries and the politics of cohesion. *Critical Policy Studies*, *4*(2), 146–163. https://doi.org/10.1080/19460171.2010.490637

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *Sage Open*, *4*(1), 2158244014522633. https://doi.org/10.1177/2158244014522633

Feroz, A. K., Zo, H., & Chiravuri, A. (2021). Digital Transformation and Environmental Sustainability: A Review and Research Agenda. *Sustainability*, *13*(3), Article 3. https://doi.org/10.3390/su13031530

Financial Times. (2019, December 20). The responsible capitalists. *Financial Times*. https://www.ft.com/content/a19a625c-1059-11ea-a225-db2f231cfeae

Flostrand, A., Pitt, L., & Bridson, S. (2020). The Delphi technique in forecasting–A 42-year bibliographic analysis (1975–2017). *Technological Forecasting and Social Change*, *150*, 119773. https://doi.org/10.1016/j.techfore.2019.119773

Fraunhofer IML. (2022, December 13). *Smart Waste Management*. Fraunhofer-Institut für Materialfluss und Logistik IML. https://www.iml.fraunhofer.de/de/presse_medien/pressemitteilungen/smart-waste-management.html

Freeman, M., & Mathison, S. (2009). *Researching Children's Experiences*. Guilford Publications.

Gomez-Trujillo, A. M., & Gonzalez-Perez, M. A. (2021). Digital transformation as a strategy to reach sustainability. *Smart and Sustainable Built Environment*, *11*(4), 1137–1162. https://doi.org/10.1108/SASBE-01-2021-0011

Guenther, L., Brüggemann, M., & Elkobros, S. (2022). From Global Doom to Sustainable Solutions: International News Magazines' Multimodal Framing of our Future with Climate Change. *Journalism Studies*, *23*(1), 131–148. https://doi.org/10.1080/1461670X.2021.2007162

Guenther, L., Meyer, H., Kleinen-von Königslöw, K., & Brüggemann, M. (2023). A Distant Threat? The Framing of Climate Futures Across Four Countries. *Environmental Communication*, *17*(7), 775–793. https://doi.org/10.1080/17524032.2023.2253500

Han, H. (2021). Consumer behavior and environmental sustainability in tourism and hospitality: A review of theories, concepts, and latest research. *Journal of Sustainable Tourism*, *29*(7), 1021–1042. https://doi.org/10.1080/09669582.2021.1903019

Hoffmann, R., Muttarak, R., Peisker, J., & Stanig, P. (2022). Climate change experiences raise environmental concerns and promote Green voting. *Nature Climate Change*, *12*(2), 148–155. https://doi.org/10.1038/s41558-021-01263-8

Jasanoff, S. (2015). Future Imperfect: Science, Technology, and the Imaginations of Modernity. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, 1–33.

Jones, N. A., Ross, H., Lynam, T., Perez, P., & Leitch, A. (2011). Mental Models: An Interdisciplinary Synthesis of Theory and Methods. *Ecology and Society*, *16*(1). https://www.jstor.org/stable/26268859

Katsamakas, E. (2022). Digital Transformation and Sustainable Business Models. *Sustainability*, *14*(11), Article 11. https://doi.org/10.3390/su14116414

Kearney, K. S., & Hyle, A. E. (2004). Drawing out emotions: The use of participant-produced drawings in qualitative inquiry. *Qualitative Research*, *4*(3), 361–382. https://doi.org/10.1177/1468794104047234

Mitchell, C., Theron, L., Stuart, J., Smith, A., & Campbell, Z. (2011). Drawings as Research Method. In L. Theron, C. Mitchell, C. Mitchell, A. Smith, A. Smith, & J. Stuart (Eds.), *Picturing Research: Drawing as Visual Methodology* (pp. 19–36). SensePublishers. https://doi.org/10.1007/978-94-6091-596-3_2

Mondejar, M. E., Avtar, R., Diaz, H. L. B., Dubey, R. K., Esteban, J., Gómez-Morales, A., Hallam, B., Mbungu, N. T., Okolo, C. C., Prasad, K. A., She, Q., & Garcia-Segura, S. (2021). Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet. *Science of The Total Environment*, *794*, 148539. https://doi.org/10.1016/j.scitotenv.2021.148539

Mulder, K. F. (2007). Innovation for sustainable development: From environmental design to transition management. *Sustainability Science*, *2*(2), 253–263. https://doi.org/10.1007/s11625-007-0036-7

Müller, P. (2016). Die Wahrnehmung des Medienwandels. In: Die Wahrnehmung des Medienwandels. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-12812-8_4

Müller, R., & Antoni, C. H. (2022). Effects of ICT Shared Mental Models on Team Processes and Outcomes. *Small Group Research*, *53*(2), 307–335. https://doi.org/10.1177/1046496421997889

Nguyen, M. H., Büchi, M., & Geber, S. (2024). Everyday disconnection experiences: Exploring people's understanding of digital well-being and management of digital media use. *New Media & Society*, *26*(6), 3657–3678. https://doi.org/10.1177/14614448221105428

Obaideen, K., Yousef, B. A. A., AlMallahi, M. N., Tan, Y. C., Mahmoud, M., Jaber, H., & Ramadan, M. (2022). An overview of smart irrigation systems using IoT. *Energy Nexus*, *7*, 100124. https://doi.org/10.1016/j.nexus.2022.100124

Ratcliffe, J. (2002). Scenario planning: Strategic interviews and conversations. *Foresight*, *4*(1), 19–30. https://doi.org/10.1108/14636680210425228

Rickheit, G., & Sichelschmidt, L. (1999). 1 Mental Models: Some answers, some questions, some suggestions. *Advances in Psychology*, *128*, 9–40. https://doi.org/10.1016/S0166-4115(99)80045-4

Robertsone, G., & Lapiņa, I. (2023). Digital transformation as a catalyst for sustainability and open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(1), 100017. https://doi.org/10.1016/j.joitmc.2023.100017

Rose, G. (2001). *Visual Methodologies*. SAGE Publications. https://uk.sagepub.com/en-gb/eur/visual-methodologies/book277282

Sarathchandra, D., & Haltinner, K. (2020). Trust/distrust judgments and perceptions of climate science: A research note on skeptics' rationalizations. *Public Understanding of Science*, *29*(1), 53–60. https://doi.org/10.1177/0963662519886089

Seele, P., & Lock, I. (2017). The game-changing potential of digitalization for sustainability: Possibilities, perils, and pathways. *Sustainability Science*, *12*(2), 183–185. https://doi.org/10.1007/s11625-017-0426-4

Statista.com. (2023). *Infografik: So viel Energie verbraucht das Internet*. Statista Daily Data. https://de.statista.com/infografik/26873/co2-vergleich-dsl-und-glasfasernetz

Taylor, C. (2004). 2 What Is a "Social Imaginary"?. In D. Gaonkar, J. Kramer, B. Lee & M. Warner (Ed.), *Modern Social Imaginaries* (pp. 23-30). New York, USA: Duke University Press. https://doi.org/10.1515/9780822385806-004

TUM Venture Labs. (2022). *Ressourceneinsparung durch „Precision Agriculture"*. https://www.tum.de/aktuelles/alle-meldungen/pressemitteilungen/details/ressourceneinsparung-durch-precision-farming

Yahya, W.K., Musa, N.D., Hashim, N.H. (2016). Understanding Environmentally Friendly Consumer

Behavior. In: Abdullah, M., Yahya, W., Ramli, N., Mohamed, S., Ahmad, B. (eds) Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014). Springer, Singapore. https://doi.org/10.1007/978-981-10-1458-1_82

Prefabricated futures
AI imaginaries between elitist visions and social justice claims

Annemarie Witschas

Abstract: This paper delves into the complex interplay between contemporary artificial intelligence (AI) ideologies and politics, focusing on how discourses on AI impede the prospects for social-ecological transformation. The analysis scrutinises the fabrication of the 'future' as a locus of power by influential industry actors and delineates how AI's prominent visions of the future function as a tool to reinforce the prevailing economic order while simultaneously curbing the potential for transformative struggles. To do so, I critically analyse the content and ideological backdrop of the proposed visions, as well as the discursive strategies applied. As I demonstrate, these visions not only engender exclusionary and unsustainable 'elitist futures', marked by the consolidation of power and the perpetuation of inequalities but also extend a colonial lineage of extractivist expansion, framing the future as the final frontier to be conquered. Furthermore, I argue that discursive tactics of determinism and distraction hinder both criticism and the exploration of alternatives. Observing how the visions propagated by big tech are gaining institutional traction illustrates how they are solidifying into collective socio-technical imaginaries. Hence, demystifying the industry's mechanisms of future fabrication emerges as a primary counter-strategy essential for steering toward convivial and sustainable futures.

Keywords: AI ideologies, AI imaginaries, future imaginaries, digital capitalism, digital colonialism, socio-ecological transformation

1. Introduction

Imaginaries of the future hold immense significance within emancipatory endeavours. Particularly within progressive social movements, they serve as focal points of collective aspirations and speculative discourse, often constituting the very object of struggle itself (Hawlina et al., 2020; Jovchelovitch & Hawlina, 2018; Schulz, 2016). Throughout history, the aspiration for a better life has served as a **potent motivational force for liberatory struggles**. In contemporary social movements, the utilisation of future-oriented rhetoric is most prominently exemplified by the global climate initiative *Fridays for Future*. Their rallying cry, 'Another world is possible', encapsulates the joint mobilising power that a vision of a more just and sustainable tomorrow can unfold.

However, what happens when the future itself has become territorialised and claimed by the powerful? When our collective imagination is deeply entrenched with the imaginaries, narratives, and aesthetics propagated by influential industry actors? Where teleological determinism and a rigid adherence to notions of innovation and progress stifle alternative visions and transformative possibilities? In such a scenario, the very notion of the **future is co-opted by dominant interests**, undermining the potential for meaningful societal change.

Among those wishing to claim stakes in the years and decades yet to come, the digital tech sector plays a prominent role, offering narratives ranging from productive optimism, steady progress, and technological saviourism all the way to dystopian take-over scenarios from artificial general intelligence (AGI). Regardless of the proposed scenario's valence, the tech sector itself will always play a leading role, rendering it seemingly impossible to conceive of a future that does not involve AI or other advanced digital technologies.

This paper delves into the complex interplay between contemporary AI ideologies and politics, with a focus on how prospects for social-ecological transformation are impeded. At its core, the analysis scrutinises the prefabrication of the **'future' as a locus of power** by influential industry actors and delineates how AI's prominent imaginaries of the future function as a tool to reinforce the prevailing economic order while simultaneously curbing the potential for transformative struggles. To do so, I build on the notion of socio-technical imaginary as developed by Sheila Jasanoff (2015) and the concept of 'AI futurism' by Paul Schütze (2024), as well as contributions from decolonial studies. This research is premised on the hypothesis that being able to curate and disseminate a potent future vision, one that can become hegemonic, allows actors to **maintain or promote a particular system or favourable conditions** within it. Accordingly, imaginaries of the future that are convincingly conveyed to social groups allow actors to interpret the future's uncertainty in ways that serve their particular interests. The arising conflict is that these mechanisms of fabricating futures are not only undemocratic in their genesis but also lead to **exclusionary and**

unsustainable orders once manifested. Thus, in its broader goal, this work contributes to debates regarding strategies that aim at keeping imaginative pathways open toward convivial and truly sustainable futures. Yet, before such an imaginative reconfiguration can happen, I contend, it is imperative to dissect the current mechanisms through which influential actors fabricate future imaginaries.

This paper proceeds in the following steps. First, it provides historical context by examining the practice of constructing future imaginaries and its connection to the current influence of the AI industry. Second, it conducts a critical analysis of the future visions advanced by the AI industry, scrutinising the underlying assumptions and implications that contribute to the creation of exclusionary and unsustainable 'elitist futures'. Subsequently, it explores the mechanisms through which these future visions are propagated, arguing that this process is facilitated by a deterministic view of progress and a strategy of distraction through exaggerated scenarios. It then delves into the underlying imperatives that drive future fabrication: With accumulation and growth imperatives urging the annexation of ever-new territories, the 'future' is positioned as yet another frontier to be conquered, thereby extending the colonial lineage. Finally, alternative future visions proposed by decolonial and Black, Indigenous, and People of Colour (BIPOC) movements are briefly introduced.

2. Context

This work investigates visions and elaborations of the future within the framework of ***socio-technical imaginaries***, as outlined by Sheila Jasanoff (2015). She conceives of these as 'collectively held, institutionally stabilised, and publicly performed visions of desirable futures', influenced by common beliefs about social life and order and supported by technological advancements (Jasanoff, 2015, p. 6). This perspective underscores the co-production of imaginaries along societal, institutional, and discursive axes and emphasises their political consequences. Recent work by Paul Schütze (2024) has elaborated on socio-technical imaginaries of AI through his concept of **'AI futurism',** which encapsulates 'the socio-cultural sentiment that AI systems will inexorably shape and transform the societies of the future' (p. 1). While Schütze focuses on the impacts of AI futurism on the climate crisis, this paper utilises the concept to explore how the discursive strategies employed by the AI industry limit the imagination of alternative futures.

Within this paper, the AI industry is understood as a conglomerate of companies and capital that jointly create digital technologies such as AI or big data systems. As critical scholars have noted, the term 'AI' itself does not encompass an exact technical definition, instead remaining malleable and plastic over time (Crawford, 2021). Accordingly, I follow an understanding of AI as a socio-technical 'megamachine' of 'large-scale computation' (Crawford, 2021), emphasising the deep entanglement of technology and its societal contexts, such as economic structures, values, discourses, and ways of relating. Within the context of this paper, the term AI is primarily used as a discursive marker.

Previous work on socio-technical futures has elaborated how *visions* – besides guiding actions, enabling communications, and identifying potentials for change – also allow actors to unfold a normative force, thereby 'stabilis[ing] and legitim[ising] existing power constellations' (Lösch et al., 2019, p. 76). The importance of power differentials and vested interests in these 'socio-epistemic practices' (Lösch et al., 2019, p. 72) has been further emphasised by Braun and Kropp (2023). They elaborate on how visions can evolve into socio-technical imaginaries through being 'institutionally anchored and shared by a variety of relevant social actors' like governments, businesses and other organisations (2023, p. 3). Publicly disseminated documents, such as policy reports and position papers, play a pivotal role in this transformation, serving as benchmarks by guiding opinions and decisions on technology-related policies. Lastly, *narratives* have been conceptualised as constituting the bridge 'between imaginaries and the present course of action' (Bazzani, 2023, p. 388), thereby combining imaginaries with hypothetical courses of action.

Tracing the fabrication of futures

To approach the investigation of fabricated futures, it is helpful to consider an example of how visions of the future have historically been utilised by industry actors to secure their interests. In her book *Overheated,* Kate Aronoff (2021, pp. 130-135) traces how the energy industry adeptly learnt to craft a narrative of inevitability. Since the late 1960s, Shell has grappled with mounting pressure from the Global South and alarming reports like the 1972 *Limits to Growth*, which underscored the unsustainability of fossil fuels. Seeking to confront this looming uncertainty, Shell embarked on scenario planning to envision future trajectories that would safeguard its interests. The strategies they adopted have evolved over time, from funding climate change denial (e.g., Grasso, 2019) and 'greenwashing' the company with an environmentally conscious image (e.g., Li et al., 2022) to promoting the narrative that extraction and consumer demand are unavoidable (cf. Supran & Oreskes, 2021). The motif of inevitability has remained, rooted in Shell's fundamental mission to ensure its own perpetuity and profitability, along with its resulting inability to envision a future without its presence. Companies like Shell have learnt to '[sell] the public their own strain of determinism' (Odell, 2023, p. 189), thereby **shaping public imaginaries of the future to align with their own interests**.

This vignette is more than a loose analogy; it attests to a shift in power dynamics in the recent past. In terms of profits and influence, AI companies are increasingly assuming the powerful positions previously held by the fossil industry over the course of the last century. This development has also gained attention under the business narrative of 'data being the new oil' and has been addressed by various governmental reports and scholars (Ben-Shahar, 2019; Szczepański, 2020; World Economic Forum, 2011).[1] This shift in influence becomes particularly palpable in shaping public opinion and policymaking. In the United States, for instance, tech giants have recently outrun the fossil industry in **lobbying spending**, setting new records in 2023 (Birnbaum, 2023). Similarly, in the EU, they have been fiercely advocating against the regulation of advanced AI systems and successfully influenced the content of the EU AI Act (Vranken, 2023). This growing institutional manifestation of AI industry agendas is explored further in Section 4. As the AI industry's visions gain public influence, it becomes paramount to scrutinise their impact on co-opting collective imaginaries of futures to align with corporate interests.

3. Elitist futures

Having established the role of future imaginaries as tools for consolidating power and the rising influence of the AI industry, in the following, I turn

1 Critical scholars have emphasised that digital capitalism extends rather than replaces fossil capitalism, stressing that, in fact, both industries productively cooperate and co-depend (Brevini, 2022; Taffel, 2023; see also Greenpeace, 2020).

toward a critical examination of future visions the AI industry propagates, particularly their underlying assumptions and implications.[2] As I argue, these visions are embedded into an **ideological tradition of social exclusion** that disregards vast shares of the planetary existence.

To begin, it is necessary to understand how a certain 'techno-utopian' vision of the future lies at the heart of the tech milieus in which AI is predominantly developed and funded (Torres, 2023) and is deeply entangled with the mindset and bundle of ideologies permeating these spheres. For instance, the tech elite's narrow and self-interested understanding of the future has previously been analysed quantitatively, revealing a dominant framing in 'meritocratic, self-affirming, or even self-serving terms' and discussing topics that centre around tech companies and entrepreneurs (Brockmann et al., 2021, p. 21).

The tech elite, most prominently exemplified by Silicon Valley, is becoming increasingly represented by and interwoven with the AI industry. In an investigation of the ideological underpinnings prevalent within these realms, Gebru and Torres (2024) coined the acronym **TESCREAL.**[3] This conceptual framework amalgamates a bundle of ideologies, tracing their shared origin from transhumanism toward the more recent incarnations of singularitarianism, effective altruism, and longtermism. Gebru and Torres offer a critical assessment of how these ideologies engender exclusionary worldviews, perpetuating systemic issues such as racism and even legitimising eugenic practices. While certainly the entirety of the AI industry does not hold these beliefs, they argue that these views are nonetheless emblematic of big tech circles (see also Daub, 2021). Notably, TESCREAL ideologies not only coincide with prominent AI industry representatives' beliefs but are posited as the driving impetus for companies like OpenAI and DeepMind to devise AGI in the first place (Torres, 2023). Aspects of these ideologies have further been criticised from various angles, such as their totalitarian character (von Redecker, 2023, p. 61). While a thorough elaboration of these critiques would go beyond the scope of this paper, it is necessary to understand the outlook into the future these ideologies compel.

2 An exhaustive analysis of AI's future visions, their common motifs and aesthetics, and how they manage to appeal to different groups, is highly promising but goes beyond the scope of this work.

3 Acronym for transhumanism, extropianism, singularitarianism, cosmism, rationalism, effective altruism, and longtermism.

Importantly, the futuristic discourse emanating from these spheres does not reflect a universal outlook that respects the plurality of human (and more-than-human) lived realities. Instead, concerns that imperil the sustenance, conviviality, and survival of the majority of humankind, such as the climate crisis, biodiversity loss, and escalating inequality, become peripherical or even disparaged (Torres, 2023). While there is little evidence to suggest that proponents of these ideologies are adequately engaging with these pressing real-world threats,[4] their immense wealth furthermore serves to insulate them from effectively experiencing any existential threats to their own bodies and livelihoods. This phenomenon has been debated as constituting an epistemic deficit in gaining access to shared worlds (von Maur & Slaby, 2024). This is further evidenced by the fact that 'prepping' for a civilisational collapse with supplies such as gold or even extreme-weather-resistant high-tech bunkers is a common practice in Silicon Valley and wider tech elite circles (Friend, 2016; O'Connell, 2018; Osnos, 2017; Rushkoff, 2022).

TESCREAL ideologies serve as the moral backdrop to legitimise such elitist positions. Most noteworthy, longtermism develops a moral theory through which harms experienced in the present or near future are relativised by the potential existence of hypothetical conscious beings in the far future, which scale beyond comparison to today's population (Bostrom, 2017). Transhumanism perpetuates a narrative of human development with racist undertones of 'liberal eugenics' (Agar, 1998), thereby rendering parts of the human population disposable. Additionally, cosmism contends that humanity's destiny is to be an interplanetary species (Goertzel, 2010), thus diminishing the urgency to maintain Earth's climate in liveable conditions. Consequently, the ideological tradition permeating big tech and AI circles is not in alignment with an inclusive, convivial, and sustainable existence. Instead, it propagates what can be termed **'elitist futures', catering primarily to a small cohort of the super-rich** and influential, predominantly consisting of white males.

4 For instance, in tweet posted in 2022, Elon Musk, a prominent figure in this realm, postulates that declining birthrates are 'a much bigger risk to civilization than global warming' (Musk, 2022).

4. Mechanisms of fabricating the future

In the following section, I explore the mechanisms through which these future imaginaries are propagated. I focus on two discursive strategies: a deterministic view of progress and a mechanism of distraction through exaggerated scenarios.

a) Narrowing future corridors through a deterministic view of progress

Contemporary futuristic AI debates are commonly framed against a backdrop of technological and historical determinism. In his investigation of AI futurism, Schütze (2024) contextualises the importance of notions such as tech-solutionism and tech-futurism in these discourses, which view technology, epitomised in recent years by AI, as the primary catalyst for societal progress.

However, I would like to take a step back and scrutinise the motif of determinism more broadly within the construction of inevitability and the future outlook it compels. This perspective hinges on a specific conception of temporality that portrays **history as a linear progression** while simultaneously narrowing the space in which conceivable trajectories for the future can unfold. Such a determinism views history as 'a unidirectional and inevitable march of progress that can never be questioned or redirected' (Odell, 2023, p. 184).[5] This belief in human history as primarily one of progress has been expressed by a number of leading figures in the AI industry, including former OpenAI manager Zack Kass (Fulterer, 2023). Further, Sam Altman, CEO of OpenAI, has conjectured that AGI does not redirect how the future unfolds but simply accelerates it along a supposedly predetermined path (Altman, 2023). Similarly, economist Robin Hanson contemplated how long a small group of post-apocalyptic survivors would take to 'return to our current level of development' (Hanson, 2008, p. 12).[6] In Hanson's narrative, history is read as a sequence of different phases of growth, with each phase accelerating the productivity of the previous through new kinds of 'capital'. While extreme circumstances like earthquakes and wars might 'set back' this progression, the implied hierarchy of societal stages already indicates the trajectory of development after

5 See also Hartley, 2016.
6 For a more extensive critique, see Mitchell & Chaudhury (2020).

a potential collapse: back to a US-/Euro-centric imaginary of 'Western civilisation' as the universalised pinnacle of human organisation. Such a view disregards a substantial part of humankind globally, as well as any previous societies, and degrades contemporary indigenous peoples with hunter-gatherer economies.

Furthermore, viewing history as continual linear progress, smoothing out struggles and contestation, results in a form of futuristic thinking where any **resistance appears futile** against the mechanical model of deterministic progress.[7] As Antti Tarvainen (2022, p. 1) underscores with reference to Walter Benjamin (1990), such visions of progress are 'founded upon a culture of amnesia that serves those in power', in which 'pasts of the oppressed become repressed and appropriated while the pasts of the rulers become fetishised into "universal" destinies'. Benjamin's (1990) analysis reveals how the ruling classes utilise the allure of industry and technology to perpetuate their hegemony, projecting a mythic future of peace and abundance. In this teleological reading of progress, the future becomes the primary casualty, subjugated to the dictates of an automatised and deterministic narrative.

This mythic future of peace and abundance finds embodiment in the concept of singularity. For the AI industry, it serves as the endpoint toward which the deterministic trajectory of progress converges. Yet, within this narrative, moments of disruption do emerge: For instance, Kurzweil prophesied that 'within a few decades, machine intelligence will surpass human intelligence, leading to The Singularity – technological change so rapid and profound it represents a *rupture in the fabric of human history*' (Teuscher, 2004, p. xx, emphasis added). Although disruption appears as a potent counterforce to determinism, the notion of the 'fabric of human history' nonetheless implies a universal and continuous trajectory whose redirection requires a force so powerful and rare that it is envisioned as *singular*. This stands in stark contrast to recognising history as a **non-linear, pluralistic, and dynamic** interplay of struggle, adaptation to crises, and regeneration, as BIPOC temporalities suggest (cf. Mitchell & Chaudhury, 2020).

7 Aptly summarised by this prediction attributed to Kurzweil: 'Though the Luddites might, at best, succeed in delaying the Singularity, the march of technology is irresistible and they will inevitably fail in keeping the world frozen at a fixed level of development' (Sephiroth, 2016).

b) Distraction through exaggerated scenarios

AI imaginaries of the future come in various shapes, yet they are found largely on an axis between saviourism and doomerism, with each of these extremes encompassing a cluster of proponents. The fields of AI-solutionism and **saviourism** feature shiny promises about how AI's vast possibilities will usher in a golden age, for instance, by **curing diseases, guaranteeing economic prosperity, and solving climate change** (Fulterer, 2023). Such futuristic speculations serve to keep radical demands and subversive potentials at a far enough distance. In this promised land of a singularity-like future, the full imaginative potential, which is currently kept at bay, is eventually unleashed. According to AI executives, we will have jobs we cannot yet imagine, may no longer need to work at all, and could even discard capitalism entirely (Fulterer, 2023). Conveniently, in the powerful narrative of AI futurism, boundless opportunities are realised *through* – and not *against* – the interests of the industry: There is no need to strive for collective liberation and wealth redistribution in the present, as the elimination of all scarcity is already nearing. Thus, these speculative, transformative pursuits are already territorialised within the greater AI futurist imaginary.

While certain applications of AI, such as its use in healthcare, may indeed hold some potential, critics caution that these grandiose claims not only tend to be dangerous exaggerations but also serve as part of a **larger distraction scheme**. For example, Naomi Klein (2023) contends that such promises serve as 'powerful and enticing cover stories' for what she perceives as potentially the 'largest and most consequential theft in human history'. In her critique, she refers to how big tech entities are seizing control of the entirety of human knowledge available in digital, scrapable form and enclosing it within proprietary products.

This strategy further diverts attention from AI's currently existing detrimental environmental toll and reliance on **inhumane labour conditions** (Schütze, 2024). AI's functionality and performance heavily rely on human workers to provide annotated training data, fine-tune models, and moderate results. A vast share of this labour is outsourced to the Global South and is facilitated by underpaid workers in precarious conditions (Miceli & Posada, 2022; see also van Doorn, 2017). These realities, also debated under the term 'Ghost work' (Gray & Suri, 2019), starkly contrast with the industry's vision of a labour-free future. Upon closer inspection of the industry's own forecasts, further discrepancies become apparent. For instance, recent research from Google projects a substantial increase in third-party annotation workers, reaching into the billions (Wang et al., 2022). Consequently, given the dependencies of AI systems on human feedback, the envisioned AI future without labour does not seem like a universal reality but rather appears attainable solely by a privileged minority. Likewise, the lofty promise that AI will solve climate change has been under heavy scrutiny by scholars studying the intense planetary impact of these systems, which intensify extractive mining practices, water scarcity, and carbon emissions (e.g., Brevini, 2022; Crawford, 2021).

On the other side of the spectrum, AI doomers paint a daunting image of the future. Yet, rather than addressing the grim questions of environmental and social justice, they caution against the **existential risks emanating from an unbound AGI** that sees humans as mere ends to its means (e.g., Yudkowsky & Bostrom, 2018). Understanding their broader approach within the framework of 'criti-hype', a term coined by Lee Vinsel (2021), is insightful. Essentially, such critiques mirror the exaggerated hype surrounding the topic through excessively dramatized scenarios or doomsday predictions, thereby inverting the enthusiasts' messages while maintaining the narrative of grand transformation. As a result, such narratives act as a smokescreen, diverting public attention and research resources away from AI's contested political grounds toward hypothetical realms such as AGI alignment (Bordelon, 2023; Clarke, 2023).

In conclusion, both saviourism and doomerism surrounding AI can be understood as discursive strategies that serve as flashy distractions, ultimately hindering the articulation of alternative futures. The allure of saviourism often manifests through the depiction of shiny yet unrealistic and exclusionary futures. These captivating promises overshadow the critiques of present-day harms, with proponents suggesting that these issues will be resolved once the systems have matured sufficiently. Conversely, doomerism amplifies concerns about the existential risks posed by AI, **diverting attention from more nuanced critiques**. In both cases, the focus on polarising scenarios obstructs urgent discussions about the ethical, social, and environmental implications of AI development and deployment and has been strategically wielded to fend off regulation (Marx, 2024). The two strategies I have investigated, a deterministic view of progress and the propagation of exaggerated scenarios, go hand in hand. Whereas the former strategy envisions the certainty of a given path, the latter builds onto it and fills the vision with content.

Gaining institutional traction

The agendas described above have recently gained increasing institutional traction. Embedded within a growing network of philanthropic organisations and think tanks, AI futurists administer considerable financial resources, amounting to billions of dollars in scale. Their lobbying efforts have wielded significant influence in **aligning national AI policies** in Britain (Clarke, 2023) and the US (Bordelon, 2023) with doomerist concerns of 'existential risk', as well as impeding regulations in the EU (Vranken, 2023). Tech companies have also enjoyed growing leverage in academia. For instance, the administration of research funds and dual affiliation positions **steers AI research in favourable directions** while keeping critical contributions at bay (Whittaker, 2021). By framing AGI as 'a natural progression in the field of AI', researchers and practitioners are mobilised into the field, employing the resulting 'veneer of scientific authority' to further legitimise their agenda (Gebru & Torres, 2024, p. 21). Analysing EU strategy papers on AI, Brevini (2021) found that the industry-aligned myths of inevitability, as well as the image of AI as a solution to humanity's greatest challenges, are already represented in policymaking discourses and are being further reproduced there. Expanding upon Gramsci's theory of cultural hegemony, Brevini explains how these industry-serving positions contribute to the consolidation of hegemonic structures, thereby rendering the 'conception of alternatives virtually impossible' (2021, p. 145). What Gramsci (1999, p. 630) termed 'folklore of the future', denoting a relatively rigid phase of narratives widely accepted as facts, underscores the profound impact of potent myths on the imagination of alternative futures. Observing how the AI industry's visions are becoming institutionally anchored in both policymaking and research marks their transformation from a discrete vision to a wider-ranging socio-technical imaginary (cf. Braun & Kropp, 2023). The successful construction of a hegemonic discourse thus enables the industry to exert a profound influence on shaping a widely accepted understanding of futures that safeguard its interests while diminishing environmental and social justice claims.

5. The future as the final frontier: extending a colonial continuity

Which underlying forces drive the AI industry's claims to the future? Crucially, the imperative of growth stands as a fundamental pillar of contemporary AI capitalism, mirroring broader trends within the capitalist system at large (Verdegem, 2022, p. 1). Sustaining this growth necessitates the continuous expansion into new domains to perpetuate itself, resulting in a relentless **pursuit of accumulation**. This drive for accumulation is also reflected in transhumanist endeavours. For instance, pursuits such as body modification through implants or the quest of mind uploading can be interpreted as efforts to surpass mortality as the ultimate barrier to accumulation (cf. Heffernan, 2020). Overcoming mortality, as Eva von Redecker (2021, p. 899) argues, drawing upon Arendt, becomes imperative in the pursuit of perpetual accumulation.[8]

The logic of accumulation is further interconnected with **colonial rationality** (von Redecker, 2021, p. 900). This connection has garnered recent attention under the terms of *data colonialism* (Couldry & Mejias, 2019; Thatcher et al., 2016) and *digital colonialism* (Coleman, 2019; Kwet, 2018).[9] While both concepts are inherently interwoven, *data* colonialism specifically highlights the framing of data, particularly personal data, as a fresh and exploitable resource. At the heart of this mechanism lies the process of 'capitalist accumulation by dispossession', as posited by Thatcher et al. (2016), wherein big data serves as a potent tool for perpetuating 'continual growth' within the capitalist machinery (p. 3). A central aspect of their critique is the increasing commodification and privatisation of various aspects of life, which has given rise to a new wave of frontierism. With ever more frontiers to be expanded and conquered, this relentless pursuit also does not recoil from temporal limitations. As one AI enthusiast put it, 'Replace the roar of rockets with the crunch of data, […] "boldly going" not to space but to a new final frontier: predicting the future' (Siegel, 2015, p. 23). Consequently, in adherence to the logics of continual expansion and accumulation, it becomes apparent that even the **future itself is subject to siege**.

Once-unattainable realms, whether distant planets like Mars or the nebulous terrain of the future, are now coveted as **territories ripe for appropriation.** As von Redecker (2023) argues, they become subject to colonisation and exploitation in pursuit of capitalist expansion, thereby being reduced

8 Arendt theorises a different way of transcending the individual's mortality in order to enable ceaseless accumulation. Looking at Hobbes' *Leviathan*, she argues that the establishment of the state serves this particular function in the emerging proto-capitalist order (von Redecker, 2021, p. 899). However, as AI futurists notoriously reject the state, it can be argued that technology assumes this role in its stead.

9 Digital colonialism emphasises Western powers' maintenance of dominance through digital technologies, such as communication networks.

to 'empty dead proto-property' (p. 64). With futures being framed as final frontiers, it is crucial to examine which subjects are envisioned to be on either side of the frontier. This entails understanding how certain bodies, characterised as 'white' and 'male', are commonly associated with attributes like reason, creativity, and 'future-making capacity', while others are depicted as the antithesis of these qualities (Tarvainen, 2022, p. 9). For example, Nick Bostrom's idealised future of 'post-humanity' (2002) aligns with mainstream norms of white masculinity: 'deeply disembodied, unattached to place, and dominant over, or independent from, "nature"' (Mitchell & Chaudhury, 2020, p. 316). Additionally, it is important to note that the underlying colonising logic stretches beyond the symbolic and discursive and rests upon an ongoing material and violent practice of **dispossession and exploitation** (Mezzadra & Neilson, 2017; Tarvainen, 2022, p. 3). This includes the continuous production of a precarious labour force, as discussed in the previous section, to uphold AI's façade of excellence. The future, as it is being propagated by the AI industry, thus reveals itself as the relentless **extension of an exclusionary and violent past and present**.

6. Building counterimaginaries

A key criticism in this paper focuses on the exclusivity of individuals or small groups shaping futures rather than embracing co-design through collective, pluralistic processes. Accordingly, this paper does not aim to propose an improved alternative vision for the future. However, in exploring alternative technological visions of the future and their contrast with elitist perspectives, **BIPOC futurism and decolonial approaches** emerge as noteworthy examples. It is necessary to note that the limited scope of this paper does not fully capture the depth of their visions, claims, and complexities.

BIPOC and decolonial futures explicitly diverge from elitist futures by prioritising human and land rights while vehemently opposing extractivism and historically grown domination (Ricaurte, 2019). With the ontological separation and hierarchisation of culture and nature being fundamentally called into question, the AI futurist aspiration to **wield technology for dominion** over nature becomes crucially challenged. Conversely, decolonial movements endorse technologies that cater to community needs and adapt to **local contexts** (Lehuedé, 2024) in contrast to pre-set solutions aimed at universal scalability, a position that is also echoed by degrowth movements (Vetter, 2023).

The motif of **technological sovereignty** is also central in defying AI elitist futures, envisioning a world order without single hegemonic powers dictating technology, infrastructure, and data governance (Lehuedé, 2024, p. 10). Rather than endorsing singular and homogenous scenarios, BIPOC and decolonial approaches embrace the possibility of multiple futures. Founded on notions such as co-dependence, conviviality (Illich, 1973), and pluriversality (Escobar, 2018), they emphasise the importance of 'diverse, plural subjectivities and forms of agency' rather than homogenised views of humanity (Mitchell & Chaudhury, 2020). Movements like the Zapatistas in Mexico aim toward 'a world in which many worlds coexist' (Mignolo, 2011, p. 273), ultimately embodying a struggle against hegemony and extractivism, and for the 'flourishing of political, epistemological, and ontological systems' (Lehuedé, 2024, p. 3). The ***AI Decolonial Manyfesto*** (Krishnan et al., 2021), signed by various academics and activists, espouses a relationship with AI that acknowledges epistemic expertise stemming from lived experience and compels historically marginalised communities to 'build their own dignified socio-technical futures'. While 'creation, art, stories, and sensitive experience' can open pathways for such decolonial imaginations, it is also essential to address the necessity of reparations for historical and ongoing dispossessions (Krishnan et al., 2021).

7. Conclusion

> *'[A]s neoliberalism took hold, writing new futures got harder and harder for all but the world's wealthiest'.*
> — Kate Aronoff, *Overheated*

In conclusion, this paper has explored how the AI industry disseminates potent visions of the future to secure and enforce its advantageous position. As demonstrated, their propagated **visions are crafted with self-serving agendas**, employing discursive tactics of determinism and distraction to deter criticism and inhibit the exploration of alternatives. These visions are gaining institutional traction and being adopted by policymakers. As a result, they are solidifying into collective socio-technical imaginaries that increasingly influence the path we are steering toward.

Dissecting the mechanisms of future fabrication employed by the AI industry thus emerges as a **primary counterstrategy** essential for paving the way toward inclusive, equitable, and sustainable futures. It is crucial to remain alert about how AI actors perpetuate powerful myths that narrow collective imaginations of the future to align with their interests. Critical analysis of AI discourse is indispensable, particularly when discussing options for addressing the climate crisis and advancing social justice. For AI practitioners, this entails questioning whether proposed solutions constitute mere technological fixes or require paradigm shifts (Klein, 2023) and critically interrogating the underlying idea of 'progress' inherent in AI-driven visions of the future.

Moving forward, the strategies employed by the AI industry must be investigated further, as they are **likely to adapt to the shifting discourse**. Additionally, there is potential to explore how prominent AI futures affect the imaginative capacities of different groups, as well as their effects on subjects, such as the experience of powerlessness or defeat. Efforts should be directed towards articulating and strengthening visions of the future that are not hegemonic but represent a **polyphonic and ever-evolving collective imagination** (Jovchelovitch & Hawlina, 2018). In light of the intensifying climate crisis and rising social inequality, it is imperative to (re-)imagine alternative lifeways beyond capitalist, fossil fuel-dependent, and growth-oriented paradigms (Hosseini & Gills, 2020, p. 1), fostering pluralistic approaches that emphasise collective and participatory decision-making, such as those offered by BIPOC futurism and decolonial approaches. Amplifying and furthering such critical visions remain crucial strategies in challenging and confronting elitist and other exclusionary fabricated futures.

Acknowledgements

Many thanks to Stefano Rosso, Giulia Hambsch, Jan Siebold, and Paul Schütze, as well as the anonymous reviewers, for their generous, thoughtful, and constructive feedback that helped the text grow into the current version.

References

Agar, N. (1998). Liberal Eugenics. *Public Affairs Quarterly*, *12*(2), 137–155.

Aronoff, K. (2021). *Overheated: How capitalism broke the planet--and how we fight back* (First edition). Bold Type Books.

Bazzani, G. (2023). Futures in Action: Expectations, Imaginaries and Narratives of the Future. *Sociology*, *57*(2), 382–397. https://doi.org/10.1177/00380385221138010

Ben-Shahar, O. (2019). Data Pollution. *Journal of Legal Analysis*, *11*, 104–159. https://doi.org/10.1093/jla/laz005

Birnbaum, E. (2023, February 1). *Tech Giants Broke Their Spending Records on Lobbying Last Year—BNN Bloomberg*. BNN Bloomberg. https://www.bnnbloomberg.ca/tech-giants-broke-their-spending-records-on-lobbying-last-year-1.1877988

Bordelon, B. (2023, December 30). *When Silicon Valley's AI warriors came to Washington*. POLITICO. https://www.politico.com/news/2023/12/30/ai-debate-culture-clash-dc-silicon-valley-00133323

Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, *9*. https://ora.ox.ac.uk/objects/uuid:827452c3-fcba-41b8-86b0-407293e6617c

Bostrom, N. (2017). *Superintelligence: Paths, dangers, strategies* (Reprinted with corrections 2017). Oxford University Press.

Braun, K., & Kropp, C. (2023). Building a better world? Competing promises, visions, and imaginaries-in-the-making of the digitalization of architecture and construction. *Futures*, *154*, 103262. https://doi.org/10.1016/j.futures.2023.103262

Brevini, B. (2021). Creating the Technological Saviour: Discourses on AI in Europe and the Legitimation of Super Capitalism. In P. Verdegem (Ed.), *AI for Everyone? Critical Perspectives* (pp. 145–159). University of Westminster Press. https://doi.org/10.16997/book55.i

Brevini, B. (2022). *Is AI good for the planet?* Polity Press.

Brockmann, H., Drews, W., & Torpey, J. (2021). A class for itself? On the worldviews of the new tech elite. *PLOS ONE*, *16*(1), e0244071. https://doi.org/10.1371/journal.pone.0244071

Clarke, L. (2023, September 14). *How Silicon Valley doomers are shaping Rishi Sunak's AI plans*. POLITICO. https://www.politico.eu/article/rishi-sunak-artificial-intelligence-pivot-safety-summit-united-kingdom-silicon-valley-effective-altruism/

Coleman, D. (2019). Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws. *Michigan Journal of Race & Law*, *24.2*, 417. https://doi.org/10.36643/mjrl.24.2.digital

Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, *20*(4), 336–349. https://doi.org/10.1177/1527476418796632

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Daub, A. (2021). *Was das Valley denken nennt: Über die Ideologie der Techbranche* (S. Gebauer, Trans.; 3. Auflage). Suhrkamp.

Escobar, A. (2018). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Duke University Press.

Friend, T. (2016, October 3). Sam Altman's Manifest Destiny. *The New Yorker*. https://www.newyorker.com/magazine/2016/10/10/sam-altmans-manifest-destiny

Fulterer, R. (2023, November 25). Ex-OpenAI-Manager Zack Kass im Interview: KI, ChatGPT und das ewige Leben. *Neue Zürcher Zeitung*. https://www.nzz.ch/technologie/kuenstliche-intelligenz-und-open-ai-zack-kass-zu-chatgpt-klimawandel-ki-tech-ld.1765858

Gebru, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*. https://doi.org/10.5210/fm.v29i4.13636

Goertzel, B. (2010). *A cosmist manifesto: Practical philosophy for the posthuman age.* Humanity+.

Grasso, M. (2019). Oily politics: A critical assessment of the oil and gas industry's contribution to climate change. *Energy Research & Social Science*, *50*, 106–115. https://doi.org/10.1016/j.erss.2018.11.017

Gray, M. L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.

Greenpeace. (2020). *Oil in the Cloud* [Report]. Greenpeace USA. https://www.greenpeace.org/usa/reports/oil-in-the-cloud/, https://www.greenpeace.org/usa/reports/oil-in-the-cloud/

Hanson, R. (2008). Catastrophe, social collapse, and human extinction. In R. Hanson, *Global Catastrophic Risks*. Oxford University Press. https://doi.org/10.1093/oso/9780198570509.003.0023

Hartley, D. (2016). Anthropocene, Capitalocene, and the Problem of Culture. In J. W. Moore (Ed.), *Anthropocene or Capitalocene? Nature, History, and the Crisis of Capitalism* (pp. 154–165). Kairos.

Hawlina, H., Pedersen, O. C., & Zittoun, T. (2020). Imagination and social movements. *Current Opinion in Psychology*, *35*, 31–35. https://doi.org/10.1016/j.copsyc.2020.02.009

Heffernan, T. (2020). Fiction Writes Back: "Limitless Profit", Artificial Intelligence, and the Immortality Industry. *Researcher. European Journal of Humanities & Social Sciences*, *3*(1), 27–46. https://doi.org/10.32777/r.2020.3.1.2

Hosseini, S. A. H., & Gills, B. K. (2020). Beyond the critical: Reinventing the radical imagination in transformative development and global(ization) studies. *Globalizations*, *17*(8), 1350–1366. https://doi.org/10.1080/14747731.2020.1736852

Illich, I. (1973). *Tools for conviviality* (1. Perennial Library paperback ed). Perennial Library.

Jasanoff, S. (2015). Future Imperfect: Science, Technology, and the Imaginations of Modernity. In S. Jasanoff & S.-H. Kim (Eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226276663.001.0001

Jovchelovitch, S., & Hawlina, H. (2018). Utopias and World-Making: Time, Transformation and the Collective Imagination. In C. de Saint-Laurent, S. Obradović, & K. R. Carriere (Eds.), *Imagining Collective Futures: Perspectives from Social, Cultural and Political Psychology* (pp. 129–151). Springer International Publishing. https://doi.org/10.1007/978-3-319-76051-3_7

Klein, N. (2023, May 8). AI machines aren't 'hallucinating'. But their makers are. *The Guardian*. https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein

Krishnan, A., Abdilla, A., Moon, A. J., Souza, C. A., Adamson, C., Lach, E. M., Ghazal, F., Fjeld, J., Taylor, J., Havens, J. C., Jayaram, M., Morrow, M., Rizk, N., Ricaurte Quijano, P., Çetin, R. B., Chatila, R., Dotan, R., Mhlambi, S., Jordan, S., & Rosenstock, S. (2021). *AI Decolonial Manyfesto*. https://manyfesto.ai/index.html

Kwet, M. (2018). *Digital Colonialism: US Empire and the New Imperialism in the Global South* (SSRN Scholarly Paper 3232297). https://doi.org/10.2139/ssrn.3232297

Lehuedé, S. (2024). An alternative planetary future? Digital sovereignty frameworks and the decolonial option. *Big Data & Society*, *11*(1), 20539517231221778. https://doi.org/10.1177/20539517231221778

Li, M., Trencher, G., & Asuka, J. (2022). The clean energy claims of BP, Chevron, ExxonMobil and Shell: A mismatch between discourse, actions and investments. *PLOS ONE*, *17*(2), e0263596. https://doi.org/10.1371/journal.pone.0263596

Lösch, A., Heil, R., & Schneider, C. (Eds.). (2019). Visionary Practices Shaping Power Constellations. In *Socio-Technical Futures Shaping the Present: Empirical Examples and Analytical Challenges* (pp. 67–88). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27155-8

Marx, P. (2024, January 24). *Sam Altman's self-serving vision of the future*. Disconnect. https://disconnect.blog/sam-altmans-self-serving-vision-of-the-future/

Mezzadra, S., & Neilson, B. (2017). On the multiple frontiers of extraction: Excavating contemporary capitalism. *Cultural Studies*, *31*(2–3), 185–204. https://doi.org/10.1080/09502386.2017.1303425

Miceli, M., & Posada, J. (2022). The Data-Production Dispositif. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–37. https://doi.org/10.1145/3555561

Mignolo, W. D. (2011). *The Darker Side of Western Modernity: Global Futures, Decolonial Options*. Duke University Press.

Mitchell, A., & Chaudhury, A. (2020). Worlding beyond 'the' 'end' of 'the world': White apocalyptic visions and BIPOC futurisms. *International Relations*, *34*(3), 309–332. https://doi.org/10.1177/0047117820948936

Musk, E. [@elonmusk]. (2022, August 26). *Population collapse due to low birth rates is a much bigger risk to civilization than global warming* [Tweet]. Twitter. https://twitter.com/elonmusk/status/1563020169160851456

O'Connell, M. (2018, February 15). Why Silicon Valley billionaires are prepping for the apocalypse in New Zealand. *The Guardian*. https://www.theguardian.com/news/2018/feb/15/why-silicon-valley-billionaires-are-prepping-for-the-apocalypse-in-new-zealand

Odell, J. (2023). *Saving Time: Discovering a Life Beyond the Clock*. Jonathan Cape.

Osnos, E. (2017, January 22). Survival of the Richest. *The New Yorker*. https://www.newyorker.com/magazine/2017/01/30/doomsday-prep-for-the-super-rich

Redecker, E. von. (2021). Autoritäre Akkumulation: Hannah Arendt über Hobbes' Leviathan und die bürgerliche Geschichte. *Deutsche Zeitschrift Für Philosophie*, *69*(6), 897–914. https://doi.org/10.1515/dzph-2021-0074

Redecker, E. von. (2023). *Bleibefreiheit* (2. Aufl). SFischer.

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, *20*(4), 350–365. https://doi.org/10.1177/1527476419831640

Rushkoff, D. (2022). *Survival of the richest: Escape fantasies of the tech billionaires* (First edition). W.W. Norton & Company.

Schulz, M. S. (2016). Social Movements and Futures Research. *World Futures Review*, *8*(2), 98–107. https://doi.org/10.1177/1946756716634310

Schütze, P. (2024). The Impacts of AI Futurism – An unfiltered look at AI's true effects on the climate crisis. *Ethics and Information Technology*.

Sephiroth. (2016, July 24). *Technological singularity—How many of you are aware of it and have heard of Ray Kurzweil's accurate predictions?* Steemit. https://steemit.com/futurism/@sephiroth/technological-singularity-how-many-of-you-are-aware-of-it-and-have-heard-of-ray-kurzweil-s-accurate-predictions

Siegel, E. (2015). *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Wiley.

Supran, G., & Oreskes, N. (2021). Rhetoric and frame analysis of ExxonMobil's climate change communications. *One Earth*, *4*(5), 696–719. https://doi.org/10.1016/j.oneear.2021.04.014

Szczepański, M. (2020). *Is data the new oil?* (PE 646.117; EPRS | European Parliamentary Research Service).

Taffel, S. (2023). Data and oil: Metaphor, materiality and metabolic rifts. *New Media & Society*, *25*(5), 980–998. https://doi.org/10.1177/14614448211017887

Tarvainen, A. (2022). The modern/colonial hell of innovation economy: Future as a return to colonial mythologies. *Globalizations*, 1–23. https://doi.org/10.1080/14747731.2022.2048460

Teuscher, C. (2004). *Alan Turing: Life and legacy of a great thinker*. Springer.

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

Torres, É. P. (2023). TESCREAL: The Acronym Behind Our Wildest AI Dreams and Nightmares. *Policycommons.Net*. https://policycommons.net/artifacts/11329870/untitled/12218687/

Van Doorn, N. (2017). Platform labor: On the gendered and racialized exploitation of low-income service work in the 'on-demand' economy. *Information, Communication & Society*, *20*(6), 898–914. https://doi.org/10.1080/1369118X.2017.1294194

Verdegem, P. (2022). Dismantling AI capitalism: The commons as an alternative to the power concentration of Big Tech. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01437-8

Vetter, A. (2023). *Konviviale Technik: Empirische Technikethik für eine Postwachstumsgesellschaft*. transcript Verlag. https://doi.org/10.1515/9783839453544

Vinsel, L. (2021, February 1). You're Doing It Wrong: Notes on Criticism and Technology Hype. *Medium*. https://sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5

von Maur, I., & Slaby, J. (2024). *Gefühle einfordern? Zur Normativität gesellschaftlicher Affektlagen. 01–2024*.

Vranken, B. (2023, November 24). *Big Tech lobbying is derailing the AI Act*. Corporate Europe Observatory. https://corporateeurope.org/en/2023/11/big-tech-lobbying-derailing-ai-act

Wang, D., Prabhat, S., & Sambasivan, N. (2022). Whose AI Dream? In search of the aspiration in data annotation. *CHI Conference on Human Factors in Computing Systems*, 1–16. https://doi.org/10.1145/3491102.3502121

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

World Economic Forum. (2011). *Personal Data: The Emergence of a New Asset Class*. https://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf

Yudkowsky, E., & Bostrom, N. (2018). The Ethics of Artificial Intelligence. In *Artificial Intelligence Safety and Security*. Chapman and Hall/CRC.

Debiasing with foresight: A horizon scanning process on the digitalisation of research and policymaking

Lorenz Erdmann and Simone Kimpeler

Abstract: Research and policymaking organisations increasingly recognize the importance of digitalisation for both, their own operations and the transformations of their operating environment. The opaqueness and openness of digital developments ask for dealing with future uncertainty. However, biases distort the perception of changes and constrain the exploration of plausible and relevant alternative futures. This chapter explains and reflects how thinking about digital futures can be debiased with foresight. We present and discuss our conceptual design and empirical material building upon the three-year foresight project 'Environmental research and governance in the digital age' that we led as a contractor of the Federal Environment Agency. We co-designed a five-step horizon scanning process and integrated debiasing measures for each step thereby systematically mitigating six salient biases in futures thinking, among them the surveillance filter and the confirming trend bias. As a result, four success factors of debiasing with foresight stand out: eye-level foresight and digitalisation competences of both the contracting authority and contractor; co-creation of the scan field definition thereby reflecting different impact pathways of digitalisation; multiple flexibly adopted horizon scanning approaches encompassing semiautomated scanning, human main and fringe source analysis, interviews with radical thinkers and targeted meta-engine search; participatory sensemaking to leverage diverse and controversial external views in tailoring the future topics and anticipating different digital future trajectories and emerging issues. The suggested debiasing measures are systematically and explicitly integrated into the horizon scanning to ensure the novelty, relevance and validity of dealing with uncertain digital futures. The process yielded a disaggregated and multi-perspective view of digitalisation with its future topics, subtopics, and related emerging issues. Such a strategic compass for research and policymaking can stimulate organisational learning through a common reference. It should be kept modifiable and amendable to account for new insights into the dynamically changing panoply of digital phenomena.

Keywords: anticipation biases, integrated debiasing, hybrid human-automated search, participatory sensemaking, strategic compass

1. Introduction

Digitalisation is changing the way how organisations design, implement, and evaluate their activities. Pertinent definitions of **digitalisation** refer to improving, changing, and enabling processes through digital technology (e.g. Bitkom, 2016) or to the deep transformation that emerges from the interplay of social and technological digitalisation processes (e.g. WBGU, 2019). All kinds of organisations, policymaking government bodies in particular, recognise the importance of digitalisation for their own operations and are increasingly aware of its transformative effects on their environment and of the opportunities and risks entailed. The future impacts of digitalisation on government activities are uncertain, as the future is uncertain to a significant degree (e.g. Loveridge, 2008).

Foresight is the 'discipline of exploring, anticipating and shaping the future [...] in a structured and systemic way' (EC, 2024). It is not about predicting the future but about dealing with future uncertainty and exploring different possible futures and strategic implications. A variety of foresight methods and approaches are available (Popper, 2008; Saritas, 2013), ranging from exploratory to normative methods, routing developments in the present (e.g. Horizon Scanning, Three Horizon Framework) or 'jumping' into the future (e.g. pictures of the future, shared vision). An understanding of foresight as a cycle of identifying current changes, imagining future changes, and promoting changes (action) is increasingly established (Cuhls, 2019).

Human perceptions of the future are biased. **Biases** can be defined as systematic distortions of perception that influence decision-making and judgement under uncertainty (Gigerenzer & Gaissmaier, 2011; Kahneman et al., 1982). The scholarship of psychology and behavioural economics has identified and examined a large range of unconscious biases in decision-making, some of which have been studied in relation to future change. For example, Ansoff (1975) identifies three perception filters that could lead to biases in the detection and assessment of weak signals of change: the surveillance filter structures an organisation's resource-constrained observation of its environment, the mentality filter shapes anyone's thoughts and judgements according to previous experiences, and the power filter distorts the perception and judgement of the importance of weak signals through power relations and organisational and hierarchical routines.

Recognising perception filters that shape biases implies the need to widen those that structure our thinking about the future (Ansoff, 1975; Holopainen & Toivonen, 2012; Miller et al., 2012; Rossel, 2012; Warnke & Schirrmeister, 2016). In scenario building, biases analysed include 'group biases' leading to the conformity of participants and 'the end of history illusion' stating that people underestimate future change (Schirrmeister et al., 2020). The tendency to confirm one's own assumptions (confirming trends bias), the preference for positive trends (overconfidence bias), and the overestimation of predictability (overprediction bias) all contribute to biases in future-oriented thinking.

Dedicated foresight processes can be measures for debiasing a future-centred approach (Schirrmeister et al., 2020; Van Woensel, 2020). In our view, **debiasing** is a conceptual approach to make biases explicit and to provide measures to mitigate them and their impacts. For example, scenario workshop moderators can provide measures to stimulate thinking of alternatives and to create a communication climate to mitigate 'group biases' and 'the end of history illusion'.

Although there is a large body of research on the social, economic, and environmental future impacts of digitalisation (e.g. Muench et al., 2022; WBGU, 2019), there is little research on how digitalisation is changing the way knowledge is produced to inform policymaking (EC, 2023; Van Woensel, 2020). There is a substantial need for insights into how interactions between digital and other system transformations, the so-called 'wicked problem', could be addressed through strategic foresight approaches (Cork et al., 2023) and how typical biases in the perception and evaluation of signals of change in an organisation can be addressed (Schirrmeister et al., 2020).

In our chapter, we **aim** to contribute to making future thinking about digitalisation and related deep transformations less biased. By outlining the importance of debiasing with foresight when exploring digital futures, a conceptual contribution to the advancement of the field is made. We draw upon a **horizon scanning** study on environmental research and governance in the Digital Age developed for a public environmental agency that wished to analyse the consequences of digitalisation beyond its usual strategic considerations (see Acknowledgements). Horizon scanning comprises the systematic search for signals of change (e.g. events, trends, issues) and the assessment of their relevance, e.g. for policy implications (Cuhls, 2019;

Hines et al., 2019). Throughout the design, conduct, and evaluation of the results of this horizon scanning study, biases were identified, and debiasing measures were provided and applied. The scope of this endeavour is large, and there are valid studies on single issues of debiasing with and within foresight. However, such a comprehensive scope in the light of digitalisation, as presented here, is exactly what government bodies might increasingly ask from foresight practitioners. We conclude with an assessment of the novelty and added value of debiasing with horizon scanning and its transferability to other policy areas.

2. The approach to debiasing with horizon scanning

Two *research questions* guided the horizon scanning:

- What are the emerging digital developments in environmental research and governance?
- How is digitalisation transforming the policy organisation's context, and what impact could this have on future environmental research and governance?

To answer these questions, we developed and applied debiasing approaches for all steps of horizon scanning in order to minimise potential biases in exploring the two research questions.

a) Steps of the horizon scanning process on digital futures

In this subsection, we briefly introduce the principal steps of horizon scanning to the extent necessary to make the debiasing approach accessible to readers.

The horizon scanning process consisted of five major steps:

1. Scoping: The client's information requirements and the scan field were defined. The scan team was composed, and the scanning methods, channels, and sources were selected.
2. Scanning: The different scanning approaches were set up and conducted, and they delivered candidates for signals of change that were stored in a digital repository. The scan team routinely discussed the relevance of signals, the need for reframing signals, and the coverage of the scan field to redirect scanning activities.

3. Participatory sensemaking: This step involved moderated discussions with the contracting organisation and two co-creative, participatory workshops. The first was a workshop which clustered signals of change into future topics that have a certain weight in policymaking. The second was a workshop which involved experts and stakeholders in exploring emerging issue candidates.
4. Desk research: Future topics and emerging issues were analysed and validated through additional targeted desk research. Preliminary findings were sharpened and deepened, and the need for action was elaborated.
5. Foresight transfer: The results were condensed in the scan report (Erdmann et al., 2024). Workshops with the contracting authorities developed ideas for addressing the various signals of change and emerging issues in policymaking and everyday work. The findings of the horizon scanning are about to serve as a strategic compass for the digital future in the organisation's activities.

Throughout the process, possible biases were identified and addressed by specific measures.

b) Debiasing with horizon scanning

In preparation for an emerging digital society, biases could lead to the ignorance of relevant signals of change and their disregard for strategic decisions. Strategic foresight supports the handling of future uncertainty and, therefore, offers the potential for reducing biases in future thinking at the individual and organisational level (Van Woensel, 2020). Biases are also at play in foresight itself (Nestik, 2018), so it is not enough to apply foresight, but it is also important to select and combine appropriate methods for debiasing in foresight processes. In this particular horizon scanning, an innovative set of scanning methods was applied, and in this chapter, we assess its potential for debiasing.

Based on the literature analysis (see section 1), three perception filters shaping biases and three judgement biases relevant to foresight have been selected and addressed by debiasing measures:

- mentality filter
- power filter
- surveillance filter

- confirming trend bias
- overconfidence bias
- overprediction bias

Table 1 gives an overview of how different perception filters and biases in future-centred thinking were systematically addressed by debiasing measures in the horizon scanning process on digital futures. The implementation of the debiasing methods is then described following the five steps of the scanning process.

Table 1. Debiasing with horizon scanning – an overview

	Scoping	Scanning	Sensemaking	Desk research	Foresight transfer
Mentality filter	Co-design challenging the contracting authority's daily routines Diversity in the scanning team in view of disciplines and sociocultural disposition	Deployment of various complementary approaches, channels, and sources balancing specific weaknesses out Diversity in the scanning team regarding disciplines and sociocultural disposition	Diversity of workshop participants (experts and stakeholders) Co-creative workshop concept, moderated group discussions, stimulate dissent	Challenges mentalities through fringe source analysis and interviews with out-of-the-box thinkers	Clarify how to address clients in their daily work (keep them biased) or in separate formats (debiasing possible)
Power filter	Identify potential blind spots of power incumbents through intense questioning Diversity in the scanning team in view of hierarchies and eye-level communication with contracting authority	Detachment of the operational scanning from unwanted interference Value the contributions of juniors, newcomers, and other contributors with poor power	Diversity of workshop participants (strategic and technical level, non-academics and practitioners) Moderation to attract voices from all parties involved	Questioning the authority of usual suspects in the assessment of issues (consider lead users, revisiting blogs, and other fringe sources)	Involvement of different operational levels, strategic and specialist departments when discussing the potential for addressing emerging issues
Surveillance filter	360° scan of all possible domains Precise discursive delimitation of the scan field in cooperation with the contracting authority	Account for vast source semi-automated scanning 'scan the scanners' Targeted desk research to validate the initial findings of weak signals	Clustering of individual signals to identify emerging patterns and trends not visible from individual signals alone	Interdisciplinary experts desk research for validation of emerging issues identified in the workshops	Identify connections to ongoing work in the scan report and substantiate them with sources
Confirming trend bias	Consideration of all-encompassing and fringe	Combination of source types	Broad involvement of experts, foresight practi-	Challenge findings through external voices	Communication and transfer keep the future uncertain

	Scoping	Scanning	Sensemaking	Desk research	Foresight transfer
	sources detecting weak signals	Doubt is key within the scan team culture	tioners, involved and affected stakeholders	(e.g. interviews with visionaries)	Learn through reception by different audiences
Overconfidence bias	Variety of sources (different disciplines, media) Consideration of 'undesirable' and marginal sources	Search for signals of negative developments Doubt is key within the scan team culture	Application of the Three Horizon Model Workshop design leveraging pessimistic views	Reflection on underlying values	Propagate negative findings in the final results
Overprediction bias	Consideration of inconvenient and fringe sources challenging 'future certainty'	Search for signals of countertrends Doubt is key within the scan team culture	Application of the Three Horizon Model Workshop design leveraging alternative views	Search for signals of countertrends	Explicit link of emerging issues to current policy agenda and future mission of the organisation

Note: Compiled by the author

aa) Scoping

The aim of the horizon scanning was to identify signals of change driven by digitalisation that could have an impact on the field of action of the contracting organisation (Figure 1). First, digital technology itself is changing (e.g. maturity of quantum technology); second, digital technology is directly impacting the core issue under investigation (e.g. an organisation's processes are being digitalised); and third, the environment of the core issue under investigation – which is part of the wider society – is being digitalised (e.g. digital working practices), thus having an indirect impact on the core issue under investigation.

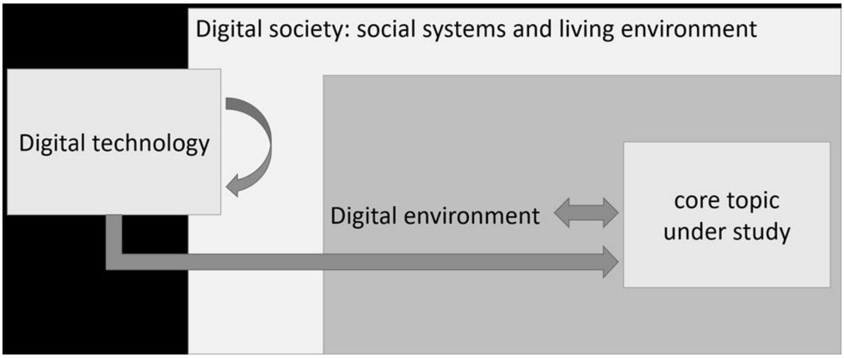

Figure 1. Generic scheme of digital change impact mechanisms (Erdmann et al., 2024)

The definition of the scan field included all three mechanisms of digital change. It covered 20 areas, ranging from the research cycle and the policy cycle (initiation, conduct, and evaluation/implementation) over digitalisation itself to social systems, such as education, and to the living environment, such as private life. The criteria for the search, selection, and assessment of signals of change were a) the relevance of the information for one of the three change mechanisms (Figure 1), b) the novelty of the information for current strategies and decision-making, and c) the coverage of the scan field.

The aim was to ensure the greatest possible diversity of expertise in the team and in the types of horizon scanning approaches, channels, and sources. To widen the filters of perception within the organisation, an interdisciplinary scan team was built, consisting of foresight experts with additional expertise in media studies, sociology, economic policy, environmental science and technology, history, data science, semiotics, business studies, psychology, and policy research. Measures to mitigate biases in the set-up of the horizon scanning included the mutual challenging of mentality and power filters of both the contracting authority and the contractor.

bb) Scanning

The scanning process tapped five major source types: main sources, fringe sources (i.e. at the fringes of perception of the contracting authority), a vast digital news repository, interviews with unconventional thinkers, and a gap-closing targeted search with a meta-search engine. The objective of this source type set was to minimise the impacts of unconscious biases in main sources and to strengthen the validity of individual signals with multiple occurrences.

The semi-automated scanning exploited the vast digital news repository using an ontology of key terms, topic modelling, and the identification of highly relevant individual articles. It aimed at being as comprehensive as possible in detecting signals of change. The tapping of digital news was based on web scraping via the provider API-News.[1] Two separate corpora, one on general news (e.g. BBC, Hindustan Times) and one on science news (e.g. PNAS, *Nature*), were built. Search strings were defined by combing terms from the realms 'digital', 'research', 'governance' and 'environment' to scrape relevant content. The two scraped news repositories were analysed separately with and without accounting for the terms for 'environment'. For both, a natural language processing (NLP) approach was carried out to identify hidden topics from the two corpora of digital texts, which could not have been identified in the short time available using a qualitative evaluation. NLP has expanded the analytical capabilities of web scraping substantially (Goddard et al., 2021; Nemorin et al., 2023; Vignoli et al., 2022). The ontology, programming, interpretation of results, and iterative refinement are the human components in this hybrid human–automated search (Geurts et al., 2021; Krigsholm & Riekkinen, 2019). The findings from the automated search were evaluated by several experts regarding their novelty for the client and relevance to the action areas. Experts identified and conducted a detailed analysis of sources that contributed substantially to a topic.

1 https://newsapi.org/

In addition, digitally available primary and secondary sources that met the criteria of high quality, timeliness, and originality were evaluated (see Table 2). This expert-based activity covered main source exploitation (e.g. archives and specialist journal articles) as well as the search and analysis of so-called marginal or fringe sources. Fringe sources are sources usually not considered in the context of strategy development, e.g. podcasts, science blogs, special interest journals, or community platforms. However, they can be important sources for early signals of change as they look at developments from a less established perspective. Interviews with unconventional thinkers contributed to debiasing the confirming trend bias, too. Table 2 shows the source types selected, as well as their coverage and processing, to extract insights.

Table 2. Coverage of the dimensions of digital change and processing by main source types (own compilation)

Source type	Coverage (What?)	Processing (How?)
Digital science news repositories	Comprehensive coverage of digital technology and its applications	Web scraping, topic modelling, and analysis of most contributing sources
Digital general news repositories	Comprehensive coverage of digital society, more precisely, the digital environment of the core topic under study	Web scraping, topic modelling, and analysis of most contributing sources
Conferences	Deliberation of emerging digitalisation issues	Search and analysis of programmes and abstracts to identify premature emerging issues
Pre-print archives	Perception of emerging epistemic issues due to digital change	Search and analysis of principal articles at the fringes of scientific consolidation
Peer-reviewed journals	Perception of digital change in the core topic under study	Search and analysis of salient articles on the core topic under study
Foresight studies	Future assumptions on digital technology, digital society, and the core topic under study	Search and analysis of signals of change such as weak signals, wild cards, key factors, trend impacts analyses, and visions
Popular media	Insights into the living environment	Search and analysis of digital living practices
Social media	Deliberation of publicly resonating changes in the living environment	Search and analysis of speculative digital change

Source type	Coverage (What?)	Processing (How?)
Interviews with visionaries	Questioning future assumptions on digital technology, digital society, and the core topic under study	Exploratory interviews with a small set of open questions
Targeted Internet search	Coverage of subtopics on digital technology, digital society, and the core topic under study	Meta-search engine switching on and off certain dominant machines

The horizon scanning was carried out in iterations, identifying blind spots and new keywords that guided subsequent scanning activities, followed by desk research to validate initial findings. The scan team regularly reviewed results and modified the signal candidates. The diversity of sources and clarity of the scan field allowed for reflection and the recognition of the maturing of the search results. A long list of signals of digital change was produced, and each (e.g. cybernetic citizenship) was described in a short summary with various references providing evidence for the emergence of the development and noting why it should be considered novel and relevant for the future.

cc) Participatory sensemaking

The horizon scanning process identified roughly 70 pertinent weak signals of change. Sensemaking activities complement the scanning by participatory, multiple-actor, and stakeholder engagement in the assessment of signals of change and future assumptions (Rosa et al., 2021). The sensemaking aimed to formulate a smaller number of relevant future topics from the panoply of weak signals and to anticipate emerging issues. The aim of the debiasing was to minimise individual and group-related biases by leveraging and combining different perspectives and points of view.

Two co-creative and participative sensemaking workshops were designed, conducted, and evaluated:

- The cluster workshop aggregated and condensed signals of change to broader future topic candidates in an open process. These future topic candidates were later consolidated by additional desk research and internal discussion.
- The future workshop explored different developments of these future topics and identified related emerging issue candidates based on the Three Horizon Framework, which explores the strategic fit of current institutions (Horizon 1), emerging signals (Horizon 3), and innovating

institutions that account for emerging signals (Horizon 2) to the world as it is (Curry & Hodgson, 2008).

Both workshops ensured the relevance and directionality of the foresight in view of the organisation's activity field.

Debiasing efforts were supported through (a) the involvement of experts for the different topics addressed in the 70 signals both in interviews and/or participation in one or both workshops, (b) the engagement of topic experts and policymakers from the contracting organisation in both workshops, and (c) the choice of the workshop methods and setting (e.g. breakout groups, whiteboards to support visual thinking, moderated discussions). In total, 45 external experts were involved in the sensemaking workshops, alongside six experts from the policymaking organisation and five experts from the research team. The workshop participants' composition supported the reflection of the internal stakeholder perspective in light of the external perspectives from the policy field and from related research.

dd) Desk research for validation of the findings

The sensemaking results were transferred into a scan report. In addition, a cross-future topic analysis yielded a landscape of shared and specific emerging issues. The repercussions and implications of these emerging issues on the remit of the contracting organisation were represented in the Three Horizon Framework, which enables the visual integration of different futures routed in the present (Curry & Hodgson, 2008). Following an understanding of signals in horizon scanning as 'observations that trigger irritation and as an indication of possible changes' (Rossel, 2012, p. 236), the analysis of the future topics and emerging issues had to be related to other observations in public and other relevant discourses.

Debiasing was integrated into the research on the selected ten future topics and eight crosscutting emerging issues: First, by juxtaposing the assumptions from the sensemaking workshops with established discourses and positions on the topic; second, by rigorously applying the Three Horizons Framework (Curry & Hodgson, 2008) to promote consideration of emerging issues beyond business-as-usual future expectations; and third, by conducting targeted desk research including conforming and counterfactual search strategies.

ee) Foresight transfer

The integration of foresight into political decision-making (Da Costa et al., 2008) necessitates a comprehensive search for indications of change and the optimal involvement of actors and interest groups (De Vito & Radaelli, 2023). Possible obstacles to the development and adoption of future-oriented knowledge in any policymaking government body include a less open organisational culture as well as mental models and prejudices that lead to a reluctance to change or inhibit the development of new capacities (Mortensen et al., 2021).

The recognition of digitalisation for transformative change requires various systemic and dynamic capabilities (Borrás & Edler, 2020; Wu et al., 2015), including the ability to develop and test new policy approaches and anticipatory capacities. The establishment of such capabilities often faces a hierarchical and bureaucratic organisational culture. Such organisations' employees are typically preoccupied with well-defined and – to a certain extent – standardised operational tasks. Embedding foresight as a strategic activity naturally conflicts with these operational activities. Hierarchical organisations face challenges when collaboration in foresight is required across different departments. Foresight can either be set aside from the daily routines as an extra activity, or it reflexively feeds into and supports the daily work of employees as their roles evolve in the Digital Age.

The horizon scanning project material was reviewed, four open interviews for transfer ideation took place, and four virtual meetings were held with the contracting organisation to explore options for the transfer. These deliberations resulted in more than a dozen ideas on how to position the results of the horizon scanning within the organisation. Each concept elaborated on specific goals, target groups, and the estimated time and effort needed to adopt the horizon scanning results in daily work. Concepts developed ranged from low-hanging formats such as signal cards over the adoption of a strategic compass to outreach to other organisations.

The landscape of weak signals, future topics, and emerging issues was used to develop a strategic compass for an organisation that considered embedding its activities intentionally into the dynamic landscape of digital change. In addition to the two layers (future topics and emerging issues), the organisational remits, its departments, and operations represent a third layer of accessibility to digital change. In the digital pilot application of the strategic compass, there is a possibility to switch between the three different layers to identify the most suitable lens through which digital change is viewed. The strategic compass is principally open to incorporating new future topics and emerging issues – and their relationships – into the established knowledge base. It allows for a systematic and transparent treatment of digital change, enabling organisational units to contribute to and benefit from accumulating experiences in dealing with digital change.

3. Reflection on debiasing in the horizon scanning process

The foresight process served as a pilot for the exploration of digital change and its implications in the policy area of the contracting organisation. It was **co-designed by the contracting organisation and the contractor**. Both sides disposed of substantial foresight literacy and were able to mobilise experts in certain newly identified realms (contractor) and actors of relevant organisational units such as the strategy department and the digitalisation department (contracting organisation). The process minimised future biases through measures consistently integrated into every step, from the set-up and conduct of the horizon scanning to the participatory sensemaking and validation of the findings in view of the foresight transfer. It addressed transformative policy goals, improved the quality of insights with transdisciplinary expertise (digitalisation and systems transitions) and collective intelligence building, supported by mutual learning between different actors and stakeholders, and thereby increased the legitimacy of the outcome for policymaking and research decisions.

A few insights from **this particular horizon scanning** on digital futures stand out: the added value of a concise scan field, the need for multiple flexibly adopted approaches to identify signals of change, and the participation in sensemaking. The concise scan field design was instrumental in relating the vast amount of information on digitalisation to concrete realms and provided orientation throughout the project. While the discovery of new signals was guided by cognitive dissonance, the assessment of signals onto the scan field was guided by resonance, thereby reducing the large number of signal candidates to a manageable set. The particularities of this scan field included a reflexive consideration of the changes of digitalisation, as well as the two impact pathways of digital change on the organisation's remit (Figure 1). These were either direct, in terms of research and governance, or indirect, in relation to the digital change of its environment (e.g. the development of digital work skills as a precondition for future recruitment).

A horizon scan produces a snapshot in time but could also be conducted repeatedly or continuously. The contribution of semi-automated scanning and of the other approaches to the outcomes differed. The semi-automated scanning of digital sources yielded results for digitalisation that made it difficult to distinguish the signal from the noise – in contrast to successful semi-automated scanning in other realms, such as biotechnology. The main reason is that terminology around the concept of digitalisation is contained in almost all digitally accessible documents.

Keyword combinations involving the concept of digitalisation and more specific topics such as artificial intelligence yielded useful single sources to be analysed in detail. However, there was still a great deal of noise in the entire repository of digital sources, so topic modelling rarely produced substantially meaningful topic landscapes. It was time- and resource-efficient to consider the other source types (Table 2) and to complement the search with a targeted search machine enquiry that made use of newly identified terms that emerged from sources identified in the topic modelling.

Meanwhile, web scraping bots have improved: Their harvest can be assessed to which extent single sources contribute to meaningful outcomes, which is key to machine learning algorithms that guide the web scraping bot to ever more useful sources.

The composition of the scan team and the involvement of expertise from the contracting organisation and other organisations proved to be key success factors for the interdisciplinary nature of the outcomes of the sensemaking and – finally – for the development of transfer formats. The evaluation of the signals using participatory sensemaking methods is a necessary but demanding task. The involvement of suitable experts contributed to a high degree of accuracy and differentiation of the future topic at hand. The dialogue formats delivered meaningful future topics and emerging issues; however, continuity of participant engagement was not always ensured. The multi-actor future workshop produced a first map of emerging issues, assuming that current trends continue. Upcoming future workshops may take workshop participants further, perhaps involving the assumption that current trends radicalise or become mainstream fast (instead of continuation). The engagement of the various actors contributed to the legitimacy of the process, which in turn could increase acceptance of the outcomes.

4. Conclusion

A horizon scanning of digital change was carried out to derive insights into how the activities of an environment agency are changing and how to seize the opportunities for transformation opening up through digital change. At its core, **debiasing measures were integrated throughout the process** to ensure the novelty, relevance, and validity of the process and of the results. The emphasis was put on how digitalisation affects the ways by which science, policymaking, and daily operations of the organisation are carried out. Signals of digital change were identified and condensed into overarching future topics, and emerging issues were uncovered. Participatory sensemaking enhanced the validity and legitimacy of the findings. The result is the first monograph on research and governance in the Digital Age for a policymaking government agency.

The approach yielded a **disaggregated and multi-perspective view** of digitalisation and its impacts and opportunities accounting for future uncertainty. The monolith 'digitalisation' was decomposed successfully into future topics, subtopics, and related emerging issues. Depending on the organisational units, different facets are of interest. Concepts on how to position the foresight results strategically within the organisation have been developed and handed over. The development of the strategic compass was a collaborative effort of the contracting organisation and the contractor. The actual use of the strategic compass with its accessibility via future topics, emerging issues, or organisational units requires further evaluation.

Promising next steps in the **research of debiasing with horizon scanning** include the development and sharing of smart keyword combinations, as well as leveraging the power of bots to reduce the noise in the web scraped content through machine learning. Debiasing in foresight requires an open and constructive working culture in exchanges between the contracting authority and contractor. The composition of the breakout groups in workshops, as well as more radical methods to drive participants towards the future journey and mitigate biases in future-oriented thinking, have been taken up in recent foresight exercises.[2]

Although the contracting organisation is active in environmental policy, the foresight approach with its debiasing measures is generic enough to **inform organisations in other policy fields** on how digital change can

2 e.g. Horizon scanning of developments relevant to climate protection (https://www.isi.fraunhofer.de/en/competence- center/foresight/projekte/klimascan.html)

be accounted for in their strategy development and daily activities. The digital trends, future topics, and emerging issues can serve other actors from research (e.g. research funding bodies, research-performing organisations, transformative research institutions) and governance (e.g. civil society organisations, government bodies at various governance levels, NGOs, business associations) to reflect their remits under the conditions of digitalisation and to adjust their activities.

Acknowledgements

The authors from Fraunhofer ISI conducted the project 'Environmental research and governance in the Digital Age', which was funded by the Federal Environment Agency under the research code number (FKZ) 3719 11 103.

References

Ansoff, I. H. (1975). Managing strategic surprise by response to weak signals. *California Management Review*, *18*(2), 21–33. https://doi.org/10.2307/41164635

Bitkom (2016). *Digitale Prozesse. Begriffsabgrenzung und thematische Einordnung* [White paper]. Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (Bitkom). https://www.bitkom.org/sites/default/files/file/import/160803-Whitepaper-Digitale-Prozesse.pdf

Borrás, S., & Edler, J. (2020). The roles of the state in the governance of socio-technical systems' transformation. *Research Policy*, *49*(5), 103971. https://doi.org/10.1016/j.respol.2020.103971

Cuhls, K. (2019). Horizon scanning in foresight: Why horizon scanning is only a part of the game. *Futures and Foresight Science*, *2*(1), e23. https://doi.org/10.1002/ffo2.23

Cork, S., Alexandra, C., Alvarez-Romero, J. G., Bennett, E. M., Berbés-Blázquez, M., Bohensky, E., Bok, B., Costanza, R., Hashimoto, S., Hill, R., Inayatullah, S., Kok, K., Kuiper, J. J., Moglia, M., Pereira, L., Peterson, G., Weeks, R., & Wyborn, C. (2023). Exploring alternative futures in the Anthropocene. *Annual Review of Environment and Resources*, *48*(1), 25–54. https://doi.org/10.1146/annurev-environ-112321-095011

Curry, A., & Hodgson, A. (2008). Seeing in multiple horizons: Connecting futures to strategy. *Journal of Future Studies*, *13*(1), 1–20.

Da Costa, O., Warnke, P., Cagnin, C., & Scapolo, F. (2008). The impact of foresight on policy-making: Insights from the FORLEARN mutual learning process. *Technology Analysis & Strategic Management*, *20*(3), 369–387. https://doi.org/10.1080/09537320802000146

De Vito, L., & Radaelli, C. M. (2023). Another brick in the wall?: The case for embedded foresight. *EUI, STG, Policy Brief* 2023/18. https://doi.org/10.2870/0686.

EC (2023). *Trends in the use of AI in science – A bibliometric analysis*. European Commission. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/418191

EC (2024). *Foresight*. European Commission. https://commission.europa.eu/strategy-and-policy/strategic-planning/strategic-foresight_en (accessed 24 March 2024).

Erdmann, L., Kimpeler, S., Gutknecht, R., Cuhls, K., & Rörden, J. (2024). Environmental research and governance in the Digital Age. Results of a horizon scanning for the environment department in tomorrow's society. German Environment Agency (UBA). https://www.umweltbundesamt.de/sites/default/files/medien/479/publikationen/uba_fb_environmental_research_and_governance_in_the_digital_age.pdf

Geurts, A., Gutknecht, R., Warnke, P., Goetheer, A., Schirrmeister, E., Bakker, B., & Meissner, S. (2021). New perspectives for data-supported foresight: The hybrid AI-expert approach. *Futures and Foresight Science*, *4*(1), e99. https://doi.org/10.1002/ffo2.99.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482.

Goddard, M. A., Davies, Z. G., Guenat, S., Ferguson, M. J., Fisher, J. C., Akanni, A., Ahjokoski, T., Anderson, P. M. L., Angeoletto, F., Antoniou, C., Bates, A. J., Barkwith, A., Berland, A., Bouch, C. J., Rega-Brodsky, C. C., Byrne, L. B., Cameron, D., Canavan, R., Chapman, T., … Dallimer, M. (2021). A global horizon scan of the future impacts of robotics and autonomous systems on urban ecosystems. *Nature Ecology and Evolution*, *5*, 219–230. https://doi.org/10.1038/s41559-020-01358-z

Hines, P., Yu, L. H., Guy, R. H., Brand, A. & Papaluca-Amati, M. (2019). Scanning the horizon: A systematic literature review of methodologies. *BMJ Open*, *9*(5), e026764. https://doi.org/10.1136/bmjopen-2018-026764

Holopainen, M., & Toivonen, M. (2012). Weak signals: Ansoff today. *Futures*, *44*(3), 198–205. https://doi.org/10.1016/j.futures.2011.10.002

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgements under uncertainty: Heuristics and biases*. Cambridge University Press.

Krigsholm, P., & Riekkinen, K. (2019). Applying text mining for identifying future signals of land administration. *Land*, *8*(12), 181. https://doi.org/10.3390/land8120181

Loveridge, D. (2008). *Foresight: The art and science of anticipating the future*. Routledge. https://doi.org/10.4324/9780203894156

Miller, R., Rossel, P. & Jorgensen, U. (2012). Future studies and weak signals: A critical survey. *Futures*, *44*(3), 95–197. http://dx.doi.org/10.1016/j.futures.2011.10.001

Mortensen, J. K., Larsen, N. & Kruse, M. (2021). Barriers to developing futures literacy in organisations. *Futures*, *132*, 102799. https://doi.org/10.1016/j.futures.2021.102799

Muench, S., Stoermer, E., Jensen, K., Asikainen, T., Salvi, M., & Scapolo, F. (2022). *Towards a green and digital future*. Publications Office of the European Union. https://doi.org/10.2760/54

Nemorin, S., Vlachidis, A., Ayerakwa, H. M., & Andriotis, P. (2023). AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development. *Learning, Media and Technology*, *48*(1), 38–51. https://doi.org/10.1080/17439884.2022.2095568

Nestik, T. (2018). The psychological aspects of corporate foresight. *Foresight and STI Governance*, *12*(2), 78–90. https://doi.org/10.17323/2500-2597.2018.2.78.90

Popper, R. (2008). How are foresight methods selected? *Foresight*, *10*(6), 62–89. https://doi.org/10.1108/14636680810918586

Rosa, A. B., Kimpeler, S., Schirrmeister, E., & Warnke, P. (2021). Participatory foresight and reflexive innovation: Setting policy goals and developing strategies in a bottom-up, mission-oriented, sustainable way. *European Journal of Futures Research*, *9*, 2. https://doi.org/10.1186/s40309-021-00171-6

Rossel, P. (2012). Early detection, warnings, weak signals and seeds of change: A turbulent domain of futures studies. *Futures*, *44*(3), 229–239. https://doi.org/10.1016/j.futures.2011.10.005

Saritas, O. (2013). Systemic foresight methodology. In D. Meissner, L. Gokhberg, & A. Sokolov (Eds.), *Science, technology and innovation policy for the future* (pp 83–117). Springer. https://doi.org/10.1007/978-3-642-31827-6_6.

Schirrmeister, E., Göhring, A. L., & Warnke, P. (2020). Psychological biases and heuristics in the context of foresight and scenario processes. *Futures & Foresight Science*, *2*(2), e31. https://doi.org/10.1002/ffo2.31

Van Woensel, L. (2020). Scientific foresight: Considering the future of science and technology. In L. Van Woensel (Ed.), *A bias radar for responsible policy-making* (pp. 39–68). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-32126-0_3

Vignoli, M., Rörden, J., Wasserbacher, D. & Kimpeler, S. (2022). An exploration of the potential of machine learning tools for media analysis to support sense-making processes in foresight. *Frontiers in Communication*, *7* (March 23, 2022). https://doi.org/10.3389/fcomm.2022.750614

Warnke, P., & Schirrmeister, E. (2016). Small seeds for grand challenges. Exploring disregarded seeds of change in a foresight process for RTI policy. *Futures*, *77*, 1–10. https://doi.org/10.1016/j.futures.2015.12.001

WBGU (2019). *Towards our common digital future* [Report]. German Advisory Council on Global Change (WBGU). https://www.wbgu.de/en/publications/publication/towards-our-common-digital-future

Wu, X., Ramesh, M., & Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. *Policy and Society*, *34*(34), 165–171. https://doi.org/10.1016/j.polsoc.2015.09.001

List of Authors (alphabetic order)

Manuel 'HonkHase' Atug has been working as a consultant and auditor in information security and the protection of critical infrastructures for over 23 years and deals with the topics of KRITIS, ethics, hybrid warfare, cyber resilience, civil protection and disaster protection. HonkHase has a degree in computer science, a Master of Science in Applied IT Security and is an engineer.

Thomas Bartoschek is post-doc at the Institute for Geoinformatics at the University of Münster and co-founder and managing director of Reedu. In his research and development work, he focuses on technologies for Citizen Science and the future of education interfacing with sustainability research. He has received a number of awards for his work, most recently the Open Data Impact Award and the ACM Eugene Lawler Award for humanitarian contributions to Computer Science.

Jana Baum is a research associate in the research programme “Digital Democratic Innovations”. Her research interest is to explore the question of subjectivity and self-determination in the context of digitality and artificial intelligence. Since 2022 she is Research Assistant at the Institute of Philosophy / University of Leipzig (Lecturer for Theoretical Philosophy).

Nikolas Becker heads the Politics & Science division of the Gesellschaft für Informatik e.V. (GI) and the Netzwerk Digital GreenTech. After earning his M.A. in Political Science (Universität Potsdam, UCSD) with a focus on the international political economy of intellectual property rights, he worked as a product manager for the Open Knowledge Foundation International.

Grischa Beier studied mechanical engineering at TU Ilmenau, UFSC (Brazil) and ITMO (Russia). Since September 2014, he has been working at RIFS Potsdam, where, as head of the research group “Digitalisation and Impact on Sustainability”, he explores the social and ecological effects of future digitalised and networked industrial production, among other topics. Since 2023, he has held a professorship in Sustainability in Digitalisation at the University of Potsdam.

Christoph Bieber is a research professor at CAIS with the research programme “Digital Democratic Innovations” and focuses on the research areas like Digitalisation and Democracy; Elections and election campaigns; Ethics and responsibility in politics; Transparency and Public Communication.

Bonny Brandenburger graduated from the University of Lille in a German-French double master's programme in the field of cultural studies and economics. She then worked for four years in various projects on the topics of international vocational training, transnational science communication and community participation of new immigrants funded by the German Federal Ministry of Education and Research (BMBF).

Francesca Bria is an economist, digital policy expert, and information technologist working at the intersection of technology, geopolitics, economics and society, who lectures at various universities and is a consultant amongst others to the United Nations and the European Commission.

Florian Butollo is Professor of Sociology with a focus on “Digitalization of the world of work” at TU Berlin and part of the research group “Working with Artificial Intelligence”. He is responsible for several projects based at the Social Science Research Center Berlin and associated with the Weizenbaum Institute - on the consequences of the pandemic for digitalization in the world of work (funded by BMAS), on the regulation of algorithmic management in the world of work (EU Commission), and on the restructuring of global value chains in the wake of the pandemic (DFG).

Anouk Cenan is a scientific assistant for the research programme Digital Democratic Innovations.

Tina Comes is a Professor in Decision Theory & ICT for Resilience at the TU Delft (NL). Since her PhD, she has been determined to better understand how digital technologies drive decision-making of individuals and groups in the context of risk and crises. At the TU Delft, she leads the Resilience Lab and heads the disaster resilience theme for the Delft Global Initiative. She serves as the Scientific Director for the 4TU.Centre for Resilience Engineering, a collaboration of the 4 technical universities of the Netherlands.

Jordana Composto is a Psychology and Social Policy PhD student at Princeton University. Her research interests include organizational behaviour, social norms, and pro-social decision making in the private sector. At Princeton, she is advised by Professor Elke Weber. She is affiliated with the Department of Psychology and the Andlinger Center for Energy and the Environment.

Teresa Cerratto Pargman is a Human-Computer Interaction (HCI) professor at the Dept. of Computer and Systems Sciences at Stockholm University (SU). With a background in Psychology, Pedagogy, and Cognitive Science, she conducts conceptual and empirical studies on emerging technologies and practices in the public sector drawing from perspectives on the philosophy of technology and science and technology studies.

Nicole Döpp is a project manager at Smart City in Süderbrarup. Here, she is passionate about digitalisation and increasing efficiency in public administration with the aim of optimising processes and making work easier for everyone involved.

Lorenz Erdmann leads the Business Unit Foresight for Research and Civil Society at Fraunhofer ISI's Competence Center Foresight. He graduated at the Technical University Berlin in Environmental Sciences and Technology. Then he worked for a decade at the Institute for Futures Studies and Technology Assessment IZT on sustainable digitalization, before joining Fraunhofer ISI's Competence Center Foresight in 2011.

Niklas Frechen is PhD student and research associate in the research programme “Digital Democratic Innovations”.

Anne Goldmann is PhD student and research associate in the research programme “Digital Democratic Innovations”.

Jennifer Haase is a research associate at the Humboldt University in Berlin, specifically at the Institute of Computer Science. Additionally, she contributes to the research group "Security and Transparency of Digital Processes" at the Weizenbaum Institute. Jennifer's research results have found their way to top-tier journals in psychology, creativity, and information systems, and she's been a notable presence at numerous conferences.

Andrea Hamm received her doctoral degree in Digital Media and Technology from the Dep. of Political and Social Sciences at Free University Berlin, Germany. She is now researcher in the group "Digitalization, Sustainability, and Participation" at the Weizenbaum Institute for the Networked Society. She is researching interdisciplinarity between the fields Human-Computer-Interaction and Journalism Studies. In her current work, she is interested in better understanding modes of participation in varying contexts of digitalization projects.

Pauline Heger is a research associate in the research program "Digital Democratic Innovations". Since August 2023, she has been on the doctoral list at TU Dresden. The working title of her dissertation is: "Mapping Social Acceptance of Green Innovation in Urban Settings". The goal: to better understand social processes that arise from the interaction between politics and civil society when innovative technologies are to be integrated into urban living spaces. She has a particular interest in climate-friendly Technologies.

Mennatullah Hendawy is an interdisciplinary urban planner working at the intersections of cities, media, and technology toward equity and sustainability. She is an assistant professor at Ain Shams University in Cairo, Egypt, and is also affiliated with the Center for Advanced Internet Studies, Technical University of Munich, and Impact Circles e.V., in Germany.

Christian Herzog is a transdisciplinary researcher with a PhD in engineering and an M.A. in applied ethics. Since 2023, he has been appointed Professor for the Ethical, Legal, and Societal Aspects of AI at the University of Lübeck. Christian Herzog investigates ethical issues by focusing on power dynamics, epistemic injustice, and a non-reductionist approach to centering on the diverse ways epistemes, identities, and values are generated, contested, proliferated, and protected within technologically mediated assemblages.

Luca Hesse recently completed a Bachelor's degree in Geoinformatics and is now continuing with a Master's program at the Institute for Geoinformatics in Münster. Currently, he works as a student assistant at con terra, focusing on WebGIS development. His recently completed Bachelor's thesis examined the integration of machine learning with me-of-flight sensor technology.

Lena Hoffmann is a Consultant at the Gesellschaft für Informatik e.V. (GI), where she works on the project “Enabling Green Computing and Digital Transformation” (ECO:DIGIT). She also served as a sustainability expert in the BMUV’s AI flagship project “Green Consumption Assistant”, which focused on the AI-supported promotion of sustainable consumption on the internet. She holds a Master’s degree in Sustainable Resource Management (M.Sc.) from the TU Munich and was involved in organising the Bits&Bäume Conference in 2022.

Sigrid Kannengießer is professor of Communication Studies with a focus on Media Sociology. Previously, she was professor of Communication and Media Studies with a focus on Media Society at the University of Bremen. She rejected appointments as a professor to the University of Paderborn and the Zeppelin University Friedrichshafen.

Benjamin Karic is a research associate and PhD candidate at the Institute for Geoinformatics, University of Münster. With a strong foundation in the Internet of Things from his bachelor's and master's degrees in geoinformatics, he now dedicates his research to enhancing the energy efficiency of Internet of Things applications. He focuses on reducing data transmission through the integration of onboard AI methods and advanced engineering of wireless sensor networks.

Ann Katzinski is a research associate in the project "Generative AI in the Workplace: Application, Impact and Design" within the research group "Globalization, Work and Production" at WZB I Berlin Social Science Center and an affiliated researcher at the Weizenbaum Institute. She previously conducted research there in the project "ELIZA reloaded", examining potential changes in knowledge work through generative language-based AI systems, using ChatGPT as a case study. Her current research at WZB builds on this, focusing on the interplay between generative AI and programmers.

Krishnan Radhika has been post-doctoral fellow at the Centre at Concurrences in Colonial and Post-Colonial Studies at Linnaeus University in Sweden. She was also a Fellow at the Indian Institute of Advanced Study (IIAS) in Shimla, India. She also has several years of leadership experience in product management collaborating with cross functional engineering teams, SMEs and stakeholders to deliver customer centric and high-quality business and IT solutions.

Caroline Krohn is co-founder and spokesperson of the Working Group Sustainable Digitalisation. This initiative has set itself the goal of focusing on protecting future generations in times of increasing digitalisation. She is also an entrepreneur in the fields of IT security, IT expert recruitment and sustainable corporate management. She is also active in digital policy with Bündnis90/Die Grünen, for which she ran for the state parliament in Hesse in 2018 and for the Bundestag in 2021.

Felix Kronlage-Dammers has been active in various open-source development communities, including DarwinPorts, OpenDarwin, OpenBSD and currently, the Sovereign Cloud Stack. His interests range from monitoring and observability to infrastructure-as-code to building and scaling communities and organisations. He is a techie at heart and enjoys empowering others to do great things. He is part of the extended board of OSBA and describes himself as a Unix/Open-Source nerd.

Anne K. Krüger is a sociologist studying the reorganization of knowledge practices and evaluation through digitalization. Her research is situated at the intersection of Critical Data Studies, Organizational Sociology, the Sociology of Valuation and Evaluation, and Science and Technology Studies. She is currently working on digital infrastructures in research evaluation and the use of generative AI in science.

Furkan Koc is a student of Political Science at the University of Duisburg-Essen. Areas of research and interest include the implementation of artificial intelligence in politics, conflict research in the Balkans and co-operation in international development aid between security communities.

Thomas Kox studied geography with minors in political science and sociology at the Rheinische-Friedrich-Wilhelms-Universität Bonn, at the University College Cork, and at the University of Potsdam (doctorate). From 2011 to 2019, he worked as researcher at Freie Universität Berlin (FUB) and from 2019 to 2023 at Ludwig Maximilian University of Munich (LMU). During the summer term 2024 he was interim professor in Earth Science Communication and Climate Education at LMU. His work is on the perception and use of uncertainty, disaster and civil protection practices, as well as weather impacts, warnings and citizen science. Since November 2022, he leads the research group "Digitalization and Networked Security" at the Weizenbaum Institute.

Verena Majuntke has worked for more than six years at Bosch Software Innovations and Bosch Connected Industry. Since 2019, she has held a professorship in Software Engineering at HTW Berlin, where she maintains a strong focus on Green Coding, Requirements Engineering and Programming Education.

Florian Meissner is Professor of Media Management and Journalism at Macromedia University of Applied Sciences in Cologne, Germany. He serves as the Chair of the Risk and Crisis Communication Section of ECREA and is the Principal Investigator of a research project called "Digital Fitness for Citizens." This project is funded by the German Federal Ministry of Education and Research with the aim of promoting cybersecurity awareness and behaviour at the societal level.

Anne Mollen's research focuses on sustainability and digitalisation as well as automation in media communication and data-centric technologies. She studied Communication and Political Science at WWU Münster (B.A.) from 2006 to 2009 and International Communications at the University of Leeds (M.A.) from 2009 onwards. She received her PhD in 2018 with a thesis on the socio-technical constitution of digital communicative spaces. She was the chair of YECREA from 2016 to 2018.

Meyke Nering Bögel is a consultant at the systems change company Systemiq in Amsterdam (NL). Prior to that, she worked as a researcher at the Resilience Lab at the Faculty of Technology, Policy and Management at the University of Technology in Delft, focusing on emergent behaviour in refugee settlements. She holds a MSc degree in Engineering and Policy Analysis from the TU Delft.

Jan Magnus Nold holds a Master degree in "IT-Security and Information Engineering". During his studies, he focused on human behaviour. Now he is a PhD candidate at the chair for Human-Centred Security at Ruhr University Bochum. His work within the project "Digital Fitness for Citizens", funded by the German Federal Ministry of Education and Research, focuses on technology assisted behaviour change for personal cyber security.

Seyi Olojo is PhD student in the School of Information with a designated emphasis in Science and Technology Studies at The University of California, Berkeley. Her research explores how colonial histories of racial classification are re-articulated within contemporary machine learning practices, specifically data collection and early-stage machine learning. Additionally, she investigates the relationship between technical failures and identity formation among historically marginalized communities.

Rebecca Panskus holds a Master Degree in "Cybersecurity & Information Technology" and is a PhD candidate in Cybersecurity and Privacy at the Digital Sovereignty Lab at Ruhr University Bochum, Germany. Her work within the research project "Digital Fitness for Citizens" funded by the German Federal Ministry of Education and Research, focuses on Cybersecurity and Privacy of children, teenagers and young adults from a human-centred perspective.

Robin Preiß is a doctoral candidate and member of the working group of the Ethical Innovation Hub at the University of Lübeck. He holds an M.A. in Science and Society from the Leibniz University of Hannover. His current research focuses on digitalisation approaches in the public administration sector. The objective of his research is to investigate ethical and social questions that emerge with technological changes and to analyse the impact of these transformations on society.

Lubna Rashid is an academic, educator, activist and artist. Lubna's work lies at the intersection of ecological sustainability, behavioral psychology, digital entrepreneurship & innovation, and social justice. She is particularly interested in understanding the drivers of (un)sustainable decision-making and behavioral change for sustainability. She has extensive experience with post-conflict and refuge settings and has eclectic backgrounds, including in ecology and the performing arts.

Angela Sasse studied psychology in Wuppertal and Great Britain in the 1980s. She completed her Master's degree in Occupational Psychology in Sheffield and gained her doctorate in Computer Science at the University of Birmingham. From 2003, she was Professor of Human-Centred Technology at University College London. In 2018, she took over the Chair of Human-Centred Security at the Horst Görtz Institute for IT Security at Ruhr University Bochum and was admitted to the German National Academy of Sciences Leopoldina in 2023.

Paula Scharf is a software developer at Reedu where she is currently mainly involved in the research project TinyAIoT. In the project she investigates applying sustainable, energy- and resource-efficient machine learning models on especially small and constrained hardware. Previously she has pursued her bachelor and master degree of geoinformatics at the University of Münster which she finished in 2023 with a final thesis about using machine learning for insect detection.

Ina Schieferdecker is Honorary Professor for Software-Based Innovations at the Technical University of Berlin. She was Director of Research for Technological Sovereignty and Innovation at the German Federal Ministry of Education and Research, focusing on modern key technologies such as quantum technologies, materials technologies, microelectronics, communication, software and AI technologies, as well as the future of work and value creation.

Josephine B. Schmitt, is scientist and certified Scrum Master at the center for Advanced Internet Studies (CAIS). Her expertise lies at the intersection of media psychology, communication sciences, media education, science and technology studies. Among others, she shapes and explores new methods of collaboration among researchers. She focuses on creating and translating scientific insights into practical applications using innovative methods. In her research, she also delves into topics such as civic education on the internet, hate speech and fake news.

Yuya Shibuya is an associate professor at the University of Tokyo, Japan, focusing on the interaction between virtual and real worlds. She is interested in how the virtual and real worlds interact with one another. In particular, her research interests have lied in how to ensure and enhance democratic participatory structures in the digital era. Mainly based on but not limited to quantitative and mixed methods, she's have researched people's socio-economic activities, people's behavioral changes, and digital media activities.

Samuel T. Simon is a certified agile coach and process facilitator. He teaches at the Ruhr University Bochum and is a researcher at the Center for Advanced Internet Studies GmbH (CAIS). He is completing his doctorate on the application of various innovation approaches in interdisciplinary scientific projects (particularly research on digital transformation). He is investigating how these approaches can be used to actively shape social transformation processes and promote sustainable change.

Marc Steen works as a senior research scientist at TNO, a leading research and technology organization in the Netherlands. He earned MSc, PDEng and PhD degrees in Industrial Design Engineering at Delft University of Technology. He worked at Philips and KPN before joining TNO. He is an expert in Human-Centred Design, Value-Sensitive Design, Responsible Innovation, and Applied Ethics of Technology and Innovation.

Felix Sühlmann-Faul is an expert in digitalization and sustainability, holds a doctorate in technological sociology, is a speaker, consultant and author. He was a test manager in Daimler customer research and a project manager at the Institute for Transportation Design. He advises the German Sustainability Award, among other things, and was involved in setting up a Germany-wide research network on digitalization and sustainability.

André Ullrich heads the research group "Digitalisation, Sustainability and Participation" at the Weizenbaum Institute for the Networked Society and is currently doing his habilitation on responsible and sustainable digitalisation. His work focuses on the opportunities and risks of digitalisation for the socio-ecological transformation of companies and society.

Martijn Warnier is full Professor of Complex System Design and head of the Multi-Actor Systems Department - Faculty of Technology, Policy and Management - Delft University of Technology. He holds a PhD degree in Computer Science from the Radboud University Nijmegen (2006). In his research prof. Warnier studies complex emergent behaviour in and of socio-technical systems, specifically of power and ICT infrastructure and the combination thereof.

Elke U. Weber is a German-Canadian psychologist and professor of psychology and public affairs at Princeton University. She is Visiting Fellowship at the Technische Universität Berlin, where she researches political decision-making powers and global action to tackle climate change through the behavioural and social sciences. Elke Weber will draw on psychological theories to investigate the extent to which personal experiences, as well as social norms and interactions, affect decision-making and action.

Ines Weigand graduated in 2020 from the Berlin University of the Arts (UdK) with a Master's degree in Communication in Social and Economic Contexts. In her collaborative Master's thesis, she investigated potentials of the Open Science movement for a change in the relationship between humans and nature. In order to investigate this problem, which is typical for the Anthropocene, she used methods of critical making and experiential learning.

Jan Wieland is a student of political science at the University of Duisburg-Essen.

Alexander Wilke holds a Master Degree in Media and Communication Management and is a Project Researcher at Macromedia University of Applied Sciences in Cologne, Germany. His work is focused on the research project "Digital Fitness for Citizens" funded by the German Federal Ministry of Education and Research. He has a professional background as a communications specialist and consultant in a nationwide PR agency and is experienced in communication conception, realization and evaluation from a practical perspective.

Annemarie Witschas is a member of the research group "Ethics and critical Theories of AI" at Osnabrück University. With an interdisciplinary background in Cognitive Science, she has an expertise in both AI and critical philosophy. Her research interests broadly span power differentials in the digital realm, ranging from the impact of AI technologies in pornography, to exploring the effects of AI discourse on collective futures thinking, as in their current MWK-funded project "Predicted Futures?"".

Herbert Zech holds the Chair of Civil Law, Technology Law and IT Law at Humboldt University, Berlin, and is a Director at the Weizenbaum Institute for the Networked Society. After his habilitation in Bayreuth, he worked at the University of Basel from 2012, initially as an associate professor for private law with a focus on life sciences law, and from 2015 as a full professor for life sciences law and intellectual property law. His focus is on technology law, intellectual property and digitalization.

Daniela Zetti is a historian of media technology at the University of Lübeck. Recent publications: Bory, Paolo, and Daniela Ze. 2022. "Digital Federalism. Balancing Autonomy, Authority and Automation" 2024. "Writing the History of Computing: The First Half Century." In Writing Computer and Information History. Approaches, Connections, and Reflections, edited by William Aspray, 163–90. Lanham: Rowman & Littlefield.