

Introduction

*Andreas Sudmann, Anna Echterhölter, Markus Ramsauer, Fabian Retkowski,
Jens Schröter, Alexander Waibel*

For more than ten years now, we have witnessed an AI boom affecting basically all areas of culture and society, including the scientific field. This book explores the potentially profound transformation in academic research.

Such a focus is not only aiming at the question of AI's impact, be it as a technology, a component of a larger infrastructure, or a tool. It is also about exploring what AI as a concept actually means, which different techniques and approaches it addresses, to what extent it might be important to continue the long tradition of problematizing it, and last but not least, how a particular understanding of AI might be transformed by the practices and conditions of its scientific situatedness and application (Suchman 2006).

As a point of departure for the following considerations, we engage with the history of AI as a contest between two fundamental approaches: the symbolic and the subsymbolic (see also Dreyfus/Dreyfus 1988: 15–43). The former, also known as GOF AI (Good Old-Fashioned Artificial Intelligence), processes knowledge and tasks based on logical or rule-based procedures. Knowledge is explicitly represented, often hard-coded and manually entered into the system by experts. The latter is characterized by the fact that corresponding procedures seek to find patterns and correlations in data automatically. This approach involves statistical and neural models to learn from data without relying on explicitly defined rules. Knowledge representation operates in an implicit manner. For example, knowledge can be implicitly encoded in the weights of a neural network. While this allows these systems to process sizable amounts of complex, unstructured data, it is also responsible for their black-box nature.

Subsymbolic AI and its scientific impact are the focus of this book. More specifically, the contributions from various fields shed light on artificial neural networks (ANNs) as the currently dominant and discourse-determining forms of AI, which are broadly inspired by the neuroinformatic model of the brain

(mostly related to humans, but also with regard to animals).¹ In fact, *prima facie* it seems as if the subsymbolic approach of ANNs (mostly even only based on backpropagation) has become synonymous with AI as it has largely supplanted symbolic AI and even other machine learning methods, including symbolic learning, statistical learning, and Hidden Markov models, among others.

However, such a thesis must be differentiated more precisely in at least two respects: on the one hand with regards to the long tradition of hybrid connections of symbolic and subsymbolic methods; on the other hand that this distinction becomes blurred when traditionally symbolic problems (e.g., language processing) are increasingly handled on a neural substrate as well (e.g., machine translation, parsing, large language models (LLMs)). The term “neuro-symbolic AI” is pertinent in this context, as it refers to hybrid systems that integrate neural models with symbolic AI. In his Robert S. Engelmore Memorial Lecture at AAAI 2020, Henry Kautz (2022) provided a taxonomy of neuro-symbolic AI systems. One of the categories he introduced was “symbolic Neuro symbolic” (*ibid.*: 118), which also directly applies to LLMs. Systems in this category have their inputs and outputs presented as a symbolic form and even natural language with its discrete tokens counts towards that. Although these systems are not widely regarded as neuro-symbolic, it does make the term more ambiguous. The “Neuro[Symbolic]” (*ibid.*: 119) category may be of greater interest and relevancy, as it embeds symbolic reasoning as part of the neural engine. A new development with contemporary LLMs, such as ChatGPT or Toolformer, is the ability to interact with plugins. One such plugin can be a symbolic reasoning engine like WolframAlpha.

However, the relations and connections between symbolic AI and the subsymbolic AI of ANNs are no impediment for focussing on the latter for the purposes of this book – on the contrary. Above all, given their specific history, the current relevance of ANNs is quite remarkable. Their technical foundations were already developed as early as the 1940s and 1950s (Sudmann 2018a; Sudmann 2018b), but more complex, foundational architectures emerged in the 1980s and 1990s as they enabled ANNs to operate on real-world problems that required context, shift invariance, or sequential processing (Waibel et al. 1987; Waibel et al. 1989; Hochreiter/Schmidhuber 1997; LeCun et al. 1998). Nevertheless, until the 2000s, ANNs were largely ignored by practitioners and struggled to find broad adoption: Because the computing of the day would

1 In addition to neural networks, other approaches or algorithms can also be subsumed as subsymbolic AI, for example k-nearest neighbors.

only permit training of small networks, simpler statistical methods could already deliver competitive performance. It would take 20 more years until computational resources and data had scaled sufficiently for ANNs to show their true potential: From networks with a dozen or hundred connections and a single hidden layer, we now see networks with 175 billion parameters (GPT-3) and dozens or hundreds of layers. And ANNs could now deliver (with the same or similar algorithms as in the 1980s) impressive performance advances over classical methods. In speech recognition, error rate reductions of 30% or more were observed on published benchmarks. In vision, significant improvements could be obtained over standardized object classification benchmarks (see ImageNet, Krizhevsky/Sutskever/Hinton 2012). And even in machine translation, performance leapt forward through the adoption of large recurrent neural encoder-decoder networks (Luong/Manning 2015). In many domains, e.g., speech (Nguyen/Stueker/Waibel 2020), vision, machine translation, performance now exceeds human capabilities over certain defined benchmarks.

Another decisive part was played by big tech. The AI renaissance was accelerated as soon as the information industry became aware of the economic potential of ANNs. This resulted in a concerted move to massively expand AI research activities, invest in computing resources, and to acquire and merge promising AI start-ups (like DeepMind, and others).

The widely broadcasted 2016 victory of the AlphaGo program over Go master Lee Sedol had a further reinforcing effect with regard to the perception of the lingering capabilities of AI. This media event significantly shaped public perception. Subsequently, experts in various scientific fields were alerted, and increasingly interested in AI and, ultimately, began to integrate the new technology into various methodological toolkits. Somewhat unexpectedly, the release of ChatGPT in 2022 proved to be another game-changer. AI could finally be experienced and utilized by a wider circle of users, an encounter that swept public perception, and made it impossible to overlook the ramifications of this new technology for the most basic practices of mainstream science, its quotation standards, and academic exams. Besides questions of authorship and reliability, one important provocation may lie with the political and moralistic overtones of these chatty AIs. Furthermore, as language models, they merely predict text based on massive amounts of past textual data and thus ethical standards or factual correctness can not be assured as of yet. Even if a majority-driven form of reinforcement learning from human feedback decides about the biases of such machines, a “mathematisation of ethics” and a quantitative

vote for majority morals is at hand (Simanowski 2023: 73). Still, large self-supervised models like LLMs can digest virtually all of humanity's textual data and thus generate predictions with surprising accuracy and relevance, resulting in a powerful illusion of human-like intelligence and clarity.

Nevertheless, it is necessary to unravel this rather event-centered and also person-centered historiography in more detail. For example, backpropagation as a central learning algorithm of ANNs was already developed in the 1970s and 1980s, some elements of it even as early as the 1960s.² Accordingly, it is difficult to attribute the development of this algorithm to just one person or one group of people at a specific time. Moreover, it has been and continues to be the case that the development of AI in many areas is based on close cooperation between industry and science, but also the military. Not least for this reason, AI research has always been, to a considerable extent, applied research.

To understand these transdisciplinary effects of the new technology, we must examine the level of data practices and scientific methodologies. Several recent publications have been addressing the impact of new AI technologies on scientific practices (Athey 2018; Fecher et al. 2023; Gethmann et al. 2022; Okerlund et al. 2022). At the same time, it seems evident that we are witnessing the effects of a much longer history of data, statistics, formalization, modeling, and simulation. Since the early days of AI, attempts were made to put 'intelligent' systems to use in various academic settings³, but the corresponding reflections, if they had their place in the sciences at all, remained, in most cases, either necessarily speculative or their lasting contribution to the development of a research field ultimately proved to be extremely limited. There were, for example, early attempts to use AI systems for specified scientific tasks such as proving theorems (see Feigenbaum/Feldman 1963; Dick 2011), but corresponding implementations of the systems were typically very far from actually advancing research in the respective areas of knowledge.

With the successive establishment of so-called expert systems starting in the 1970s, the application-oriented perspective of AI finally gained some relevance, but this upswing ultimately did not last either. It is quite telling that Pamela McCorduck's relevant study on the history of AI – *Machines who Think* –

2 For a technical history of backpropagation related to ANNs, see for example Schmidhuber (2022).

3 For discussing AI in the context of psychology, see for example Hunt (1968: 135–168); for organic chemistry, see Feigenbaum (1968: 23–27).

contains a separate chapter titled “Applied Artificial Intelligence”, which introduces two of these early expert systems and their respective application contexts. But remarkably enough, this chapter begins by pointing out how AI is derided and mocked in terms of its supposed potential on a regular basis (McCorduck 1979: 272f.).

The latter has not fundamentally changed today, even in light of the considerable achievements of large language models like ChatGPT. There continues to be a pronounced interest as well as a certain pleasure to expose the shortcomings of even the most advanced AI systems. Nevertheless, there is a significant shift in this respect: Currently, AI is no longer a speculative concept at its core; the relevant point of reference for (critical) reflections now is the concrete implementation of corresponding systems, not only with respect to areas of academic knowledge but all areas of culture and society.

There is little doubt about the fundamental importance of AI in all spheres of social life, given the prevailing assessments in public discourse. Furthermore, there seems to be no sign of an imminent end to today’s AI boom. Following many booms and busts of previous AI excitement and promised revolutions, AI has now found its firm and sustainable footing. This is especially true for applications of AI in various fields of science, as countless research examples demonstrate (for an exemplary overview of AI research projects in Europe, see “How is Artificial Intelligence Changing Science?” 2023).

Unsupervised and self-supervised algorithms and the increasing use of simulations and data augmentation have advanced practical AI applications to astonishing performance levels and opened new applications. Sharing of open-source code, tools and large pretrained models now also accelerate progress by leapfrogging from one accomplishment to another at unprecedented speed. Google DeepMind, for example, has released a series of specialized models that aim to assist researchers in their respective fields, including AlphaFold (Jumper et al. 2021) which is able to predict 3D structures of proteins more accurately than previous models and, more importantly, is in many cases accurate enough to replace real-life experiments. AlphaFold is arguably the organization’s biggest success so far and is now deeply ingrained as a tool in medicine and life sciences (Varadi/Velankar, 2022). More recently, Google DeepMind published AlphaTensor (Fawzi et al. 2022) and AlphaDev (Mankowitz et al. 2023), both of which have been used in the research area of computer science to optimize algorithms and low-level code such as matrix multiplication and sorting algorithms. In the case of AlphaTensor, the model was able to find an algorithm to reduce the number of multiplications neces-

sary for certain types of matrix multiplications. On the related blog posting, Google DeepMind's headline "optimising the world's code one algorithm at a time" (Mankowitz/Michi 2023), aptly describes its current approach.

At present, the contributions of AI to scientific challenges are not always as spectacular as in the case of AlphaFold; often enough, standard AI technologies are used as elements of methods or in everyday applications (although usually at much better performance). However, it is remarkable how diversely and broadly AI is now being applied in various fields of research. In sports science, ML-based pattern recognition is increasingly used for the performance analysis of athletes, players and teams (Araújo et al. 2021). In art history, a computer vision system has been able to identify connections between artworks by analyzing poses of human subjects in paintings (Jeniczek & Chum, 2019). ML has also, for a long time, been used in particle physics, due to the enormous datasets analyzed in this field. In 2012, one of the important discoveries, the Higgs boson, owed much to the application of machine learning (Radovic et al. 2018; Bourilkov 2019).

Even though the general AI boom has been felt in many scientific fields for years now, one should note that the application of AI in many disciplines is still in its infancy. In our view, it is therefore even more important and timely to recognize, reflect on, and historically document this transformation of the sciences by AI in *statu nascendi*. To address this challenge, our transdisciplinary research group, encompassing the disciplines of media studies, computer science, and the history of science, has started its work in 2019, respectively 2022, to investigate the ways in which research is conducted not only on/about AI but *with* AI, in various fields from the natural and social sciences to the humanities. In particular, we are interested in exploring how AI interacts with the established practices and methods of science, whether they are complemented, modified, and/or potentially replaced.

Three disciplines or domains of research are at the center of our inquiry: environmental sciences/climatology, social sciences/sociology, and film studies. Three additional fields – literary studies, medicine and economics – are investigated to broaden the range of disciplines to be studied, partly in order to capture the heterogeneous range of uses of AI more accurately and to better generalize our results across scientific disciplines. In a first programmatic paper, the research group has already discussed some key challenges and perspectives, as well as some general considerations (Echterhölter/Schröter/Sudmann 2021).

The question of the transformations of the sciences through AI requires description of precisely their different scales and dimensions, as well as the general heterogeneity of the aspects addressed by them. One way of illustrating the range of conceptualizations can be the marking of extreme positions and ways of thinking, thus allowing for a more nuanced perspective. For example, one might argue that as an advanced Artificial General Intelligence (AGI)⁴ system evolves, it would also be capable of handling any (new) scientific problem. Another option would be to develop a system, however specialized, that is used for more or less specific scientific tasks or only within a certain domain or discipline. Both concepts can be imagined as systems of a “superintelligence”, to pick up Bostrom’s (2014) term, insofar as the abilities and skills of human systems are (or can be) clearly surpassed in both scenarios.

In the emphatic sense, AI stands for the possibility of a computer being able to gain its own insights, formulate questions and hypotheses at some point, and thus also complete all other steps along this path more or less autonomously. AI systems used for scientific (research) purposes can be further differentiated according to how human-like they have been designed and oriented. Explainable AI requirements make it at least likely for machine communication to remain connectable to human understanding and control.⁵ This also applies to future machine-machine communication. AI processes can also be differentiated according to the extent to which they organize individual components/phases of scientific research processes autonomously or automatically, from formulating a research question to collecting data, analyzing and evaluating data, as well as presenting and disseminating research findings.

Furthermore, there is the fundamental question of which scientific problems seem a priori suitable to be addressed by AI at all. DeepMind (Hassabis 2022) has developed three criteria in this respect:

-
- 4 The notion denotes a hypothetical AI-system with cognitive, creative etc. capabilities comparable to or even exceeding those of humans. There are no realizations of such systems yet. Their development, if it will be possible at some point, is repeatedly discussed as a great danger.
 - 5 In recent years, research on XAI systems has become increasingly important, and this applies in particular to scientific applications of AI. An overall very promising project in the German-speaking area was recently initiated with the Transregio “Constructing Explainability” of the Universities of Bielefeld and Paderborn.

1. massive combinatorial space,
2. clear objective function (metric) to optimise against,
3. either lots of data and/or an accurate and efficient simulator.

As can be seen from these criteria, the use of ML must be carefully considered, especially with regard to the significant resources and costs involved.

A relatively recent phenomenon to reflect within our study is the fact that more and more explicitly AI-driven tools or apps are either directly intended for scientific work or can be indirectly used for it. A plethora of commercial applications like SciSpace Copilot or Elicit has been launched, promising to automate certain research workflows and help with literature research or understanding literature. Language models like ChatGPT are actively used by researchers as assistance in the writing process of scientific documents, prompting repositories and journals like arXiv to define a ‘use of generative AI language tools’ policy for authors. Towards the end of last year, the domain-specific language model Galactica (Taylor et al. 2022) caused a stir among the research community. It is exclusively trained on scientific data like research papers, chemical formulas, and DNA sequences. The generated text of the model sounded convincingly scientific but triggered concerns that it could easily spread inaccuracies. At the same time, there is a class of models, such as Minerva (Lewkowycz et al. 2022) and AI Descartes (Cornelio et al. 2021) that are used in research itself and are intended to automate reasoning processes.

Beyond such specific applications, it seems important to us to explore the general tool character and principal potential of current data-driven, statistical AI systems in methodological terms. A few years ago, computer scientist Pedro Domingos described ML as the “scientific method on steroids” (Domingos 2015: 13). Such a description strikes us as highly questionable as it conceptualizes ML per se as a scientific method. In addition, the metaphor “on steroids” suggests that ML allows an almost illegal and unhealthy form of performance enhancement in this respect. Nevertheless, it is obvious that the performance level of learning algorithms significantly increases when corresponding systems are trained with more and more data and computational power. The present publication, therefore, is also motivated by an interest in discussing AI through the lens of the ways in which learning algorithms potentially reconfigure the epistemic relationship of qualities and quantities. More specifically, we would like to shift the perspective on this relationship by highlighting epistemic aspects *beyond quantity* and thus also illuminate

perspectives beyond the dominant relation of AI and big data. Two aspects are particularly important to us in this respect:

Firstly, current approaches to AI, i.e., subsymbolic AI in the form of ANNs, are not merely capable of extracting information from large amounts of data and making it productive, but they can solve problems that can also be described as qualitative. They involve dealing with qualitative questions of content, aesthetics, style, e.g., in the field of natural language processing or computer vision, in ways that were unimaginable until recently.⁶

Secondly, current approaches in AI research are increasingly focused on reducing or avoiding dependence on large amounts of labeled data, e.g., through strategies of self-supervised learning, zero- or one-shot learning, transfer learning, or even the use of synthetic data or simulation data. Contemporary LLMs are, for example, the result of causal language modeling which is a type of self-supervised learning in the course of which the model is tasked with predicting the next token in a sequence while requiring no additional labels.

As a result of our project's opening conference hosted by the Sorbonne Center for Artificial Intelligence (SCAI), we present first explorations on the subject of AI in the natural sciences and humanities at a point in time where qualitative problems seem to come into reach to be handled by machines. At the same time, these discussions of European scientific applications tie in with concerns that lie beyond this subject area and concern general preconditions of digital humanities (DH) or also of STS. Certain problems of transformative processes in the scientific field, which are closely related to AI, emerge *mutatis mutandis* in other constellations as well.

It is important to keep in mind in this context that research on scientific practices in AI has been conducted in a wide variety of disciplines and analytical perspectives, such as science and technology studies, sociology, infrastructure studies, cultural anthropology, philosophy of science, and data science (Baurmann/Mans 1984; Carley 1996; Groß/Jordan 2023; Krämer 1994; Ligo et al. 2021; Manhart 1995), but also specifically in the disciplines involved in this project: media studies (MS), history of science (HS) and computer science (CS) itself.⁷

6 For the details, see our (the editor's) contribution to this book as well as the essay by Schröter and Sudmann, also published here.

7 For media studies, see for example Engemann/Sudmann 2018; Ernst et al. 2019; Mackenzie 2017; Mann/Matzner 2019; Pasquinelli 2017, 2023; Sudmann 2023. For the

Current investigations into AI research stem from various disciplines involved with the reflection of the sciences. The philosophy of science has deployed its specific expertise for problems of AI (about cognition, consciousness, etc.) and proliferates in the field of AI ethics in particular, a field in which, among other things, various critical perspectives on AI and its research are normatively negotiated (algorithmic biases, surveillance, opacity of technology, etc.; for an overview, see Coeckelbergh 2020; Dimock 2020; Mann/Matzner 2019). Moreover, AI clearly resonates with and functions as a catalyst for the research perspectives of the digital humanities (Jannidis 2013; Manovich 2017; Flückiger 2021).

To further enhance these discussions for the field of AI-based methods in the sciences, a more thorough investigation of scientific practices and infrastructures seems in order (Star 1999; Schabacher 2022). To keep track of current developments an integrated dialogue with computer science is of the essence. In addition to that, it seems highly desirable to observe, document and reflect the current shifts in scientific practices through AI-based methods. To capture these developments up close, a media ethnography of selected AI research projects is the most viable option and will be conducted as the research project unfolds (Dracklé 2014; Dippel 2017; Schüttpelz/Gießmann 2015; Bareither 2019). The integrated approach to scientific practices will further draw on the strengths of media archaeology to situate technically mediated knowledge production in larger frameworks. To this end, we emphasize the technological aspect as well as the social embeddedness of the emerging technology (Dotzler 2006; Schröter 2020; Ernst/Schröter 2020). Historical depth is provided for these findings on scientific practices by recent results from the history of data use in various disciplines (Aronova/von Oertzen/Sepkoski 2017; Schlicht/Ledebur/Echterhölter 2021). In this newly developing field within the history of science, separate instances in data journeys are consulted (Leonelli/Tempini 2020), the emergence of specific algorithms are traced (Evans/Johns 2023) or models are investigated in and of themselves. One of the best researched cases may be weather models, which took a stunning trajectory from decentralized weather observers to dynamic climate models and eventually, their integration into the vast machines of computer simulations (Coen 2018; Edwards 2000; Edwards 2010; Gramelsberger 2010).

history of science, see Seising 2021; Cave/Dihal/Dillon 2020; Evans/Johns 2023. For computer science, see Vaswani et al. 2017; Devlin et al. 2019; Brown et al. 2020; Rombach et al. 2022; Kirillov et al. 2023.

A new technical option for the sciences and humanities calls for a critical reflection of emerging forms (such as databases, algorithms, frameworks, interfaces, etc.) related to the production of knowledge. An engagement with possible transformations of scientific practices demands a methodological approach which refrains from creating *prima facie* distinctions between internal and external factors shaping these transformations, namely approaches from media ethnography, media archaeology, or the history of quantification. Establishing an account of what factors are important for the origin, the implementation (or non-implementation), and not least the retention of AI technology can possibly serve as a gateway for criticizing these very conditions in which the scientific endeavor takes place.

Various contemporary debates on AI technologies revolve around their social and cultural effects. Problems of algorithmic biases, data privacy, or opacity of infrastructures are commonly placed in the normative framework of AI ethics. Critical discussions of the high hopes invested in AI, as well as its present limitations, also continue to play a crucial role in ongoing debates (Broussard 2018). There is still little knowledge, however, about the relationship between the assumed problematic aspects of AI and the ways in which AI affects research practices, methodologies, and outcomes across different sciences. Adequate assessment of the impact of AI on science, including reference to its socio-political implications, is therefore a major research desideratum.

As has been pointed out here, research on the research of AI is confronted with significant challenges. The transdisciplinary view on the problems of AI in science requires distinctive expertise in very heterogeneous fields. However, there is no such thing as universal competence. Therefore, the research group hosting these discussions is all the more dependent on the dialogue and support of scholars from different disciplines and has benefited considerably from their civic engagement across the disciplines. The main focus of this publication is to explore different ways of thinking about the uses of AI in a broad set of scientific fields. At the same time, and in relation to selected disciplines, we want to exemplarily demonstrate the application of AI in specific academic contexts.

List of contributions

In their joint paper, the **members of the project “How is Artificial Intelligence Changing Science?”** discuss nine preliminary theses regarding the possible ef-

fects of the use of different AI technologies in the sciences. I) It is questioned if the widespread rhetoric of an “AI revolution” is helpful to describe the shifts that occur with the introduction of AI technologies in the sciences. II) It is emphasized that AI technologies can only be understood by understanding their embeddedness in infrastructures and social contexts. III) It is stated that AI systems can process fuzziness and uncertainty in a new way. In IV) the conflict between the big tech industry and academia in the development of AI is being highlighted. Thesis V) elaborates on how the fast introduction of AI technologies causes an expert crisis. In VI) it is discussed that many disciplines split into a computational and a non-computational branch. VII) points to the connection of AI technologies with data extraction and data colonialism. In VIII) the thesis is formulated that the introduction of AI will alter the labor landscape profoundly. IX) asks how the self-improvement and the self-evaluation of AI have to be conceptualized.

Mathematics struck gold when employing infinitesimal quantities to solve practical problems towards the end of the 17th century. In a further decolonial reflection on the inherent problems of AI and pattern recognition and discrimination, **Clemens Apprich** investigates this calculus in historical and present debates about incalculability. The calculus, which still performs reliable approximations within the schemes of artificial neural networks, should not be tamed into absolute congruence. On the contrary, it might be the imperfections and approximations which may help us to cultivate procedural and plural approaches. In this sense, the immanence residing at the end of all approximations (which would be tantamount with mastering the visual and compositional realms of quality by new AIs) does not appear fully desirable. Apprich acknowledges the insurmountable incongruity of the mathematical setup of AI and suggests strategic uses, such as Ramon Amaro’s for possible “black totalities”.

Matteo Pasquinelli’s approach to AI is informed by the joint traditions of materialistic epistemology and media theory. He argues strictly against “folk AI”, a perception of this new technology which all too readily accepts a new and contextless entity and its miraculous abilities. Instead, a much longer history of mediated thought and neoliberal entanglement of AI is in order. Without a shadow of a doubt automated and mechanical ways of reasoning have been part and parcel of the scientific endeavor long before artificial neural networks. The paper revisits Rosenblatt’s 1957 strategy to facilitate pattern recognition via the modeling of the labor of perception and supervision. It integrates this historical analysis of AI with Peter Damerow’s theory of mental representation,

the dialectics of tools and knowing, as well as neo-Gramscian approaches towards formalization, and the Hessen-Grossmann thesis of the labor dependency of all science. AI's advanced algorithms are not unique. They are the latest result in a long history of confluences and attempts at "epistemic scaffolding".

Markus Ramsauer offers a genealogy of the development of Early Warning Systems and the potential enhancement in the detection of danger via the use of AI. Taking the trope of birds as sentinels for future catastrophic developments as leitmotif, it is argued that the discovery of latent danger often depends on the use of non-human sensors or kinds of intelligence; be they animalistic or machinistic. This offers a lens for thinking about the concept of 'artificial' and 'non-artificial' intelligence beyond the question if machines can pass as human.

Jean Gabriel-Ganascia's text discusses AI not only in terms of a tool for scientific practice but as a science itself. As such, the author claims, it evades classifications as 'theoretical science', 'science of nature' or 'science of culture'. The reason for this special status can be explained by the history of AI development of which the author provides a brief outline. As a second strand, the article explores the possibilities of 'epistemological ruptures' through the use of AI in the humanities as well as in the 'hard sciences'. Whereas for the former, these tools can assist in assessing individual cases, it contributes to an 'automation of induction' for the latter.

Gabriele Schabacher discusses in her essay the centrality of the notion of "pattern" for subsymbolic artificial intelligence. She asks what the power of patterns in contexts of cognition or application is, by distinguishing two ways of conceptualizing patterns, namely *template* and *correlation*. The reconstruction shows how these two forms are peculiarly blended in the horizon of AI technologies. The first example is the application domain of security research and how the blending of template and correlation works there. The focus will be on German pilot projects in Berlin and Mannheim that test the use of intelligent video analysis. Finally, Schabacher comments on the statistical creativity of AI image generators such as DALL-E, highlights four overarching aspects associated with the work of patterns of AI technologies, and describes their effects on scientific understanding, but also on culture and society in general.

The revolutionary potentials of AI in healthcare are covered in detail in the overview by **Urvi Sonawane** and **Matthieu Komorowski**. The usage of AI-based technologies is currently quite limited, the authors discover, despite its enormous potential. Responsible bottlenecks like technical, ethical, legal, and hu-

man aspects are examined and the need for a multidisciplinary approach involving regulatory bodies, clinicians, government, and patient committees is argued for.

In her position paper, **Isabelle Bloch** argues that a hybrid point of view of designing AI, considering both knowledge data representation and reasoning, offers opportunities towards explainability. This idea is illustrated on the example of medical image understanding, formulated as a spatial reasoning problem.

In his contribution, **Giacomo Landeschi** shows how computer-based applications had a profound impact on the discipline of archaeology and how different methods and techniques, such as satellite remote sensing, geophysical prospections, and more recently, airborne laser scanning (LiDAR), have been employed for surveying purposes. Nowadays, artificial intelligence has also started to play an important role in the analysis of archaeological contexts. In the case of Sweden, approximately 70 per cent of its land comprises forests where a substantial number of archaeological sites remain hidden beneath the vegetation, undiscovered and unmapped. Landeschi explains how a team of scientists from Lund University recently undertook a project to showcase the potential of utilizing deep learning-based analysis and convolutional neural networks for automatically identifying a specific category of archaeological features called ‘clearance cairns’ in LiDAR-derived raster imagery.

Sabine Wirth’s paper sheds light on the ways how the concept of the interface matters for a critical understanding of AI technologies in use. From a media and culture studies perspective she discusses how research on machine learning techniques can profit from a critical perspective on interfaces. Drawing on the emerging field of critical interface studies, Wirth describes two examples of popular apps that rely on machine learning, and she outlines potential lines of inquiry and critical questions that address the central role of interfaces as mediators of AI within the field of popular media culture. Ultimately, this allows her to ask how critical interface studies can inform research on AI in science by providing an additional analytical layer.

In their contribution, **Andreas Sudmann** and **Jens Schröter** shed light on the role of media related to how AI is used in and potentially transforms different fields of academic research. Furthermore, they draw attention to some important problems of applied AI which thus require critical reflection, especially from a media studies perspective.

Johannes Breuer poses the question of how AI is changing scientific practice in the realm of the social sciences. His contribution “Putting AI into social

science” highlights the importance of tools for different stages of the scientific endeavor. The author discusses a variety of AI-driven research tools which are suitable for the social sciences, emphasizing their potentially transformative potential as well as ethical challenges that go hand in hand with this transformation. The chapter concludes with an invocation to focus on partnerships *with* AI, rather than on replacement *by* AI.

The paper by **Evangelos Pournaras** reviews the specific epistemological challenges and also the ethical and integrity risks related to generative AI and LLMs. In particular, Pournaras discusses emerging practices for research ethics, proposing ten recommendations that shape a response for a more responsible research conduct in the era of AI.

In his paper, **Fabian Retkowski** aims at concisely indicating the current state of the art in abstractive text summarization. The current paradigm shifts towards pre-trained encoder-decoder models and large autoregressive language models are outlined and the challenges of evaluating summarization systems and the potential of instruction-tuned models for zero-shot summarization are discussed in further detail. Additionally, the work gives a brief overview of how summarization systems are currently being integrated into commercial applications.

Sabina Leonelli maintains in her chapter that despite ever larger amounts of data and proclaimed bias-reducing algorithms, the employment of AI tools in scientific research is still heavily affected by the quality of the training data. The hardly traceable origin of data, combined with their often diverse nature and purpose, leads to what the author calls “in-practice opacity”. Instead of focusing on quantitative modes of reproducibility as a panacea for making science transparent, the author calls for extended attention to questions about the quality and the funding of research data.

The use of data has been a key element of statistics, yet the dimensions of current data usages constitute a new situation. **Gérard Biau** is in a unique position to answer a set of questions about the changes affected by AI in this particular field of mathematics: He works at the Probability, Statistics, and Modeling Laboratory (LPSM), serves as director of the Sorbonne Center for Artificial Intelligence (SCAI), and was president of the *Société française de statistique*. Biau states that the impact of AI on mathematics is decisive. Some statistical tools, which have been stable for decades, are currently being revised. AIs start to make suggestions regarding results, or are instrumental in verifying the most advanced new proofs.

In his interview with **Sybille Krämer**, Jens Schröter poses nine questions which closely follow Krämer's writings over the decades. Her work has been, from the very beginning, revolving around questions that are of special relevance to understanding subsymbolic AI today. This starts with the question on the culturally shaped exteriority of the human mind, the relation of AI to the fundamental role of the analog and the digital, or the connection of AI to the field of digital humanities. Further fundamental points are discussed like the question if AI can be understood as a "cultural technique", especially when we observe the increasing role of computers in science. Finally, Krämer addresses questions of explainability and critique.

List of references

- Araújo, Duarte/Couceiro, Micael/Seifert, Ludovic/Sarmento, Hugo/Davids, Keith (eds.) (2021): *Artificial Intelligence in Sport Performance Analysis*, London: Routledge.
- Aronova, Elena/von Oertzen, Christine/Sepkoski, David (2017): "Introduction: Historicizing Big Data." In: *Osiris* 32/1, pp. 1–17.
- Athey, Susan (2018): "The Impact of Machine Learning on Economics." In: Ajay Agrawal/Joshua Gans/Avi Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press, pp. 507–547.
- Baurmann, Michael/Mans, Dieter (1984): "Künstliche Intelligenz in den Sozialwissenschaften." In: *Analyse & Kritik* 6/2, pp. 103–159.
- Bostrom, Nick (2014): *Superintelligence: Paths, Dangers, Strategies*, Oxford: University of Oxford Press.
- Bourilkov, Dimitri (2019): "Machine and Deep Learning Applications in Particle Physics". In: *International Journal of Modern Physics A*, 34/35, 1930019.
- Broussard, M. (2018): *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: The MIT Press.
- Brown, Tom/Mann, Benjamin/Ryder, Nick/Subbiah, Melanie/Kaplan, Jared D./Dhariwal, Prafulla/Neelakantan, Arvind/et al. (2020): "Language Models are Few-Shot Learners." In: *Advances in Neural Information Processing Systems* 33, pp. 1877–1901.
- Carley, Kathleen M. (1996): "Artificial Intelligence within Sociology." In: *Sociological Methods & Research* 25/1, pp. 3–30.

- Cave, Stephen/Dihal, Kanta/Dillon, Sarah (eds.) (2020): *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford: Oxford University Press.
- Coeckelbergh, Mark (2020): *AI Ethics*. Cambridge, MA: The MIT Press.
- Coen, Deborah (2018): *Climate in Motion: Science, Empire, and the Problem of Scale*, Chicago: The University of Chicago Press.
- Cornelio, Cristina/Dash, Sanjeeb/Vernon, Austel/Josephson, Tyler/Goncalves, Joao/Clarkson, Kenneth/Megiddo, Nimrod/et al. (2021): *AI Descartes: Combining Data and Theory for Derivable Scientific Discovery*, arXiv Preprint (<https://arxiv.org/pdf/2109.01634.pdf>).
- Devlin, Jacob/Chang, Ming-Wei/Lee, Kenton/Toutanova, Kristina (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MS, USA, pp. 4171–4186.
- Dick, Stephanie (2011): "AfterMath: The Work of Proof in the Age of Human-Machine Collaboration." In: *Isis* 102/3, pp. 494–505.
- Dimock, Wai Chee (2020): "AI and the Humanities." In: *PMLA: Publications of the Modern Language Association* 135/3, pp. 449–454.
- Dippel, Anne (2017): "Das Big Data Game: Zur Spielerischen Konstitution kollaborativer Wissensproduktion in der Hochenergiephysik am CERN." In: *NTM – Zeitschrift Für Geschichte der Wissenschaften, Technik und Medizin* 25/4, pp. 485–517.
- Domingos, Pedro (2015): *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, New York: Basic Books.
- Dotzler, Bernhard J. (2006): *Diskurs und Medium. Zur Archäologie der Computerkultur*, München: Wilhelm Fink Verlag.
- Dracklé, Dorle (2014): "Medienethnologie/Medienethnographie." In: Jens Schröter (ed.), *Handbuch Medienwissenschaft*, Stuttgart: Metzler, pp. 393–404.
- Dreyfus, Hubert/Dreyfus, Stuart (1988): "Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint." In: Stephen R. Graubard (ed.), *The Artificial Intelligence Debate. False Starts, Real Foundations*, Cambridge, MA, and London: The MIT Press, pp. 15–43.
- Echterhölter, Anna/Schröter, Jens/Sudmann, Andreas (2021): *How Is Artificial Intelligence Changing Science? Research in the Era of Learning Algorithms*, MediarXiv Preprint (<https://doi.org/10.33767/osf.io/28pnx>).

- Edwards, Paul (2000): "The World in a Machine. Origins and Impacts of Early Computerized Global Systems Models." In: Agatha C. Hughes/Thomas Parke Hughes (eds.), *Systems, Experts, and Computers*, Cambridge, MA: The MIT Press, pp. 221–253.
- Edwards, Paul (2010): *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge, MA: The MIT Press.
- Engemann, Christoph/Sudmann, Andreas (eds.). (2018): *Machine Learning – Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld: transcript.
- Ernst, Christoph/Schröter, Jens (2020): *Zukünftige Medien: Eine Einführung*, Wiesbaden: Springer.
- Evans, James/Johns, Adrian (eds.) (2023): *Beyond Craft and Code: Human and Algorithmic Cultures, Past and Present*, *Osiris* 38/1.
- Fawzi, Alhusein/Balog, Matej/Huang, Aja/Hubert, Thomas/Romera-Paredes, Bernardino/Barekatin, Mohammadamin/Novikov, Alexander/et al. (2022): "Discovering Faster Matrix Multiplication Algorithms with Reinforcement Learning." In: *Nature* 610/7930, pp. 47–53.
- Fecher, Benedikt/Hebing, Marcel/Laufer, Melissa/Pohle, Jörg/Sofsky, Fabian (2023): *Friend or Foe? Exploring the Implications of Large Language Models on the Science System*, arXiv Preprint (<https://doi.org/10.48550/arXiv.2306.09928>).
- Feigenbaum, Edward A. (1968): *Artificial Intelligence: Themes in the Second Decade*, Memo Number 67, AD680487, Stanford: Stanford University.
- Feigenbaum, Edward A./Feldman, Julian (eds.) (1963): *Computers and Thought*, New York: McGraw-Hill.
- Flückiger, Barbara (2017): "A Digital Humanities Approach to Film Colors." In: *The Moving Image: The Journal of the Association of Moving Image Archivists* 17/2, pp. 71–94.
- Flückiger, Barbara (2020): "Digitale Werkzeuge zur ästhetischen Analyse von Filmfarben." In: *Montage AV* 29/1, pp. 157–172.
- Flückiger, Barbara (2021): "Color and Expressive Movement. A Computer-assisted Analysis." In: *Mediaesthetics – Zeitschrift für Poetologien Audiovisueller Bilder* 4 (<https://doi.org/10.17169/mae.2021.86>).
- Gethmann, Carl Friedrich/Buxmann, Peter/Distelrath, Julia/Humm, Bernhard G./Lingner, Stephan/Nitsch, Verena/Schmidt, Jan C./Spiecker genannt Döhmann, Indra (2022): *Künstliche Intelligenz in der Forschung. Neue Möglichkeiten und Herausforderungen für die Wissenschaft*, Berlin and Heidelberg: Springer.

- Gramelsberger, Gabriele (2010): *Computerexperimente. Zum Wandel der Wissenschaft im Zeitalter des Computers*. Bielefeld: transcript.
- Groß, Richard/Jordan, Rita (2023): *KI-Realitäten. Modelle, Praktiken und Topologien maschinellen Lernens*, Bielefeld: transcript.
- Hassabis, Demis (2022): “Dr Demis Hassabis: Using AI to Accelerate Scientific Discovery.” University of Oxford, uploaded to YouTube on August 3, 2022 (<https://youtu.be/AU6HuhrC65k>).
- Hochreiter, Sepp/Schmidhuber, Jürgen (1997): “Long Short-Term Memory.” In: *Neural Computation* 9/8, pp. 1735–1780.
- How is Artificial Intelligence Changing Science? (2023): “AI Projects in Europe. Selected Disciplines.”, last updated February 14, 2023 (<https://howisaichangingscience.eu/ai-projects-in-europe/>).
- Hunt, Earl (1968): “Computer Simulation: Artificial Intelligence Studies and Their Relevance to Psychology.” In: *Annual Review of Psychology* 19/1, pp. 135–168.
- Jannidis, Fotis (2013): “Computerphilologie.” In: Thomas Anz (ed.), *Handbuch Literaturwissenschaft, Band 2, Methoden und Theorien*, Stuttgart and Weimar: Metzler, pp. 27–40.
- Jenicek, Tomas/Chum, Ondrej (2019): *Linking Art through Human Poses*, arXiv Preprint (<http://arxiv.org/abs/1907.03537>).
- Jumper, John/Evans, Richard/Pritzel, Alexander/Green, Tim/Figurnov, Michael/Jumper Ronneberger, Olaf/Tunyasuvunakool, Kathryn/et al. (2021): “Highly Accurate Protein Structure Prediction with AlphaFold.” In: *Nature* 596/7873, pp. 583–589.
- Kautz, Henry (2022): “The Third AI Summer: AAAI Robert S. Engelmore Memorial Lecture.” In: *AI Magazine* 43/1, pp. 105–125.
- Kirillov, Alexander/Mintun, Eric/Ravi, Nikhila/Mao, Hanzhi/Rolland, Chloé/Gustafson, Laura/Xiao, Tete/et al. (2023): *Segment Anything*, arXiv Preprint (<https://arxiv.org/abs/2304.02643>).
- Krämer, Sybille (ed.) (1994): *Geist, Gehirn, künstliche Intelligenz: Zeitgenössische Modelle des Denkens*. Ringvorlesung an der Freien Universität Berlin, Berlin and New York: de Gruyter.
- Krizhevsky, Alex/Sutskever, Ilya/Hinton, Geoffrey E. (2012): “ImageNet Classification with Deep Convolutional Neural Networks.” In: *Advances in Neural Information Processing Systems* 25, pp. 1097–1105.
- LeCun, Yann/Bottou, Leon/Bengio, Yoshua/Haner, Patrick (1998): “Gradient-Based Learning Applied to Document Recognition”. In: *Proceedings of the IEEE* 86, pp. 2278–2324.

- Lewkowycz, Aitor/Andreassen, Anders/Dohan, David/Dyer, Ethan/Michalewski, Henryk/Ramasesh, Vinay V./Slone, Ambrose/et al. (2022): Solving Quantitative Reasoning Problems with Language Models, arXiv Preprint (<https://arxiv.org/pdf/2206.14858.pdf>).
- Ligo, Alexandre K./Rand, Krista/Basset, Jason/Galaitsi, S.E./Trump, Benjamin D./Jayabalasingham, Bamini/Collins, Thomas/Linkov, Igor (2021): "Comparing the Emergence of Technical and Social Sciences Research in Artificial Intelligence." In: *Frontiers in Computer Science* 3, pp. 1–13.
- Luong, Minh-Thang/Manning, Cristopher (2015): "Stanford Neural Machine Translation Systems for Spoken Language Domains". In: *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign, Da Nang, Vietnam*, pp. 76–79.
- Mackenzie, Adrian (2017): *Machine Learners. Archaeology of a Data Practice*, Cambridge, MA: The MIT Press.
- Manhart, Klaus (1995): *KI-Modelle in den Sozialwissenschaften. Logische Struktur und wissensbasierte Systeme von Balancetheorien*, München: Oldenbourg Wissenschaftsverlag.
- Mankowitz, Daniel J./Michi, Andrea (2023): "AlphaDev Discovers Faster Sorting Algorithms." June 7, 2023 (<https://www.deepmind.com/blog/alphadev-discovers-faster-sorting-algorithms>).
- Mankowitz, Daniel J./Michi, Andrea/Zhernov, Anton/Gelmi, Marco/Selvi, Marco/Paduraru, Cosmin/Leurent, Edouard/et al. (2023): "Faster Sorting Algorithms Discovered Using Deep Reinforcement Learning." In: *Nature* 618/7964, pp. 257–263.
- Mann, Monique/Matzner, Tobias (2019): "Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-discrimination in Responding to Emergent Discrimination." In: *Big Data & Society* 6/2, pp. 1–11.
- Manovich, Lev (2017): "Cultural Analytics, Social Computing and Digital Humanities." In: Mirko Tobias Schäfer/Karin van Es (eds.), *The Datafied Society. Studying Culture Through Data*, Amsterdam: Amsterdam University Press, pp. 55–68.
- McCorduck, Pamela (1979): *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*, San Francisco: W.H. Freeman and Company.
- Nguyen, Thai-Son/Stueker, Sebastian/Waibel, Alex (2020): Super-Human Performance in Online Low-latency Recognition of Conversational Speech, arXiv Preprint (<https://doi.org/10.48550/arXiv.2010.03449>).

- Okerlund, Johanna/Klasky, Evan/Middha, Aditya/Kim, Sujin/Rosenfeld, Hannah/Kleinman, Molly/Parthasarathy, Shobita (2022): “What’s in the Chat-terbox? Large Language Models, Why They Matter, and What We Should Do About Them, Research Report Ford School of Public Policy.”, April 2022 (<https://stpp.fordschool.umich.edu/research/research-report/what-s-in-the-chatterbox>).
- Pasquinelli, Matteo (2017): “Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference.” In: *Glass Bead 1*, (<https://www.glass-bead.org/article/machines-that-morph-logic/>).
- Pasquinelli, Matteo (2023): “Labour, Energy, and Information as Historical Configurations: Notes for a Political Metrology of the Anthropocene.” In: *Journal for the Interdisciplinary History of Ideas* 22/11 (<https://doi.org/10.13135/2280-8574/7340>).
- Radovic, Alexander/Williams, Mike/Rousseau, David/Kagan, Michael/Bonacorsi, Daniele/Himmel, Alexander/Aurisano, Adam/et al. (2018): “Machine Learning at the Energy and Intensity Frontiers of Particle Physics.” In: *Nature* 560/7716, pp. 41–48.
- Rombach, Robin/Blattmann, Andreas/Lorenz, Dominik Lorenz/Esser, Patrick Esser/Ommer, Björn (2022): “High-Resolution Image Synthesis with Latent Diffusion Models.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022)*, New Orleans, LA, USA, pp. 10674–10685.
- Schabacher, Gabriele (2022): *Infrastruktur-Arbeit. Kulturtechniken und Zeitlichkeit der Erhaltung*, Berlin: Kadmos.
- Schmidhuber, Jürgen (2022): *Annotated History of Modern AI and Deep Neural Networks*, arXiv Preprint (<https://arxiv.org/ftp/arxiv/papers/2212/2212.11279.pdf>).
- Schröter, Jens (2020): “Digitale Medien und Methoden. Jens Schröter zur Medienarchäologie der digitalen Medien.”, July 10, 2020 (<https://zfmiedienwissenschaft.de/online/open-media-studies-blog/digitale-medien-und-methoden-6-schroeter>).
- Schüttpelz, Erhard/Gießmann, Sebastian (2015): “Medien der Kooperation. Überlegungen zum Forschungsstand.” In: *Navigationen – Zeitschrift für Medien- und Kulturwissenschaften* 15/1, pp. 7–55.
- Seising, Rudolf (2021): *Es denkt nicht! Die vergessenen Geschichten der KI*, Frankfurt a.M.: Büchergilde Gutenberg.

- Simanowski, Roberto (2023): "Narrative der Weltbeglückung. Die neue Sprach-KI und die Mathematisierung der Ethik." In: *Blätter für deutsche und internationale Politik* 6, pp. 64–73.
- Star, Susan L. (1999): "The Ethnography of Infrastructure." In: *American Behavioral Scientist* 43/3, pp. 377–391.
- Suchman, Lucy (2006): *Human-Machine Reconfigurations: Plans and Situated Actions*, Cambridge: Cambridge University Press.
- Sudmann, Andreas (2018a): "Zur Einführung. Medien, Infrastrukturen und Technologien des maschinellen Lernens." In: Christoph Engemann/Andreas Sudmann (eds.), *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld: transcript, pp. 9–23.
- Sudmann, Andreas (2018b): "Szenarien des Postdigitalen. Deep Learning als MedienRevolution." In: Christoph Engemann/Andreas Sudmann (eds.), *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld: transcript, pp. 55–73.
- Sudmann, Andreas (2023): "Zum Einsatz von KI als Methode und Werkzeug in der Medienwissenschaft." In: Sven Stollfuß/Laura Niebling/Felix Raczkowski (eds.), *Handbuch Digitale Medien und Methoden*, Wiesbaden: Springer (https://doi.org/10.1007/978-3-658-36629-2_16-2).
- Varadi, Mihaly/Velankar, Sameer (2022): "The Impact of AlphaFold Protein Structure Database on the Fields of Life Sciences." In: *Proteomics*, e2200128 (<https://doi.org/10.1002/pmic.202200128>).
- Vaswani, Ashish/Shazeer, Noam/Parmar, Niki/Uszkoreit, Jakob/Jones, Llion/Gomez, Aidan N./Kaiser, Łukasz/Polosukhin, Illia (2017): "Attention Is All You Need." In: *Advances in Neural Information Processing Systems* 30, pp. 5998–6008.
- Waibel, Alex (1987): "Phoneme Recognition Using Time-Delay Neural Networks." In: *Proceedings of the Meeting of the Institute of Electrical, SP87-100, Information and Communication Engineers (IEICE), Tokyo, Japan*.
- Waibel, Alex/Sawai, Hidefumi/Shikano, Kiyohiro (1989): "Modularity and Scaling in Large Phonemic Networks." In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37/12, pp. 1888–1898.