Editorial

Introduction to the Special Issue "The Datafication of Communication – New Methodological Approaches and Challenges"

Julia Niemann-Lenz / Tim Schatto-Eckrodt / Emese Domahidi / Merja Mahrt*

Over the past decade, methodological developments in communication research have accelerated considerably. New approaches and tools have been introduced under different labels, including "big data," computational communication science, and, more recently, generative artificial intelligence. Each of these terms signals shifts in how communication processes are empirically accessed and analyzed, often accompanied by promises of increased analytical power.

These methodological developments are rooted in a fundamental transformation of communicative environments and research conditions. The critical foundation of these transformations lies in processes of *datafication*. Datafication describes a broader sociotechnical process through which social action, communication, and interaction are translated into quantifiable data that can be stored, processed, and repurposed (Mayer-Schönberger & Cukier, 2013). Importantly, datafication is not a neutral process, but is embedded in asymmetrical power relations: Decisions about which forms and aspects of communication are captured, how they are classified, and who gains access to the resulting data are shaped by technical infrastructures as well as organizational and economic interests. These infrastructures and interests exert a substantial influence on communication processes themselves. For instance, a growing body of research shows that the affordances of algorithmic infrastructures in social media affect different forms of communication, including interpersonal, journalistic, and political communication (Hase et al., 2023; Li et al., 2024; Teocharis et al., 2023; Welbers & Opgenhaffen, 2019). Moreover, central concepts of media and communication research—such as selection, gatekeeping, visibility, and the formation of public discourse—are increasingly intertwined with datafied and algorithmically mediated processes (Couldry & Hepp, 2017; van Dijck et al., 2018).

The datafication of communication not only alters the object of media and communication research itself but also affects which aspects of communication are considered empirically observable and, consequently, researchable. As communicative practices are translated into digital traces, forms of communication that are easily captured, quantified, and processed tend to gain analytical prominence, while others remain less visible or difficult to access. New forms of data access, automated data collection, and computational methods of analysis thus increasingly influence how communication phenomena are opera-

* Dr. Julia Niemann-Lenz, German Centre for Higher Education Research and Science Studies (DZHW), Lange Laube 12, 30159 Hannover, Germany, lenz@dzhw.eu, https://orcid.org/0000-0003-1991-8535;
Dr. Tim Schatto-Eckrodt, University of Hamburg, Journalism and Mass Communication, Von-Melle-Park 5, 20146 Hamburg, Germany, tim.schatto-eckrodt@uni-hamburg.de, https://orcid.org/0000-0003-1658-4373;
Prof. Dr. Emese Domahidi, Technische Universität Ilmenau, Institut für Medien und Kommunikationswissenschaft, Fachgebiet Computational Communication Science, Ehrenbergstraße 29, 98693 Ilmenau, emese.domahidi@tu-ilmenau.de, https://orcid.org/0000-0003-3530-797X;
PD Dr. Merja Mahrt, Weizenbaum Institute e. V., Hardenbergstraße 32, 10623 Berlin, Germany, merja.mahrt@weizenbaum-institut.de, https://orcid.org/0009-0009-4246-734X.

DOI: 10.5771/1615-634X-2026-1-3

https://doi.org/10.5771/1615-634X-2026-1-3 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

tionalized, measured, and interpreted. Under the label of computational communication science, digital trace data have moved to the center of empirical research in the field (Domahidi et al., 2019). Approaches such as automated content analysis, network analysis, machine learning, and, more recently, large language models (LLMs) promise new insights into communicative dynamics in a datafied society. These approaches build on large-scale datafied communication as their empirical foundation and further automate processes of classification, modeling, and interpretation. In doing so, they extend established methodological repertoires while simultaneously challenging existing conventions of data collection, analysis, and validation.

The increasing prominence of data-driven research has also prompted critical debate. Scholars have questioned the assumption that social reality can be adequately captured through large-scale data alone, pointing to the socially constructed nature of data and the socio-technical conditions under which they are produced (boyd & Crawford, 2012). From this perspective, digital data are not neutral representations of human behavior but are shaped by platform architectures, algorithms, institutional practices, and power relations. Similarly, the methodological convenience of accessing social media data has led to a proliferation of studies focusing on easy-to-research spaces. This may have contributed to an overestimation of these platforms' role in misinformation research and to the neglect of other areas, such as the role of established media in disseminating false information (Altay et al., 2023). Critical accounts have further emphasized the role of large platforms in extracting and controlling data, conceptualizing these practices as forms of "data colonialism" that parallel historical patterns of resource extraction (Couldry & Mejias, 2019). Relatedly, the ideological foundations of datafication have been described as "dataism," referring to the belief that human actions can be objectively quantified and understood through data (van Dijck, 2014). In addition to these epistemological critiques, the use of new data sources and analytical techniques raises legal and ethical concerns, particularly with regard to privacy, consent, transparency, and accountability (Spirling, 2023).

In light of these critiques and acknowledging the dual role of datafication as both a methodological precondition and a constitutive context of communication, the Digital Communication and Methods of Media and Communication Research sections of the German Communication Association (DGPuK) jointly organized a conference in Hamburg in 2024 to address these developments. The conference brought together 70 scholars working on different aspects of datafied communication research in order to reflect on methodological innovations, empirical practices, and critical perspectives across a range of subfields. The presentations and discussions quickly made clear that the topic engages colleagues from both sections, as they are centrally affected by datafication. Large language models were a focal point of many panels, as scholars employed different approaches to explore new methodological opportunities (as well as potential pitfalls) and to carve out appropriate roles for these approaches. This applies both to digital communication research and to the development of methods for data collection, organization, and analysis. It also became apparent that the field is currently highly productive and fast-moving, underscoring the need for more long-term investigations to complement the often tightly focused and practically oriented studies on developing or applying new datafied approaches.

Following the fruitful discussions at the conference, this Special Issue addresses some of the broader methodological implications of datafication for communication research. At the same time, it brings together contributions that focus more specifically on different stages of the research process, examining how digital data are generated, accessed, analyzed, and critically reflected upon within the discipline. Rather than advocating for a singular methodological position, the Special Issue aims to highlight the diversity of approaches that

https://doi.org/10.5771/1615-634X-2026-1-3 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

currently shape datafied communication research. The contributions engage with different stages of the research process and illustrate how methodological innovation and critical reflection can be productively combined. While this results in a broad thematic scope, the Special Issue deliberately embraces this breadth in order to capture the heterogeneity of datafied communication research and the multiple ways in which datafication affects empirical practice.

The Special Issue opens with a contribution in German language by Denise Sommer, Jörg Hagenah, and Anna-Sophie Brucks who trace the long-term effects of datafication on empirical communication research. They conducted a content analysis of the three major German-language journals from 2003 to 2023 to investigate changes in method and data choices over time. Computational approaches have only recently seen an—albeit steep—increase in popularity, while the growing availability of larger datasets for secondary analysis does not appear to have resulted in more studies drawing on these resources. The authors conclude that compared to neighboring disciplines such as sociology, communication research, at least in these three journals, does not yet appear to be strongly data-driven, and that data infrastructures should be strengthened to enable scholars in German-speaking countries to keep pace with international developments.

Next are Patrick Zerrer, Paul Pressmann, Cornelius Puschmann, and Philipp Krieter, who present a holistic framework for researching mobile media use. In their article, Zerrer et al. highlight the relevance of mobile device use for media use research more broadly and reflect on the current state of the field. Accessing mobile media use involves various hurdles for researchers, ranging from non-trivial questions of technical access to participants' devices, the ethical considerations concerning user privacy to the methodological challenge of operationalizing media exposure beyond the analysis of timestamped browser histories alone.

Through two case studies, Zerrer et al. demonstrate the possibilities and limits of mobile media use tracking by combining screen recording, a keyword tracker that automatically detects relevant terms appearing on screen, and the matching of observed exposure to media content through scraping and (computational) analysis of, for example, relevant social media accounts. Based on these case studies, the authors derive five conceptual core principles for in-app tracking, concerning 1) user privacy, 2) the management and interpretation of the raw data generated by these methods, 3) potential biases arising from the interconnectedness of the socio-technical systems under study, 4) the need to track not only singular media exposures but also cascades of subsequent exposure in order to understand how information travels across platforms, and 5) the assurance of replicability through precise and comprehensible documentation of the reported approaches.

In the third article, Philipp Kessling and Felix Victor Münch address a central methodological challenge in the study of the digital networked public sphere: the systematic and transparent collection of data across multiple platforms. Rather than focusing on specific technologies or access points, the contribution proposes a conceptual framework that abstracts from concrete data sources and instead structures the data collection process itself. By conceptualizing data collection as a sequence of modular components, the article provides a common language for describing, comparing, and documenting exploratory data collection in cross-platform research. This modular perspective is particularly valuable in research contexts characterized by heterogeneous data access conditions and rapidly changing technical environments. It allows researchers to disentangle conceptual decisions from technical implementations and thereby enhances transparency, reproducibility, and interoperability. Beyond its immediate applicability, the framework contributes to ongoing

https://doi.org/10.5771/1615-634X-2026-1-3 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

methodological debates by offering a way to systematically reflect on how networked communication is sampled and empirically reconstructed under conditions of datafication.

Nadezhda Ozornina and Mario Haim examine the influence of target language selection on topic modeling in multilingual settings—an issue that has so far received little attention in computational communication research. The study compares two methodological approaches: consolidating texts into one of the original document languages versus translating all texts into an intermediary language that is not part of the original corpus. These approaches are tested on a parallel corpus of United Nations documents in Russian and German and are subsequently replicated using a second bilingual dataset. The findings indicate that translation into an intermediary language—English in this case—offers certain advantages, including a more symmetrical topic distribution and a higher degree of overlap among top-ranked terms. At the same time, this strategy substantially reduces the available vocabulary compared to consolidation in an original language, which may have implications for downstream analyses, such as cross-lingual emotion detection. Particularly noteworthy are the detailed discussion of the findings and the formulation of best-practice recommendations, which provide valuable guidance for future research in this area.

Last but not least, Annekatrin Bock explores the benefits of combining the theoretical framework of critical data studies (CDS) with the research agenda of communication studies, while offering insights into the general assumptions of CDS, their relevance for communication research, and the methodological application of these principles. In her article, Bock critically reflects on the use of digital and computational methods by highlighting the implications of methodological frameworks that rely on platform data, which can be understood as made (i.e., not raw or value-free) and invisibly entangled within the socio-technical systems that both produce these data and exercise interpretative sovereignty over them.

By outlining two approaches from the methodological canon of CDS—the *walkthrough approach* and *data journeys*—Bock demonstrates how empirical communication research can benefit from adopting a perspective of critical data literacy. This perspective allows researchers to foreground the dynamic and relational aspects of data, as well as the ways in which data practices are embedded in broader social and technological systems. In an outlook, Bock addresses how researchers might deal with the implications of different data sources and calls for the adoption of open data practices and robust documentation to further transparency and accountability, as well as embracing interdisciplinary methodological (re-)adjustments where necessary.

The collection of articles thus illustrates how deeply and in manifold ways datafication affects communication research. The data, methods, and tools that are and become available steer research trends and may, in the long run, even shift entire paradigms. During the conference on datafication in Hamburg in the fall of 2024, this was often discussed through comparisons between the approaches presented and potential opportunities, similarities, and differences vis-à-vis related generative AI-based approaches. At the time, this line of research was still in its early stages, but the field has since been highly productive, as evidenced by numerous conferences and publications. We assume that many of the issues discussed in the five contributions collected here will require continued investigation, while others, such as the intransparency of especially commercial generative AI applications in particular, are likely to become even more pressing.

https://doi.org/10.5771/1615-634X-2026-1-3 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

References

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media + Society, 9*(1), 205630512211504. https://doi.org/10.1177/20563051221150412

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Couldry, N., & Hepp, A. (2017). *The Mediated Construction of Reality*. Polity.

Couldry, N., & Mejias, U. A. (2019). T*he Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

Dijck, J. van. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance & Society, 12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Computational Communication Science | Outlining the Way Ahead in Computational Communication Science: An Introduction to the IJoC Special Section on "Computational Methods for Communication Science: Toward a Strategic Roadmap". *International Journal of Communication, 13*, 3876–3884. https://ijoc.org/index.php/ijoc/article/view/10533

Hase, V., Boczek, K., & Scharkow, M. (2023). Adapting to Affordances and Audiences? A Cross-Platform, Multi-Modal Analysis of the Platformization of News on Facebook, Instagram, TikTok, and Twitter. *Digital Journalism, 11*(8), 1499–1520. https://doi.org/10.1080/21670811.2022.2128389

Li, M., Suk, J., Zhang, Y., Pevehouse, J. C., Sun, Y., Kwon, H., Lian, R., Wang, R., Dong, X., & Shah, D. V. (2024). Platform Affordances, Discursive Opportunities, and Social Media Activism: A Cross-Platform Analysis of #MeToo on Twitter, Facebook, and Reddit, 2017–2020. *New Media & Society, 28*(1), 119–147. https://doi.org/10.1177/14614448241285562

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (S. 242). Houghton Mifflin Harcourt.

Spirling, A. (2023). Why Open-Source Generative AI Models Are an Ethical Way Forward for Science. *Nature, 616*(7957), 413. https://doi.org/10.1038/d41586-023-01295-4

Theocharis, Y., Boulianne, S., Koc-Michalska, K., & Bimber, B. (2023). Platform Affordances and Political Participation: How Social Media Reshape Political Engagement. *West European Politics, 46*(4), 788–811. https://doi.org/10.1080/01402382.2022.2087410

Welbers, K., & Opgenhaffen, M. (2019). Presenting News on Social Media: Media Logic in the Communication Style of Newspapers on Facebook. *Digital Journalism, 7*(1), 45–62. https://doi.org/10.1080/21670811.2018.1493939

© Julia Niemann-Lenz / Tim Schatto-Eckrodt / Emese Domahidi / Merja Mahrt

https://doi.org/10.5771/1615-634X-2026-1-3 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY