Human Computation (2025) 12:1:43-55

ISSN: 2330-8001, DOI: 10.15346/hc.v12i1.160

Are hybrid human-AI-systems a more humane and responsible version of AI? Empirical inquiries into the conundrums of value-by-design in language technology

GERTRAUD KOCH, University of Hamburg

ABSTRACT

The development of hybrid human–artificial intelligence approaches offers the prospect of creating a more humane and responsible form of AI. However, technological anthropologists question whether these approaches represent a significant shift. The idea that good and bad, humane and inhumane, or responsible and irresponsible versions of AI can be achieved through hybrid human–AI systems oversimplifies the complexity of technological design and implementation. Focusing on 'intelligence performance' is too narrow, distracting from the conundrum of value-oriented design and the crucial dimensions of socio-technical systems. These systems intertwine the various socio-technical layers of AI technology design with specific stakeholders and practices in social reality. Value orientation requires practices at all layers involved in technology design and implementation, and is therefore a continuous technological activity rather than emerging from a particular AI system architecture. Furthermore, the development of a humane and responsible version of AI requires a fundamental realignment of AI technology and research paradigms, challenging the concept of superior intelligence and power for the purpose of providing a rational and reflective evaluation of technology that goes beyond Turing's black boxing.

1. INTRODUCTION: RESPONSIBILITY IN HYBRID HUMAN-AI-SYSTEMS

In current debates on artificial intelligence (AI), human-centred AI has received particular attention (Shneiderman, 2020; Spiekermann & Winkler, 2020). The hybrid intelligence design approach emphasises the contributions of both human and AI to the system, favouring a 'weak definition' of AI that does not compete with or pretend to be more powerful than humans (Jarrahi et al., 2022). The approach proposes augmenting human intellect with collaborative, adaptive, re-

sponsible and accountable AI. Such hybrid human-AI approaches thus promise a superior, more responsible and humane version of AI that is still better than humans (Zeynep et al., 2020). However, this opinion piece contests the idea that hybrid human-AI systems are a superior and better version of AI, arguing that they cannot resolve the conundrums of value-by-design approaches, which consider value in systems engineering methodologies as going beyond the economic value towards "the multi-domain effects on processes, environments, and stakeholders" (Lavi & Reich, 2024, p. 1). Furthermore, the paper asserts that the distinction between weak and strong AI is, in technical terms, arbitrary, and that the roles and outcomes for the human element are often unsatisfactory. To this end, this paper introduces a socio-technical model of technological design.

This opinion piece is written from an anthropology of technology perspective, with a long-standing ethnographic research interest in AI dating back to the 1990s (Koch, 2005), and more recently from within the hybrid human-AI project DWISE (*Digitale Wissenssoziologische Diskursanalyse*) in the digital humanities. The DWISE project was carried out from 2021 to 2024 in collaboration with Chris Bieman's Language Technology Group Hamburg to create the Discourse Analysis Tool Suite (DATS) for multimodal digital discourse analysis. In this project, the two disciplines explored how AI technological innovation can be combined with the innovation of hermeneutic processes in the sociology of knowledge discourse analysis, which is led by grounded theory and involves large amounts of digital materials (Koch et al., 2026). The project's tangible outputs include DATS as a proof of concept (https://github.com/uhh-lt/dats) as a pre-state of a running software and new approaches to working with large data sets in qualitative social science research (Koch et al., 2022). Further outputs include insights into the limited capacity of individual digital humanities projects to address issues of responsibility and humane design in AI, even within a collaborative, cross-disciplinary laboratory structure dedicated to integrating humanities knowledge. Despite all the dedication to value-oriented design, in the end, the project reproduced the structural problems of AI. By building on pre-configured large language models (LLMs), DATS perpetuates the path dependencies of language technology (LT) and thus scales and reproduces the widely discussed problems of bias. The project exploited human capacity by making skilled experts perform tedious work to advance the technology, thereby reproducing established systems of wealth and reputation allocation.

The DWISE project shows that hybrid human-AI systems reproduce AI ethics problems and raise many questions without providing tangible answers regarding humane and responsible AI design. Who is responsible for what, and how should this responsibility be carried out (Constantinescu et al., 2021; Himmelreich & Köhler, 2022; Rismani & Moon, 2023)? Which frameworks and governance approaches facilitate responsibility in AI (Batool et al., 2023; Gianni et al., 2022; Lu et al., 2024; Papagiannidis et al., 2025)? What role do context and culture play in responsible AI development, and how does this align with AI trajectories, structures, and institutions in AI research (Himmelreich, 2019; Stahl, 2023; Tzachor et al., 2022)? The problem of responsible and humane AI technology requires much broader consideration than can be offered by a mere focus

on hybrid human-AI, as the following descriptions of a collaborative project between LT and cultural anthropology will demonstrate.

2. THE CONUNDRUMS OF VALUE-ORIENTED DESIGN IN LANGUAGE TECHNOLOGY

Collaborating with the language technology research group — computer scientists working on the automation of speech — on the DWISE project included first-hand ethnographic observation of the design process for a hybrid human-AI software project, which was revelatory for my long-standing research interests in technological design as a social process and in the social shaping of technology development (Koch et al., 2022). As expected, the processes and practicalities of collaboration were complex. What was unexpected were the dominance of structures that oppose responsible and humane AI technology development, such as path dependencies or exploitative and extractive work practices. That was an unforeseen and peculiar aspect of a collaborative, cross-disciplinary approach. The DWISE project reproduced the structural problems and conundrums of creating more responsible and humane versions of AI in complex social contexts, which are:

i. AI technologies (like all technologies) are unevenly distributed globally (Ong & Collier, 2005), with global distributions of labour, inequalities and ontological differences across scales and sites (Wahlberg, 2022).
ii. The unconscious: invisible cultural and social inscriptions in technologies affect ontologies and presuppositions about what the technology can and cannot do (Akrich, 1992; Araujo de Aguiar et al., 2022).
iii. The implementation of technologies is the point at which they become a political and technological fact in everyday life, when competing interests, normative orientations and intentions to change the allocation of wealth, power and social status feed back into design processes (Butot & van Zoonen, 2024; Pfaffenberger, 1992).

This was also evident in the DWISE project. The following discusses these three conundrums in more detail, before the paper suggests rethinking the social and technological facticity of AI along these lines.

2.1 AI as a global technological form – Who is in the loop and with what conditions?

AI is a globally distributed technology (Collier & Ong, 2005), involving infrastructures, expertise and capacities in research, technology, implementation and impact on everyday life. However, the capacity to drive developments is unevenly distributed, and there is ongoing criticism of neo-colonial tendencies in the Global South (Gurumurthy & Chami, 2019; Lazem et al., 2022). These

tendencies are not necessarily related to place or global distribution, but rather to power relations and the recognition of work and resources, and they can occur anywhere.

In the DWISE project, conceptual innovations were achieved through the collaboration of LT and anthropologists, who met bi-weekly to share knowledge, develop a common language and organise programming and testing. Quarterly strategic planning sessions reflected these results and determined system features and priorities. From a humanities perspective, the anthropologists contributed methodological knowledge of grounded theory (Bryant & Charmaz, 2019), numerous hours of testing and discussing the results to improve LT, which fed into the dissertation projects of the three language technologist. These AI innovations have the potential to enrich digital ethnographic discourse analysis by offering new ways to filter and structure research materials, facilitating hermeneutic processes and providing new perspectives on research fields, as well as enhancing the possibilities for triangulating research data (Eiser et al., 2023; Fischer et al., 2024; Koch et al., 2026). However, realising these improvements to ethnographic research demands going beyond the proof of concept by developing it into a user-friendly software system and introducing and implementing it in the anthropological scientific community. This did not materialise because the project has yet to progress beyond the proof-of-concept stage. For anthropologists, the idea of human and AI knowledge growth in hybrid systems proved to be illusory when the language technologists moved on to their next projects. It is a small consolation that the creative and repetitive work the anthropology researchers had to do in the project, which supported machine learning and was necessary for improving AI technology, was carried out under slightly better working conditions than the exhausting global divisions of labour (Reddyhoff, 2022; Siapera, 2022).

2.2 Ontological bias in AI technology – the question of diversity

In the anthropology of technology, bias is known as 'social inscription' in technology. Highlighting and elaborating on the fact that technologies are not neutral but come with social inscriptions has been an important step in the social studies of science and technology, as has the development of methods and approaches to identify these inscriptions. In AI, the question of inscription extends the perspective from the training materials to the core of the machine itself: its software and hardware architecture, which also come with biases and limitations regarding how the diversity of the social world can be modelled (Blodgett et al., 2020; Hovy & Prabhumoye, 2021), thus constituting a structural problem. Due to path dependencies, English vocabulary forms the backbone of LL

Ms and thus perpetuates the cultural knowledge inscribed in English. Lexical gaps challenge the quality of LLM translations and cannot be compensated by adding more languages or training alone (see Figure 1); it is a representation problem (Giunchiglia et al., 2017). Modelling language diversity as a general linguistic principle would require different approaches at the level of data

models and algorithms that would at least work towards resolving the problem of lexical gaps (Khalilia et al., 2023).[1]

TABLE 1 Lexicalizations of nine meanings around the concept of (sibling) in eight languages.

Meaning	English	Japanese	Arabic	Italian	Indonesian	Hindi	Hungarian	Javanese
sibling	sibling	GAP	GAP	GAP	saudara	सहोदर	testvér	sedulur
elder sibling	GAP	GAP	GAP	GAP	kakak	GAP	nagytestvér	GAP
younger sibling	GAP	GAP	GAP	GAP	adik	GAP	kistestvér	adhi
brother	brother	GAP	أَخْ	fratello	GAP	भैया	GAP	GAP
sister	sister	GAP	أُخْت	sorella	GAP	बहन	GAP	GAP
elder brother	GAP	あに	GAP	fratellone	abang	भैया	báty	kangmas
elder sister	GAP	あね	GAP	sorellona	GAP	दीदी	növér	mbakyu
younger brother	GAP	おとうと	GAP	fratellino	GAP	भाई	öcs	GAP
younger sister	GAP	いもうと	GAP	sorellina	GAP	बहन	húg	GAP

Figure 1: Lexical gaps in the translation of kinship (Khalilia et al. 2023)

At the level of hardware and software, the problem is that these biases may not become immediately apparent to users or developers. They are frequently labelled as colonial bias, as the large languages — English, French, Spanish, etc. — have a colonial tradition and dominate LT, while smaller languages are not so well represented.

These gaps are inextricably linked to the current global form of LT; they are present on a material level, the level of resources. Only a significant realignment of LT in this sense could initiate a shift towards a more humane and responsible version of AI.

2.3 The implementation of AI as technological dramas and discourses

Another conundrum arises when AI enters 'the social wild' and becomes embedded in everyday life or specific fields of application. This is not a self-driven process, but is instead motivated by the interests of various stakeholders, as well as by the redistribution of status, power and wealth. The launch of OpenAI's ChatGPT is a prime example of this, generating enormous interest in

[1] Symbolic AI, top-down AI as a specific path and design approach, today complements machine learning and materialises in the research agendas of "responsible AI", "explainable AI", "adaptive AI", or "model-based diagnosis by AI", often aimed at improving LLM results. Coalitions between the two design paradigms, top down and bottom-up approaches to AI, allow their different capacities and strengths to feed into each other. Lexical databases such as LL, built top-down by humans, are themselves language resources that can be used as training materials for LLMs, and have historically been an important starting point for their training.

LLMs and initiating their widespread implementation in everyday life. This has resulted in multiple new applications in various contexts, both personal and professional, often with disruptive consequences, such as the automatic generation of fake content on the internet or job losses in the service sector. Given the disruption that new technologies can cause, which largely escapes the control of LLM designers, the anthropologist of technology Brian Pfaffenberger (1992) conceptualised implementation processes as “technological dramas” and a “technological” mode of discourse. In this sense, technology is always incomplete and any activity in the implementation of LLMs creates new ideas and needs. This feeds back into design processes, as emphasised by Michel Callon, one of the founding fathers of science and technology studies (Callon, 2004). However, the resulting disruptions take on a life of their own when fake information feeds into new versions of LLM technology, which is not necessarily conducive to the development of responsible and humane AI.

These three conundrums demonstrate the dilemma of hybrid human-AI systems, as the path to a more responsible and humane version of AI is multifaceted and obviously does not align with the current concepts of AI technology. They raise serious doubts as to whether such a version of AI can exist at all without a fundamental realignment of AI research and development. Such a shift in AI would require reinvention and massive revision based on a much broader recognition of the social and cultural context of knowledge and technology, as well as comprehensive knowledge of the multiple socio-technical dimensions and consideration of the intersecting dimensions in technology design. A presupposition for such a realignment would be abandoning the myth, that there can be an artificial version of intelligence, which not only ignores the lack of a scientific definition of the term, but also disregards the social embeddedness of both, intelligence and technology. Furthermore, responsibility requires us to turn back to scientific evidence rather than following speculations, as per Turing's black boxing (Halpern, 2006).

3. RETHINKING LANGUAGE TECHNOLOGY ALONG THE LINES OF SOCIAL AND TECHNOLOGICAL FACTICITY

Rather than envisaging responsible and humane hybrid human-AI that performs knowledge work with AI or humans in the loop, the socio-technical systems perspective (Ropohl, 1999) shifts the focus from intelligent machinery to the development of robust technology that can be used in a humane and responsible way in different contexts. This focus highlights that, like other socio-technical systems, AI always emerges from an interplay of human and machine activity. Shifting the focus away from intelligence and towards the technological practices that bring an AI LT system into being broadens the view to encompass the many social contexts and dimensions through which a system gains social and technological facticity, and thus its value orientation, embedded and distributed in diverse socio-technical system practices.

The socio-technical model of technological design described in the following provides a heuristic for studying, reflecting on and implementing diverse practices through which value-oriented, responsible and humane AI systems are constituted. It is the result of a collaboration with the LT group of Fausto Giunchiglia, and further colleagues Gabor Bella and Paula Helm, as part of the LL initiative (Koch et al., 2024). The project investigated how language diversity beyond colonial languages can be incorporated into LT, and how LT can be de-linked from colonial trajectories of epistemic injustice and exploitative and extractive data governance (Giunchiglia et al., 2017; Helm et al., 2023). The model (see figure 2) provides an overview of the relevant activities in the socio-technical system of AI LT design. These activities are arranged in layers

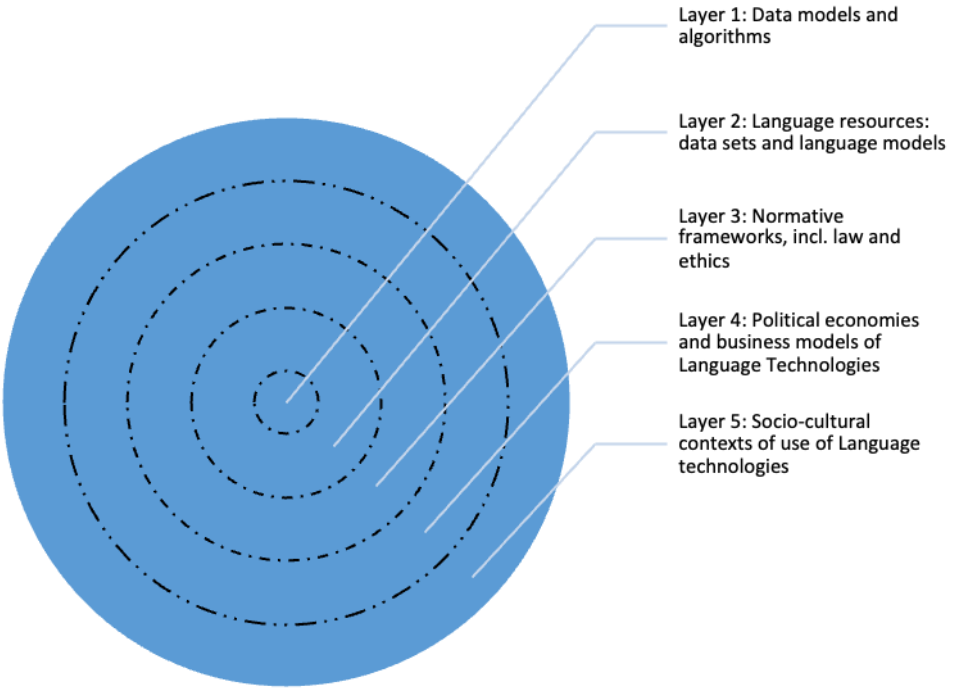

Figure 2: Layers of technological design (Koch et al., 2024)

around the data models, algorithms and language resources, which form the technical core of LL systems. Each layer is characterised by specific bundles of practices involving typical actors, while remaining permeable to the exchange of information and forming part of the global field of AI. The model serves as a heuristic for ethnographically describing technological activities[2] with-

[2] The term "technological activity" refers to the methodological approach for observing technical activities on a micro level (Coupaye, 2022), which is expanded here to the meso level.

in the socio-technical system, and acts as a boundary object (Star & Griesemer, 1989) for facilitating transdisciplinary conversations between stakeholders involved in the design of AI technology, as well as for investigating the presence of humane and responsible values within AI design. Layers 1 and 2 form the technological core consisting of data models, algorithms and language resources, including data sets and language models. These two layers form the operational unit and central core of the socio-technical system. The design approaches in layers 1 and 2 depend heavily on the expertise of computer scientists and knowledge domain experts, in this case in language and culture. This knowledge must be translated into computer data logics and algorithms. These layers provide a 'playground' for value-oriented design approaches and interdisciplinary expertise in computer technology and linguistics. Diversity is an important value in LT as it responds to language diversity, which requires adequate representation and consideration in hardware and software architecture.

Layers 3 to 5 are where value-oriented technology meets the 'social wild'; these layers are the spaces in which technological dramas unfold (Pfaffenberger, 1992). Further experts and stakeholders join in with specific interests in the distribution of wealth, power and social status. These interests are negotiated in local technological dramas and feed back into the technological design of layers 1 and 2. AI and language experts participate in the negotiations of technological dramas, competing for their interests and motivations. The world around the technological realm explores the value-oriented design of layers 1 and 2 for its own purposes. Diversity is embedded in language and thus a requirement emerging from the knowledge domain modelled by LT — enriched through scientific approaches to diversity, such as those in anthropology — which expand the concept of diversity beyond language to encompass social and cultural processes, global power relations, and the associated normative debates. In this context, the postcolonial anthropologist Arturo Escobar (2018) speaks of “diversality” to promote the diversity of knowledge as a principle in human life, while other voices point to the problematic implications of diversity as a concept (Zembylas, 2023).

Layer 3 illustrates how technological ideas and design paradigms interact with different normative frameworks and laws, such as the European AI Act, as well as ethics more broadly. While these are the most apparent frameworks, the range of normative frameworks is much broader and encompasses everyday moral orientations, meaning systems and epistemic systems. The normative layer is distinct from yet intertwined with the core layers. It embraces the unconscious dimensions of embedded worldviews, yet it can also function as a corrective or reflexive level for technological options or path dependencies in the core.

Layer 4: Designing and running LT is costly and time-consuming, for example in terms of training LLMs. As LT requires resources and is embedded in political economies, big American companies dominate and govern LT development in terms of economic output and power, rather than following social and cultural ideas of maintaining diversity, enabling social equality and epistemic justice or avoiding the extractivism of specific knowledge embedded in small or indigenous

languages. For shifting agendas towards new participatory and community-driven forms of LT, adequate business models and forms of resourcing are required.

Layer 5: The contexts in which a technology is used are another important design area. Pfaffenberger points out that people accept, change or avoid new technologies in different ways. Anthropology has extensively studied what people make of/with technologies (Bruun et al., 2022), as well as how design anthropological approaches can be used to intervene in technological design (Koch et al., 2024).

The layers are permeable ideal types. At a more abstract level, each layer constitutes a trading zone of knowledge (Gorman, 2002) for value-oriented design and diversity in the LL initiative. Each layer also constitutes a boundary object (Star & Griesemer, 1989) coordinating cooperation among transdisciplinary stakeholders with different knowledge backgrounds and communities and specific stakes in AI development. Nevertheless, frictions, contingency and messiness are prominent social forms (Tsing, 2004) in each layer, in which the global connections of AI development unfold as technological activities. Therefore, any technological activity at any level of design can affect design paradigms at levels 1 and 2, and vice versa. This can influence value-oriented AI design through negotiations, explorations and constraints imposed by multiple factors in the 'social wild'.

4. CONCLUSION: THE FUTURE OF VALUE-ORIENTED DESIGN IN HYBRID HUMAN-AI APPROACHES

The development of humane and responsible AI is a highly relevant issue in democratic societies, and there is a consensus that more effort is needed in this area. The current focus on 'hybrid human–AI' is too narrow, as demonstrated by insights from the anthropology of technology. Humane and responsible AI are not outcomes of hybrid human-AI systems, but ongoing tasks that require continuous effort at all levels of design and implementation. Social processes in technological activity deserve more appreciation and consideration. Recognising the problem-solving capacity of social processes requires the systematic integration of social processes at all levels of technological design, rather than labelling design approaches such as hybrid human–AI systems as ethical and responsible versions of AI. Strengthening the social research and design of technological activities in AI is essential to closing a basic gap and overcoming a major challenge in stabilising humane and responsible AI technologies, ensuring their democratic use and implementation. A first step towards responsible AI would be to abandon the evocative and misleading concept of 'intelligence' in the name of the research field, and instead adopt more precise terms or defined scientific terminology that highlights the socio-technical nature of AI rather than the idea of a superpower.

Such a shift towards socio-technological AI design is characterised by the fact that it does not require an evocative, imaginary term such as 'artificial intelligence' to describe a range of methods in fields of computer science that are only loosely connected, such as robotics, LT or logic reasoning. Instead, it uses designations that develop a differentiated socio-technical understanding of the relationship between humans and technology, allowing precise designation of the various relations, practices and fields currently combined under the umbrella of AI, a term with more visionary than paradigmatic strength.

This would mean that AI technology design would no longer have to rely on imitation games based on a black box for evaluation — a logic massively amplified by machine learning — but could instead systematically name and highlight the technology's achievements alongside the various dimensions of responsibility within this socio-technical system, including its more problematic aspects.

5. REFERENCES

Akrich, M. (1992). The de-scription of technical Objects. In *shaping technology/building society. Studies in sociotechnical change* (pp. 205–224). MIT Press. https://shs.hal.science/halshs-00081744/

Araujo de Aguiar, C. H., Pinch, T., & Green, K. (2022). De-scription at early phases of artifact design. In *Proceedings of the 25th International Academic Mindtrek Conference* (pp. 179–191). ACM. https://doi.org/10.1145/3569219.3569380

Batool, A., Zowghi, D., & Bano, M. (2023). *Responsible AI governance: a systematic literature review*. arXiv. https://arxiv.org/abs/2401.10896v1

Blodgett, S. L., Barocas, S., Daumé, H., III, & Wallach, H. (2020). *Language (technology) is power: A critical survey of "bias" in NLP.* In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485

Bruun, M. H., Wahlberg, A., Douglas-Jones, R., Hasse, C., Hoeyer, K., Kristensen, D. B., & Winthereik, B. R. (Eds.). (2022). *The Palgrave handbook of the anthropology of technology*. Palgrave Macmillan. https://doi.org/10.1007/978-981-16-7084-8

Bryant, A., & Charmaz, K. (2019). *The SAGE handbook of current developments in grounded theory*. SAGE.

Butot, V., & van Zoonen, L. (2024). Contesting infrastructural futures: 5g opposition as a technological drama. *Science, Technology, & Human Values*, *49*(5), 1017–1044. https://doi.org/10.1177/01622439221147347

Callon, M. (2004). The role of hybrid communities and socio-technical arrangements in the participatory design. *Journal of the Center of Information Studies*, *5*(3), 3–10. https://www.comm.tcu.ac.jp/cisj/05/5_01.pdf

Constantinescu, M., Voinea, C., Uszkai, R., & Vică, C. (2021). Understanding responsibility in Responsible AI. Dianoetic virtues and the hard problem of context. *Ethics and Information Technology*, *23*(4), 803–814. https://doi.org/10.1007/s10676-021-09616-9

Coupaye, L. (2022). Making 'technology' visible: technical activities and the Chaîne Opératoire. In *The Palgrave Handbook of the anthropology of technology* (pp. 37–60). Palgrave Macmillan. https://doi.org/10.1007/978-981-16-7084-8_2

Eiser, I., Fischer, T., Schneider, F., Koch, G., Biemann, C., & Petersen Frey, F. (Eds.) (2023). *Open Science Prinzipien und interdisziplinäre Kollaboration in D-WISE: Zwischen Hermeneutik und Digitaler Methode in der Diskursanalyse*. Zenodo.

Escobar, A. (Ed.). (2018). *New ecologies for the twenty-first century. Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Duke University Press. https://doi.org/10.1215/9780822371816

Fischer, T., Schneider, F., Petersen-Frey, F., Haque, A. S. M., Eiser, I., Koch, G., & Biemann, C. (2024). Extending the Discourse Analysis Tool Suite with whiteboards for visual qualitative analysis. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 7017–7022). ELRA and ICCL. https://aclanthology.org/2024.lrec-main.615/

Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of responsible AI: From ethical guidelines to cooperative Policies. *Frontiers in Computer Science*, *4*, Article 873437, 873437. https://doi.org/10.3389/fcomp.2022.873437

Giunchiglia, F., Batsuren, K., & Bella, G [Gabor] (2017). Understanding and exploiting language diversity. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence.* International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2017/560

Gorman, M. E. (2002). Levels of expertise and trading zones. *Social Studies of Science*, *32*(5-6), 933–938. https://doi.org/10.1177/030631270203200511

Gurumurthy, A., & Chami, N. (2019). *The wicked problem of AI governance* (Artificial Intelligence in India, No. 2). Friedrich-Ebert-Stiftung India Office. https://doi.org/10.13140/RG.2.2.14753.22886

Halpern, M. (2006). The trouble with the Turing test. *The New Atlantis*(11), 42–63.

Helm, P., Bella, G [Gábor], Koch, G., & Giunchiglia, F. (2023). *Diversity and language technology: How techno-linguistic bias can cause epistemic injustice*. arXiv. https://doi.org/10.48550/arXiv.2307.13714

Himmelreich, J. (2019). Responsibility for killer robots. *Ethical Theory and Moral Practice*, *22*(3), 731–747. https://doi.org/10.1007/s10677-019-10007-9

Himmelreich, J., & Köhler, S. (2022). Responsible AI through conceptual engineering. *Philosophy & Technology*, *35*(3). https://doi.org/10.1007/s13347-022-00542-2

Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, *15*(8), e12432. https://doi.org/10.1111/lnc3.12432

Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society*, *9*(2), Article 20539517221142824. https://doi.org/10.1177/20539517221142824

Khalilia, H., Bella, G., Freihat, A. A., Darma, S., & Giunchiglia, F. (2023). Lexical diversity in kinship across languages and dialects. *Frontiers in Psychology*, *14*, 1229697. https://doi.org/10.3389/fpsyg.2023.1229697

Koch, G. (2005*). Zur Kulturalität der Technikgenese: Praxen, Policies und Wissenskulturen der künstlichen Intelligenz*. Röhrig.

Koch, G., Bella, G [Gábor], Helm, P., & Giunchiglia, F. (2024). Layers of technology in pluriversal design decolonising language technology with the live language initiative. *CoDesign*, *20*(1), 77–90. https://doi.org/10.1080/15710882.2024.2341799

Koch, G., Biemann, C., Eiser, I., Fischer, T., Schneider, F., Stumpf, T., & Tijerina García, A. (2022). D-WISE Tool Suite for the sociology of knowledge approach to discourse analysis. In M. Rauterberg (Ed.), *Culture and Computing: 10th International Conference, C&C 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings* (1st ed. 2022, Vol. 13324, pp. 68–83). Springer International Publishing; Imprint Springer. https://doi.org/10.1007/978-3-031-05434-1_5

Koch, G., Stumpf, T., & Haque, A. S. M. (2026). Grounded Theory geleitetes Strukturieren in KI-gestützten Forschungsprozessen am Beispiel digitaler Diskursanalysen. In C. Pentzold, N. Heise, & Bischof Andreas (Eds.), *Praxis Grounded Theory* (2. edition). Springer VS.

Lavi, E., & Reich, Y. (2024). Cross-disciplinary system value overview towards value-oriented design. *Research in Engineering Design*, *35*(1), 1–20. https://doi.org/10.1007/s00163-023-00418-2

Lazem, S., Giglitto, D., Nkwo, M. S., Mthoko, H., Upani, J., & Peters, A. (2022). Challenges and paradoxes in decolonising HCI: a critical discussion. *Computer Supported Cooperative Work (CSCW)*, *31*(2), 159–196. https://doi.org/10.1007/s10606-021-09398-0

Lowenhaupt Tsing, A. (2005). *Friction: An ethnography of global connection*. Princeton University Press. https://doi.org/10.1515/9781400830596

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys*, *56*(7), 1–35. https://doi.org/10.1145/3626234

Ong, A., & Collier, S. (Eds.). (2005). *Global assemblages: technology, politics, and ethics as anthropological problems*. Blackwell Publishing.

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Pfaffenberger, B. (1992). Technological dramas. *Science, Technology, & Human Values*, *17*(3), 282–312. https://doi.org/10.1177/016224399201700302

Reddyhoff, D. (2022). *Dependency, data and decolonisation: a framework for decolonial thinking in collaborative AI research*. arXiv. https://doi.org/10.48550/arXiv.2206.03212

Reichert, R., & Richterich, A. (2015). Introduction: Digital materialism. *Digital Culture & Society*(1), 5–18. https://www.degruyter.com/document/doi/10.14361/dcs-2015-0102/html

Rismani, S., & Moon, A. (2023). What does it mean to be a responsible AI practitioner: An ontology of roles and skills. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 584–595). ACM. https://doi.org/10.1145/3600211.3604702

Ropohl, G. (1999). Philosophy of socio-technical systems. *Society for Philosophy and Technology Quarterly Electronic Journal, 4(3), 186–194. https://doi.org/10.5840/techne19994311.*

Shneiderman, B. (2020). Human-centered artificial intelligence: reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Siapera, E. (2022). AI content moderation, racism and (de)coloniality. *International Journal of Bullying Prevention*, *4*(1), 55–65. https://doi.org/10.1007/s42380-021-00105-7

Spiekermann, S., & Winkler, T. (2020). Value-based engineering for ethics by design. *SSRN Electronic Journal*. Advance online publication. https://doi.org/10.2139/ssrn.3598911

Stahl, B. C. (2023). Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems. *Scientific Reports*, *13*(1), 7586. https://doi.org/10.1038/s41598-023-34622-w

Tzachor, A., Devare, M., King, B., Avin, S., & Ó hÉigeartaigh, S. (2022). Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nature Machine Intelligence*, *4*(2), 104–109. https://doi.org/10.1038/s42256-022-00440-4

Zembylas, M. (2023). A decolonial critique of 'diversity': Theoretical and methodological implications for meta-intercultural education. *Intercultural Education*, *34*(2), 118–133. https://doi.org/10.1080/14675986.2023.2177622.