



DISCUSSION

ChatGPT and Beyond: Exploring the Responsible Use of Generative AI in the Workplace

An Interdisciplinary Perspective

Matthias Söllner · Thomas Arnold · Alexander Benlian · Ulrich Bretschneider · Caroline Knight · Sandra Ohly · Lena Rudkowski · Gerhard Schreiber · Domenik Wendt

Published online: 12 February 2025
© The Author(s) 2025

Keywords ChatGPT · Generative AI · Artificial intelligence · Interdisciplinary · Algorithms · Workplace

1 Motivation

Matthias Söllner

Artificial intelligence (AI)-based systems are increasingly pervading most parts of our everyday life. Whether we are shopping online or looking for information at work, providers nowadays rely on AI-based models and seek to provide us with tailored support or guidance. Recently, a novel class of AI-based systems has gained widespread attention around the world. At the latest when Open AI's

ChatGPT reached 100 million users in just over 2 months after launch (by comparison, it took TikTok 9 months, and Instagram 2.5 years to reach this level (Bhaimiya 2023; UBS 2023)), the potential as well as the challenges associated with Generative AI (GenAI) are widely discussed in academia, industry, and the public.

When speaking of GenAI, we refer to "...computational techniques that are capable of generating seemingly new, meaningful content such as text, images or audio from training data" (Feuerriegel et al. 2024, p. 111). Recent studies show that GenAI has great potential when it comes to increasing the productivity of knowledge workers. For example, an experimental study observed significant effects in terms of time saving as well as quality improvement when ChatGPT was used in the context of mid-level professional writing tasks. The same study also observed that using ChatGPT decreased the inequality between participants, since participants with lower ability profited more, indicating that GenAI systems could also foster equality (Noy and Zhang 2023). Another experimental study focusing on software developers reports comparable findings. With the use of GitHub's Copilot, the average completion time of a standard programming task decreased by more than 50%, with no significant difference in success rate. In this study, less experienced developers, developers with high coding loads, as well as older developers benefited the most from using Copilot (Peng et al. 2023). Consequently, first studies suggest several benefits of leveraging the potential of GenAI in the workplace.

At the same time, GenAI also presents several challenges that need to be addressed. A prominent challenge vividly discussed is so-called *hallucination*, which refers to circumstances in which GenAI systems provide incorrect information, e.g., fabricated references or fabricated

M. Söllner (✉) · U. Bretschneider · S. Ohly
Research Center for IS Design (ITeG), University of Kassel,
Pfannkuchstraße 1, 34121 Kassel, Germany
e-mail: soellner@uni-kassel.de

T. Arnold · A. Benlian
Technical University of Darmstadt, Karolinenplatz 5,
64289 Darmstadt, Germany

C. Knight
The University of Queensland, Colin Clark, 39 Blair Dr,
St Lucia, QLD 4067, Australia

L. Rudkowski
Justus Liebig University Gießen, Licher Straße 72,
35394 Gießen, Germany

G. Schreiber
Helmut Schmidt University (HSU/UniBwH), Holstenhofweg 85,
22043 Hamburg, Germany

D. Wendt
Frankfurt University of Applied Sciences, Nibelungenplatz 1,
60318 Frankfurt Am Main, Germany

content from existing references (Bhattacharyya et al. 2023). To solve this issue, computer science researchers are working on ways to mitigate hallucination if possible (Vynck 2023). A second challenge relates to the exact opposite of hallucination – plagiarism. For example, in December 2023, the New York Times announced that it will sue Open AI and Microsoft over copyright violations, because ChatGPT reproduced large parts of New York Times articles (Grynbaum and Mac 2023). Both these challenges emphasize a technical challenge in the development of GenAI systems, when it comes to finding the right balance between avoiding hallucination and plagiarism at the same time. Additionally, GenAI also comes with challenges beyond the technical domain. Universities, for example, struggle with the impact of GenAI on exam formats, such as written assignments. One case that received significant media attention was a report by a University of Pennsylvania professor that ChatGPT had generated results that he would grade as a B or B- on questions of the final exam of his Wharton MBA course (Terwiesch 2023). Furthermore, GenAI is also considered to have a significant impact on industry. For example, analysts predict that hundreds of millions of jobs will be lost or degraded by GenAI (Kelly 2023). While other analysts reach different conclusions and emphasize that GenAI will rather influence or reshape jobs (Chui et al. 2023; Gownder and O’Grady 2023), we can conclude that GenAI will have an impact on jobs across industries. As a result, organizations need to assess the impact of GenAI on, e.g., their business model, core processes, and the job profiles of current and future employees.

Consequently, leveraging the potential of GenAI and accounting for its challenges at the same time, should be viewed as a sociotechnical design challenge. Given the tradition of our discipline, information systems researchers are well-positioned to address this sociotechnical design challenge, and to engage in interdisciplinary collaborations with colleagues from both the technical and social sciences with the goal of responsibly designing GenAI models and systems.

This discussion paper is one outcome of such an interdisciplinary endeavor. In November 2022, the Hessian Centre Responsible Digitality funded an interdisciplinary project group on “Responsible Algorithmic Decision-Making in the Workplace” that combines expertise from the domains of information systems, business psychology, computer science, law, marketing, and technology ethics. Over the course of 1.5 years, we had intense discussions among the members of the project group and organized different workshops and conferences to broaden the perspective beyond the expertise of our project group, e.g., inviting other researchers, but also industry experts, as well as employer and employee representatives. The concluding

conference of our project group focused specifically on the impact of GenAI in the workplace and its responsible design and use. Consequently, this discussion paper mainly draws on the insights and discussions from this conference but is also informed by all discussions we had within the project group.

The following contributions to the discussion seek to shed light on the topic from the various perspectives present at our last conference:

- Information Systems (Alexander Benlian and Ulrich Bretschneider)
- Computer Science (Thomas Arnold)
- Business Psychology (Sandra Ohly and Caroline Knight)
- Law (Lena Rudkowski and Domenik Wendt)
- Technology Ethics (Gerhard Schreiber)

Given the sociotechnical nature of the design challenge at hand, we believe that engaging with these different perspectives and finding synergies, as well as discussing and solving potential conflicts, is a key component of responsible design and use of GenAI in the workplace. Figure 1 illustrates this logic by highlighting that even though every discipline has its own perspective, there will be areas where disciplines overlap, and interdisciplinary discourse is valuable. From the perspective of responsible interdisciplinary design and use, we expect that the best solutions are to be found near the center of the diagram, where all perspectives merge.

2 Promises, Perils and Pathways to Responsible Use of Generative AI in the Workplace – Insights from Information Systems Research

Alexander Benlian, Ulrich Bretschneider

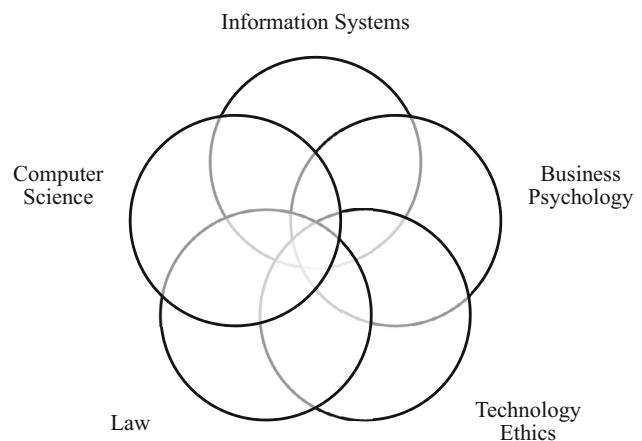


Fig. 1 Interdisciplinary perspective on responsible design and use of generative AI in the workplace

GenAI is spearheading a revolutionary change in various organizational areas, fundamentally altering the landscape of value creation. In the realm of business process management and optimization, GenAI is a game-changer, streamlining operations and sparking a wave of innovative methodologies. Recent research by Kecht et al. (2023) and van Dun et al. (2023) underscores GenAI's prowess in enhancing process documentation and inspiring novel process development. In ideation, GenAI leverages extensive online textual data to substantially broaden the knowledge repositories of project teams, thus democratizing and economizing innovation (Bouschery et al. 2023). Additionally, GenAI is a boon to the overtaxed customer support sector, paving the way for improved workforce productivity (Reinhard et al. 2024). These examples are just a snapshot of how GenAI is empowering workers in organizations and transforming their traditional roles.

However, the promise of GenAI is tempered by concerns over the accuracy of its output. As several researchers have pointed out, these systems can fall prey to producing errors, or hallucinations, when trained on datasets that contain inconsistencies or errors (e.g., Banh and Strobel 2023; Feuerriegel et al. 2024). This problem of AI reliability has gained global attention, with the World Economic Forum's 2024 Global Risk Report citing AI-generated misinformation as a significant risk (WEF 2024). An example of this issue could be seen in AI systems used for medical diagnosis, where overreliance on incorrect AI-generated content has led to misdiagnosis or inappropriate treatment plans. This emphasizes the criticality of diligent and thoughtful examination of GenAI applications in designs that involve human collaboration (e.g., human-in-the-loop), to ensure their outputs are stringently checked.

The advent of GenAI also comes with the issue of potentially infringing on copyrights and the moral rights of original creators, as noted by Smits and Borghuis (2022). For example, consider a GenAI that produces a novel by closely mimicking the style and plot of a well-known author. This could constitute copyright infringement and violate the author's so-called moral right of attribution, which allows an author to be attributed for his work. Similarly, if an AI-generated painting closely resembles a famous artist's work, it could be seen as a breach of the artist's moral right to integrity, particularly if the AI's version is presented in a context that damages the original artist's reputation. These examples underscore the need for careful consideration of both legal and ethical implications when using GenAI.

Another critical concern with GenAI is its susceptibility to replicating biases present in its training data. When these systems are not adequately equipped to identify and eliminate inherent human biases, they risk perpetuating damaging stereotypes or offensive language (Spiekermann

et al. 2022). For instance, research by Hartmann et al. (2023) highlights this problem, and further studies into AI models like CLIP and the LAION dataset (Birhane et al. 2021; Wolfe et al. 2022) have confirmed the presence of such biases. A practical example of this could be an AI system trained on historically gender-biased job data, which might then suggest careers to individuals based on gender stereotypes. Incorporating such biased AI outputs in everyday business activities not only poses ethical dilemmas but could also tarnish a company's reputation.

In light of these emerging challenges, the responsible use of GenAI applications in the workplace is becoming increasingly paramount. Employees must consider several critical aspects to use GenAI responsibly: First of all, it is essential for knowledge workers to acquire professional AI literacy, which is key to responsibly evaluating the information and content produced by AI (Pinski and Benlian 2024). For instance, a data analyst must be adept at discerning the accuracy of AI-generated data interpretations. Next, workers also have a greater moral responsibility, particularly with regard to the ethical use and possible copyright infringements of content created by AI (Mikalef et al. 2022). For example, it is up to a graphic designer to ensure that AI-generated designs do not infringe on existing copyrights, so he or she has to bring in a greater sense of responsibility. On the whole, employee intuition is becoming increasingly important for the responsible review, identification and correction of underlying biases in AI-generated content. As a result, employees will find their roles evolving toward greater responsibility when working with GenAI.

However, this transformation of work imposes responsibilities not only on employees but also on managers. Managers need to develop their own AI knowledge and establish frameworks for responsible GenAI-based work (Pinski et al. 2024). This means, for example, to increase employees' awareness for the responsible use of GenAI, which includes that managers have to provide employees with knowledge about not only the opportunities but particularly the risks and the consequences of a wrong use of GenAI for daily value creation. Managers must also implement corporate policies to guide the acceptable use of GenAI. These policies might include guidelines on data handling, AI-generated content review processes, and protocols for reporting AI misuse. Furthermore, offering comprehensive training programs is crucial. These programs could feature practical modules on integrating GenAI into daily tasks, ranging from automating routine data analysis to enhancing creative projects. Such training would empower employees to leverage GenAI's benefits while effectively navigating potential challenges and risks.

In addition to the imperative for both employees and managers to utilize GenAI responsibly within

organizations, there is a critical need for research to explore how to support this responsible use. This requires an in-depth examination of the dynamics of human-AI collaboration and the rise of hybrid intelligence (e.g., Dellermann et al., 2019). Such inquiry necessitates a nuanced understanding of the interplay between human responsibility and AI efficiency. To effectively foster potentially symbiotic relationships, three lines of research deserve particular attention:

Optimizing Collaboration Models: Future inquiries should investigate the most effective models for human-AI collaboration in various organizational contexts (e.g., Adam et al. 2023). This includes understanding how AI can augment human work and decision-making without undermining it, and identifying scenarios where AI should take the lead versus those where human insight is paramount (e.g., Baird and Maruping 2021; Fügener et al. 2021). Studies should focus on developing frameworks (e.g., human-in-the-loop, human-on-the-loop, human-out-of-the-loop) that delineate clear roles and responsibilities for both humans and AI in various business contexts. Additionally, research could explore how to balance AI's computational power with human ethical judgment and creativity, ensuring that AI acts as a complement rather than a substitute.

Understanding Hybrid Intelligence Dynamics: Another crucial research direction is to unravel the dynamics of hybrid intelligence (e.g., Akata et al. 2020; Fabri et al. 2023). This involves studying how human socio-cognitive abilities and GenAI's content-generating capabilities can be synergistically combined for enhanced problem-solving and innovation. Key themes include examining the psychological aspects of human-AI interaction, such as trust, reliance, and emotions, as well as the technical (i.e., design science) aspects, such as interface design, feedback mechanisms, and shared control systems. Understanding these dynamics is vital for creating collaborative environments where humans and AI can co-create and learn from each other effectively (Berger et al. 2021).

Enhancing Algorithmic Management: A third avenue for future research lies in exploring how GenAI could fundamentally change the domain of algorithmic management and control (e.g., Benlian et al. 2022; Wiener et al. 2023). This research could focus on how GenAI can automate and optimize decision-making processes in organizational settings. Specifically, it may investigate the role of GenAI in developing dynamic and adaptable management systems that respond in real-time to changes in the workplace, enhancing operational efficiency. Additionally, this research could delve into the ethical implications of GenAI-driven management, ensuring that such systems uphold principles of fairness, security, and transparency,

while also preserving the autonomy, safety, and morale of employees (Mikalef et al. 2022; Cram et al. 2024).

These research directions underscore the importance of a balanced and thoughtful integration of GenAI into the fabric of organizational operations, ensuring that the partnership between humans and AI is harmonious, ethical, and mutually enriching.

3 Hallucinations and Explainability – Insights from Computer Science

Thomas Arnold

GenAI has emerged as a powerful tool with the potential to revolutionize various aspects of work. From automating content creation to streamlining product design, GenAI offers exciting opportunities for increased efficiency and innovation. However, as with any powerful technology, responsible use is paramount, and computer science considerations play a crucial role in ensuring GenAI integration benefits both businesses and employees.

One of the key challenges from a computer science perspective lies in GenAI's inherent tendency towards hallucination (Ji et al. 2023). Unlike traditional AI models trained on existing data for pattern recognition, GenAI can create entirely new content. This creativity, while valuable, carries the risk of generating outputs that are factually incorrect, misleading, or simply nonsensical.

Here's a closer look at the technical hurdles related to hallucination and ongoing efforts to overcome them:

3.1 Data Biases

Data biases pose a significant challenge in GenAI, as these models can amplify and reflect the biases present within their training data. For instance, a model trained on a dataset with a gender imbalance might consistently generate content where male doctors are depicted as the norm, while female doctors are portrayed in stereotypical nurturing roles. This can perpetuate existing societal biases and lead to discriminatory outputs (Perez 2020).

To mitigate this issue, researchers are exploring several techniques. Data augmentation involves artificially increasing the diversity of the training data (Shorten and Khoshgoftaar 2019). This can be achieved by techniques like oversampling underrepresented groups, generating synthetic data that reflects missing demographics, or employing data mixing which combines data from various sources. Bias detection algorithms are another approach. These algorithms can scan the training data for potential biases, such as skewed word usage or correlations between certain attributes and specific outcomes. By identifying these biases before training, we can take steps to mitigate

their impact. Additionally, researchers are exploring fair learning techniques that explicitly aim to train models that are robust to biases in the data. These techniques may include modifying the training to penalize biased outputs or incorporating fairness constraints into the learning process. It's important to note that data bias is a complex issue, and a combination of these approaches might be necessary to achieve truly fair and unbiased GenAI models.

3.2 Evaluation Metrics

Currently, there's a significant gap in the realm of robust and standardized metrics to accurately assess the factual correctness and coherence of generated content. Existing metrics like Inception Score (IS) and Frechet Inception Distance (FID) primarily focus on the visual quality and realism of generated images, offering limited insight into factual grounding or adherence to real-world data (Barratt and Sharma 2018). Similarly, language-based metrics like BLEU score prioritize fluency and similarity to the training data, potentially overlooking factual inconsistencies or internal contradictions.

To address this shortcoming, researchers are actively exploring new evaluation methods that encompass a broader range of criteria. One promising approach involves leveraging factual knowledge bases to assess the alignment of generated content with established truths. Additionally, researchers are developing metrics that measure internal consistency within the generated content itself. This could involve analyzing the coherence of arguments presented in text or the logical flow of a narrative. Furthermore, efforts are underway to incorporate human evaluation into the metric development process (Gao et al. 2024). By involving human experts to judge the factual accuracy, coherence, and overall quality of generated content, researchers can create more comprehensive evaluation methods that reflect real-world considerations. These advancements in evaluation metrics will be crucial for ensuring that GenAI outputs are not only aesthetically pleasing or fluent, but also factually grounded, internally consistent, and aligned with the intended purpose.

3.3 Explainability of Outputs

Explainability of outputs is a critical challenge in GenAI. Unlike traditional AI models that classify existing data points, GenAI creates entirely new content, making it difficult to understand the reasoning behind its outputs (Zhao et al. 2024). This lack of transparency poses significant hurdles:

Debugging and Error Correction: If a generated output is factually incorrect or misleading, pinpointing the root cause within the complex internal workings of the model

can be a challenge. Without explainability tools, debugging becomes a time-consuming and resource-intensive process. Imagine a model tasked with generating weather forecasts. An unexplainable prediction of a sudden snowstorm in July could be due to various issues – a bias in the training data towards historical anomalies, a malfunction in the model's internal logic, or even a simple data entry error. Explainability techniques can help isolate the issue, allowing developers to fix the model and prevent similar errors in the future.

Bias Detection: As mentioned earlier, biases present in training data can be amplified in generated content. Explainability methods can help identify which parts of the training data or model architecture contributed to a biased output, allowing for targeted interventions. For instance, an explainability tool might reveal that a model generating job descriptions consistently assigns leadership qualities to male-oriented terms and nurturing qualities to female-oriented terms (Shetty et al. 2022). This insight allows developers to adjust the training data or the model architecture to mitigate such biases.

Trust and User Confidence: When users don't understand how a model arrives at its outputs, trust diminishes. Explainable AI techniques can shed light on the model's decision-making process, fostering trust and user confidence in the generated content (Ferrario and Loi 2022). Imagine a designer using a GenAI tool to create marketing materials. Without explainability, the designer might be hesitant to trust the model's suggestions, fearing hidden biases or nonsensical reasoning. Explainability tools can show the designer why the model recommends specific visuals or phrases, building trust and allowing the designer to make informed decisions.

Beyond these core benefits, explainability can unlock further advantages. By understanding a model's reasoning, humans can collaborate with GenAI systems more effectively (Theis et al. 2023). Imagine a scientist using GenAI to generate research hypotheses. Explainability tools can help the scientist understand the rationale behind the model's suggestions, leading to a more productive partnership where human expertise refines the AI's creativity. Explainability can also guide the development of better GenAI models. Analyzing how models arrive at outputs can reveal weaknesses in their architecture or limitations in their training data (Ali et al. 2023). This knowledge can then be used to refine future iterations of the model, leading to more accurate and reliable outputs.

Several promising approaches are being explored to address explainability in GenAI:

Attention Mechanisms: These techniques highlight the specific parts of the input data that the model focused on when generating the output (Chefer et al. 2021). This can provide insights into the model's reasoning process and

identify potential biases based on which data points were emphasized.

Counterfactual Explanations: These methods explore alternative scenarios where slight changes to the input data would result in different outputs. By analyzing these “what-if” scenarios, we can gain a better understanding of how the model arrives at its results (Wachter et al. 2017).

Gradient-based Explanation Techniques: These approaches analyze the gradients of the model’s output with respect to the input data. This allows us to understand how changes in the input data influence the final output, providing insights into the model’s sensitivity to specific features (Selvaraju et al. 2020).

As research in explainable AI progresses, we can expect even more sophisticated techniques to emerge, paving the way for a future where GenAI operates with greater transparency and fosters a more trusting and productive relationship with its human users.

In conclusion, GenAI presents a powerful tool for the workplace. However, responsible use requires a multi-faceted approach that addresses technical limitations and fosters a human-centric environment. By collaborating between computer scientists, social scientists, ethicists, and business leaders, we can harness the potential of GenAI while mitigating risks and ensuring it benefits both businesses and individuals. As the field of computer science continues to develop solutions for data bias, evaluation metrics, and model interpretability, GenAI can become a reliable and trustworthy partner in the evolving world of work.

4 Work Design Changes and Responsible AI Use – Insights from Business Psychology

Sandra Ohly, Caroline Knight

4.1 Effects and Unintended Side Effects

AI has been reshaping the nature of work for some time, yet there is an ongoing debate about the positive and negative impacts on work design (Parker and Grote 2022). Here, we focus on two aspects of work design which seem particularly affected by AI, including newer forms such as GenAI: (1) job complexity; and (2) relational work characteristics.

4.1.1 AI and Job Complexity

GenAI has the potential to lead to a new form of human-AI division of labor (Noy and Zhang 2023), with GenAI producing a first draft of a piece of work (e.g., a report,

email, diagram) and humans being responsible for reviewing, revising, and developing the work.

In current workplaces, such as customer service, human-AI collaboration is wide-spread: (non-generative) AI is responsible for performing tasks such as responding to standard customer questions, while humans are responsible for tasks that require problem-solving, for example, responding to idiosyncratic customer requests where standard responses are not readily available. It has been argued that this human-AI division of labor is likely to occur not only in customer service, but also in medical diagnosis and recruitment (Jia et al. 2024). With the increasing availability of GenAI, this form of human-AI division of labor is becoming more likely, and could lead to enriched jobs that are characterized by higher problem-solving demands, more job complexity, and overall more information processing (Humphrey et al. 2007). Job complexity is the degree to which a job requires multiple high level skills, deep thought and information processing, rendering the job difficult to carry out. In addition to providing stimulating and motivating work tasks, work characterized by high job complexity might be experienced as overwhelming and mentally taxing (Elsbach and Hargadon 2006), leading to stress and burnout (Humphrey et al. 2007). Routine tasks (also called mindless tasks, Elsbach and Hargadon 2006)) are a way to mentally take a break while working as they can be accomplished automatically, without thinking or making conscious decisions (Ohly et al. 2006, 2017). Consequently, reducing routine tasks or eliminating them altogether through the implementation of AI could mean greater mental workload without the opportunity to take mental breaks, raising the question of how individuals can counteract the increased mental demand. Granting autonomy (e.g., deciding on whether to take a short break before talking to the next customer) might be relevant here.

With the advent of GenAI, a greater variety of tasks can be outsourced which are less routine and more complex in nature. For example, GenAI can be used to support the stimulation of new ideas and the creation of outputs such as text, computer code, data analyses, and art and design. To be effective, workers need to trust the AI, perceive it to be helpful and appropriate, and learn how to interact with it effectively to generate useful outputs (Chowdhury et al. 2022), which might explain why individuals do not see the benefits when having limited experience and no feedback (Noy and Zhang 2023). While this may make some previously required skills and tasks redundant, new skills will need to be learned meaning that the overall complexity of the job will not necessarily change, but the nature of it will. Demands on the worker may increase while this new learning takes place (Verma and Singh 2022).

4.1.2 AI and Relational Work Characteristics

Besides changing job complexity, the adoption of AI can help or hinder the relational quality of work. Relational work design refers to the social connections and interactions that individuals engage in with others at work, the relationships they form, and the networks they build (Grant and Parker 2009). On the one hand, if humans are no longer required to interact with customers, clients, and colleagues as much, for example, because AI is able to respond to queries, this could decrease workers' sense of connection with others and the value they see in their work, reducing the satisfaction of the basic human need for belonging (Ryan and Deci 2000). Further, individuals could lose important social skills and the ability to collaborate with others, impeding team co-ordination and collaboration. The reduction in social contact could also hamper learning and network establishment, promoting professional isolation and stagnating career progression (Yang et al. 2022). For example, relying on AI for medical diagnoses could mean that a medical practitioner no longer needs to confer with a multidisciplinary team so regularly, reducing the opportunity for incidental learning and collaboration. More generally, an erosion in the relational quality of work is likely to contribute to the growing loneliness crisis exacerbated by the increasing trend for virtual work (Knight et al. 2022).

On the other hand, AI can be used to improve the relational quality of work. For distributed workforces, information communication technologies can improve co-ordination and collaboration across space and time (Kellogg et al. 2006). Yet, while technologies such as instant messaging and video conferencing encourage connectivity, it can be difficult to convey tone, meaning and empathy, and to pick up on social cues, which can hinder relationship building and foster a sense of isolation (Hill et al. 2022). This suggests that how and when AI is used is critical to consider when advancing the use of AI in workplaces. Although some might view AI as a potential teammate of humans (Seeber et al. 2020; Georganta and Ulfert 2024), it is currently unclear whether this perspective on human-AI interaction is valid, given the inherent limitations of AI decision-making ability (see below), or if this rather represents a case of anthropomorphism, attributing human attributes to AI.

4.2 What Shapes the Decision to Adopt AI?

The decision to adopt AI may be shaped by both top-down, manager-led formal processes, and bottom-up, individual-led, informal processes (Parker et al. 2017). Regarding the former, and from a strategic perspective, organizations often design jobs that are most conducive to organizational

objectives, including effective and efficient performance, but may view employee wellbeing as a secondary aim, which may lead to job design that is less conducive to optimizing both outcomes.

From a work psychology perspective, in addition to formal decision-making processes on job design, emergent, bottom-up work design processes such as job crafting can shape the nature of work tasks (Parker et al. 2017). For example, employees may craft their work tasks to either reduce demands such as high workload or enhance resources such as job autonomy, with effects on outcomes such as work engagement and work performance (Rudolph et al. 2017; Oprea et al. 2019), or even organizational financial gains such as reduced labor costs (Oprea et al. 2019). Emerging research suggests that employees might voluntarily use AI to craft their jobs, depending on the context (Cheng et al. 2023). In addition, they could also use GenAI to improve work meaningfulness, or relatedness with others (Newman et al. 2014). For example, employees might use GenAI to create better results by saving time on mundane tasks and concentrating on problem-solving, in order to please customers, to shape the relationships with their co-workers and avoid conflict by communicating more effectively or more empathetically.

In contrast, there is a possibility that employees may also craft their jobs away from using AI if they do not understand or trust it or have not been trained in how to effectively use it. For a comprehensive assessment of how work will be changed by the adoption of AI, these emergent work design processes and their underlying motives will need to be considered in addition to formal processes.

4.3 Conclusion

To conclude, AI in general, as well as novel approaches such as GenAI, offer the potential to enrich or erode the quality of jobs. Different industries and occupations will be impacted in very different ways depending on the type and extent of AI introduced. It will be important for AI developers to work with managers and organisations who design work, as well as with workers themselves, to ensure that human-AI interactions are fit for purpose, ethical and responsible. The responsible implementation and management of AI in the work context needs to consider the changes that will occur in individuals' work design, with managers and employees proactively shaping jobs to promote optimal outcomes for workers and organisations (Parker and Grote 2022).

5 Protection of Fundamental Rights, Existing Data Protection Law and the New Risk-Based AI Act – Insights from Law

Lena Rudkowski, Domenik Wendt

From a legal perspective, awareness of the fundamental rights of users characterizes the responsible use of new technologies in the workplace, including the use of GenAI systems.

Responsible use of technical systems, of any kind, preserves the fundamental rights of those who interact with them. Safeguarding human dignity is a core commitment of our legal system (Art. 1 para. 1 GG). The processing of personal data is always an encroachment on the general personal rights of the person concerned (Art. 2 para. 1, Art. 1 para. 1 GG). Both legal positions are particularly affected by technical systems.

The German Basic Law (GG) assumes the self-determination of the individual. The individual himself is not to be regarded as an object of state action, this also applies to the actions of other private individuals. If people are degraded to the status of objects, if they are dehumanized or if their human dignity is violated, the state has a duty to protect them and must intervene to preserve the individual's quality as a subject, for instance, through legislation. The general right of personality, which also includes the right to informational self-determination – “invented” by the German Federal Constitutional Court (BVerfG) on the occasion of a state census – is also based on human dignity.

As a special guarantee, the right to informational self-determination protects against the collection and processing of data by the state and by private individuals. If a private individual has knowledge about another person, such as an employer about their employee, this does not initially cause any harm – data processing does not cost the person concerned anything and does not hurt them. However, the information regularly becomes the basis for further measures. Anyone who is “transparent” is controllable. Entire areas of law (such as corporate reporting) are based on this idea. At the same time, those who are being observed are biased, thus control creates a balance of power. Therefore, everyone should decide for themselves how much they want to reveal personally – with the awareness of becoming tangible and perhaps also vulnerable, or at least controllable.

Indirectly, the protection of fundamental rights for privacy of personal data therefore serves to protect the individual from being controlled by others. Informational self-determination is protected for the sake of actual, real self-determination. Those who can determine what others find out about them can also determine the extent to which they influence them. Data protection is therefore the basis for protecting one's own self-determination.

The opportunities that AI offers for the workplace are enormous: regarding demographic developments, it can replace workers who would not be available anyway and thus redirect them towards jobs where they are needed. It can relieve the physical and mental strain on employees by taking over heavy tasks, controlling and monitoring complex operations and instructing employees. It can reduce sources of error and simultaneously speed up, clarify and simplify operational processes. This also applies to activities in the legal field (Block et al. 2023).

However, new technologies and AI in particular also have the potential to dehumanize people – either because they become mere “data suppliers” or because they are controlled by technology as they are subordinate to it because it is (supposedly) objective and infallible. From a legal perspective, this should be avoided. However, jurisprudence has not yet determined exactly where the limits lie. Is there an indispensable “core of fundamental rights” that is not subject to the discretion of the entitled person, or does this result in patronism? How should this be determined? Clarifying this question in a broad interdisciplinary and social discourse will be the task of legal scholars in the coming years.

From a labor law perspective, the potential for employee monitoring plays a particularly important role in the evaluation of AI, even if monitoring is not the aim of the measure. For example, a cobot (collaborating robot) must record the employees' movements around it so as not to injure anyone (e.g., Ebert et al. 2023). This creates a movement image of its human colleagues. An intelligent measurement software can determine exactly which worker has made incorrect adjustments and when. Handheld scanners form the basis for fair pay, because they accurately record the performance of each employee, and for perfect scheduling by the employer. However, it must be considered that scanners monitor every move the employee makes and can put pressure on them through performance evaluations.

In order to take advantage of the opportunities offered by the new technologies, too much regulation should not be imposed immediately. Contrary to fears in legal literature and the European Parliament technological development is still far away from the “robot boss”, who uses and instructs employees like tools in an unemotional and sober manner. It is important to wait and see and only counteract specific undesirable developments if the legal situation does not already provide sufficient mechanisms *de lege lata*.

However, this is regularly the case at present: the General Data Protection Regulation (GDPR), which regulates the processing of personal data, offers sufficient protection for employees in conjunction with national data protection law. It meets the requirements set out in

fundamental rights, at least in the area of employment. “Total surveillance” is already inadmissible under current law.

In addition to data protection limits, the product safety requirements of the new AI Act by the European Union (EU) must also be observed in future. The EU AI Act lays down uniform EU-wide requirements for the production and use of AI systems (Evas 2024). The directly applicable EU regulation follows a risk-based regulatory approach that categorizes AI systems according to their risk to the health, safety and fundamental rights of EU citizens (Wendt and Wendt 2024). The EU AI Act is the world’s first comprehensive legal framework regulating the use of AI systems. In addition to product safety requirements for AI systems, prohibitions are regulated, and a new supervisory system is established. Particularly extensive requirements are placed on so-called high-risk AI systems. These include AI systems that are related to employment, personnel management and access to self-employment. Regarding GenAI systems, the specific requirements for so-called general purpose AI models and systems must also be taken into account. For general purpose AI models and systems (AI models and systems that have a wide range of possible uses) which are also relevant in the employment context, a tiered regulatory concept is envisaged depending on the systemic risk (Wendt and Wendt 2024). In addition to transparency and documentation obligations, additional requirements in the areas of cybersecurity and risk management are planned for more powerful general-purpose AI models. Violations of the requirements of the EU AI Act – as in the GDPR – are subject to severe sanctions based on the company’s annual turnover or fines of up to EUR 35 million.

The GDPR and the EU AI Act have therefore already formulated important requirements and limits that must be observed when using AI systems in the workplace. One of the tasks of legal research over the next few years will be to remove any ambiguities here, but above all it will be important to determine in a broad scientific and social discourse exactly where the fundamental legal boundaries are to be drawn – here the legal situation interacts with actual technical and social developments. A social discussion about transparency and self-determination is needed, which is reflected in science through interdisciplinary research in law, philosophy, social sciences and psychology.

Interdisciplinary cooperation, in close collaboration with computer science and information systems, must create awareness of data protection requirements in the practice of software development. Data protection and product security “by design” is the best way to achieve well-balanced regulatory boundaries and requires close cooperation between law, psychology, philosophy and social sciences

on the one hand and computer science as well as information systems on the other.

6 Beyond Code – The Crucial Role of Responsibility in Technology Ethics

Gerhard Schreiber

Technology ethics explores the ethical implications of technology in the complexity of its interrelationships in the areas of production, implementation, and consequences. Technology is never considered in isolation, but always in a network of interrelationships that go far beyond its immediate function. By analyzing and evaluating these eco-socio-technical interrelationships from an ethical perspective, technology ethics makes an indispensable contribution to understanding how technology, as a “form of life” (Winner 1983), forms our lives and is in turn formed by our social structures and value systems. This is particularly evident in the light of the increasing datafication of our existence – the translation of virtually every aspect of human life into a format that can be processed, analyzed, and monetized by algorithms. This makes the question of responsibility indispensable.

6.1 Responsibility as Response-Ability

In the context of technology ethics, responsibility does not refer to a simple cause-and-effect relationship, but rather to a normative relationship that has both a prospective and a retrospective side (Werner 2021). This two-sidedness is, in a sense, inherent in the concept of responsibility itself. Responsibility, understood as “response-ability” (Perles 1971, p. 30), is always also a response to something that happens to us, an engagement with something that affects us (Buber 1947, p. 16). Thus, we are responsible not only for the actions we initiate as authors, but also for how we respond to situations and circumstances that unfold around us, recognizing that *any* response we make may itself carry significant weight and may have far-reaching consequences.

6.2 Key Aspects of Assessing (Ir)Responsibility

When considering the responsible use of new information and communication technologies in general, and GenAI systems in particular, two key aspects can be highlighted in the form of questions. These serve as guidelines for assessing what constitutes a responsible use of technology and what marks the boundary between responsible and irresponsible use from a technoeethical perspective.

Firstly, it is crucial to ask *in what way* and *to what extent* the technology in question affects human agency

(Schreiber 2024), encompassing capabilities, skills, and competencies. It's not only a question of *how* this impact manifests itself, whether as an expansion or a reduction of human autonomy, but also of *the degree to which* it occurs or should occur. The latter includes consideration of whether human authorship, and thus (the assumption of) responsibility, is, could, or should be partially or completely delegated to software systems (German Ethics Council 2023, pp. 164–186). These dimensions of quality (*in what way*) and quantity (*to what extent*) are closely intertwined but need to be distinguished, both with respect to the users and with respect to those affected by the technology in question.

Any delegation of (previously) genuinely human activities and actions to digital technologies and AI-based systems is therefore subject to the ethical reservation that these technologies not only open up new possibilities of constructive, beneficial use, but also always open up new possibilities of (their) destructive misuse: to control, mislead, restrict, and suppress humans and their potential opportunities for individual development and realization. Thus, along with the first question about the nature and degree of influence on human authorship, we must simultaneously – and *secondly* – ask about the gain in human freedom promised or hoped for in the technology in question, which threatens to turn out to be a self-incurred loss of maturity in the Kantian sense (Kant 1991, p. 54): no longer being able to use one's own understanding without the guidance of artificial neural networks. A digital revolution in the way of thinking. *Sapere aude!*

6.3 Talking with Adorno

Reflecting on a thought almost prophetically expressed by Theodor Adorno in the 1960s regarding the interaction with “cybernetic machines,” the humane significance of AI-based systems can indeed be perceived in their capacity to alleviate “the thinking of living beings” in such a manner “that thought would gain the freedom to attain a knowledge that is not already implicit” (Adorno 1998, pp. 127f.). Adorno’s optimistic vision reflects the transformative potential of technology to free human cognition from the drudgery of routine tasks and to create space for independent thinking, which, according to Adorno, can be regarded as philosophical “as soon as it ceases to content itself with cognitions that are predictable and from which nothing more emerges than what had been placed there beforehand” (Adorno 1998, p. 128).

However, the datafication of our existence inevitably presents a paradox. While intended to enhance our freedom, technological innovations can inadvertently construct a maze of constraints in which the digital freedom gained, whether consciously or not, comes at the cost of real-world

unfreedom. Applying Adorno’s vision to our current society, where digital visions become realities and digital realities become visions, thus highlights a critical point: embracing the emancipatory potential of GenAI systems requires a careful assessment of how these technologies might interfere with our actual freedoms. This juxtaposition serves as a sobering reminder of the trade-offs inherent in our march toward the digital transformation of almost all areas of human work and life, urging us to find a balance between technological development and the preservation of fundamental human freedoms.

Such ethical considerations may seem like “a bicycle brake on an intercontinental airplane” (Beck 1988, p. 194), perceived as foolish quixotry, yet they represent an expression of realistic insight into the all-pervasive ambivalence of human existence.

6.4 Non-Delegability of Responsibility and Decision-Making Authority

The two key aspects discussed above also serve as guidelines for the responsible use of GenAI systems in the workplace 4.0 (Mütze-Niewöhner and Nitsch 2019). Despite the undeniable benefits that GenAI systems offer, it is imperative to remember that their use becomes irresponsible not at the point where it leads to violations of the principle of data sovereignty (Gehring and Augsberg 2022) or when these systems inflict psychological, social, or economic harm on individuals based on (blind reliance on) unconsciously biased or intentionally manipulated, hence “toxic” data (Schreiber 2022). The use of GenAI systems already reaches the limits of irresponsibility when it restricts and corrupts the freedom of human decision-making processes by promoting excessive dependence on these systems.

The attribution of decision-making ability to AI systems is based on a functional perspective, in which algorithms are often consciously or unconsciously anthropomorphized as “decision makers” based on their ability to process input data and produce an output. However, this view tends to oversimplify the decision-making process by ignoring the deeper cognitive and emotional aspects of human decision-making. While human decision makers are able to flexibly incorporate a variety of nuances and contextual information into their deliberations, algorithms are limited to the data provided to them and follow predefined rules without having their own “understanding” of the consequences of their “decisions”. To put it bluntly: AI systems are not only incapable of making more consistent or neutral decisions than humans; they cannot make decisions *at all*. As a result of this insurmountable gap between algorithms and the more comprehensive, contextual, and intentional nature of human decision-making, the use of terms such as “output,”

“result,” “deduction,” or “conclusion” in the context of algorithmic processes would be more appropriate than the term “decision.” Human beings cannot afford to lose the freedom and responsibility for decision-making with which they are entrusted.

6.5 The Human Factor

While delegating repetitive and error-prone tasks to GenAI systems can be useful and efficient in many work settings (Schreiber and Ohly 2024), delegating decision-making responsibility to these systems remains counterintuitive. Human decision-making is not just a process of data processing; it involves participation, engagement based on bodily experience, and a deep understanding of contextual factors and nuances. These dimensions of human experience, which enable action beyond conscious planning and calculation, remain beyond the reach of AI systems. This brings us to the fundamental limits of the formalizability and simulability of human reason (German Ethics Council 2023, p. 29). While recognizing the potential of technological advances to shift or transform these limits, it remains critical to emphasize the irreplaceable value of human insight and oversight in contexts where ethical considerations, emotional intelligence, and moral judgment are indispensable. AI systems, as useful as they may be as tools to enhance and extend human capabilities, can *support* human decision-making by analyzing, filtering, categorizing, prioritizing, summarizing, or visualizing information, but they cannot make decisions or take responsibility themselves – in either a normative or a descriptive sense. They “act”, if the concept of action is applied to machines, “irresponsibly”, i.e., without even a glimpse of responsibility. This is also critical because humans tend to follow AI recommendations or may find it challenging to argue against AI suggestions, especially when they lack task knowledge or perceive AI as more capable (Vodrahalli et al. 2022). This overreliance can persist even when AI suggestions are incorrect (Buçinca et al. 2021). Therefore, even when humans appear to be the final decision-makers, this may not be the case in practice, as their decisions may be significantly influenced or corrupted by AI systems.

The responsible use of GenAI systems requires not only technological expertise and interdisciplinary collaboration between computer science, business informatics, the humanities, social sciences, and law, but also profound ethical reflection and a clear delineation of the role of machines in the decision-making process. There is no doubt that human decision-making and taking responsibility are inherently fraught with uncertainty, imponderability, and susceptibility to error. And yet: These capabilities, reflecting *both* human strengths and

fallibilities, must be protected in an increasingly automated world of work.

7 Moving Forward – A Sociotechnical Perspective on Future Research on the Responsible Use of Generative AI in the Workplace

Sandra Ohly, Gerhard Schreiber, Matthias Söllner

In this section, we seek to provide a sociotechnical perspective on research challenges that need to be answered to foster responsible use of GenAI in the workplace. While the previous sections of the discussion took a disciplinary view of our discussion topic, this concluding section seeks to provide an interdisciplinary perspective, highlighting open research opportunities that call for collaboration across fields. By considering diverse viewpoints on sociotechnical systems, we can build a holistic understanding of how GenAI can be responsibly and effectively integrated into workplace settings, balancing both technical and social considerations to meet organizational and human needs.

Given the variety of perspectives on sociotechnical systems, we first outline our perspective on sociotechnical systems, which draws upon works such as Mumford (2006), Lee et al. (2015) and Sarker et al. (2019). Consequently, we view the technical and social components of sociotechnical systems as interdependent and of equal importance (Bostrom et al. 2009). The technical component consists of data, software, hardware and the necessary techniques to complete tasks in the workplace (Ryan et al. 2002). In turn, the social components include individuals or groups, and their interactions with each other through which they attempt to either solve problems, achieve goals, or serve purposes (Lee et al. 2015). This attitude reflects the need for a human-centered approach (Shneiderman 2022) that prioritizes dignified work and employee well-being.

7.1 Human-Centered AI

In addition, a human-centered approach in AI design is emphasized in the EU AI Act, underscoring the alignment of regulatory frameworks with sociotechnical values. In line with this regulation, we propose that the deployment of GenAI in the workplace should go beyond efficiency and productivity goals to address broader ethical, social, and legal dimensions, including the protection of human dignity, autonomy, and privacy – values foundational to the sociotechnical approach (Sloane 2019; Floridi and Cowls 2019). This approach ensures that AI serves as an empowering tool for employees, enhancing their capabilities rather than constraining them. Consequently, achieving

an optimal fit between the technical and social components is crucial, since this is expected to result in improved instrumental outcomes, such as higher productivity, and better human outcomes, such as job satisfaction and employee well-being (Wallace et al. 2004; Nathan 2022). Furthermore, a human-centered approach emphasizes that AI-powered systems should be designed to be transparent and understandable, allowing users to interpret and, if necessary, challenge and override AI decisions.

7.2 Adapting System Design to User Needs and Organizational Processes

From a sociotechnical perspective, system design must account for both the technical and the social subsystem (Bostrom et al. 2009). The design of a system should be adapted to the processes of the organization and the needs of the users to best support the achievement of objectives. If the social subsystem is insufficiently considered in the design of the technology, there is a risk that the organization's goals will not be achieved. In fact, careless design decisions in the area of the technical subsystem could lead to unintended and undesired consequences in the social subsystem. Consequently, following the logic of technochange (Markus 2004), design decisions need to relate to both subsystems during change processes to ensure the achievement of desired outcomes. Such a comprehensive human-centered design also helps to preserve and even has the potential to enhance employees' work processes, well-being, and professional identity.

7.3 A Holistic Consideration of Potential AI Consequences

The sociotechnical systems perspective also implies a holistic consideration of possible consequences of AI, including decision quality, efficiency and productivity, but also user satisfaction and employee well-being, including the analyses when these outcomes may conflict with each other. The EU AI Act further mandates that, in high-risk scenarios involving hybrid decision systems, human beings must retain authority to accept or revise AI recommendations. This requirement ensures that technology remains a supportive tool, preserving users' decision-making capabilities and critical judgment. To ensure decision quality, it is essential to support humans in their decision-making ability and reliably identify the cases where AI might be wrong or insufficient, to prevent overreliance on AI. At the same time, it is also important to empower humans to understand in which situations they should rely on AI recommendations to avoid the so-called *Verschlimmbesserung*, which can be observed when humans

revise correct AI recommendations, and thus impair decision quality (Wardlaw et al. 2022).

7.4 Generative AI Literacy as Enabler of Positive Change

The foundation for avoiding the abovementioned challenges and achieving the outlined goals lies in whether all stakeholders involved can be empowered to build up the necessary GenAI literacy – which includes updating their skills on a regular basis (Pinski and Benlian 2024; Pinski et al. 2024). Managers need to understand the potential benefits and limitations of GenAI systems to take well-informed business decisions. Employees need to understand where GenAI systems can help them to do their jobs better, and how they can stay competitive in the job market in the long term, for example through vocational education and training. Finally, designers need to understand not only the technical capabilities of GenAI, but also how the application of GenAI systems might shape job profiles and organizational processes to ensure that the technical decisions they take are in line with the organization's strategy and the needs of all stakeholders involved. While each group of stakeholders mentioned here faces different challenges, most likely resulting in different needs when it comes to GenAI literacy, it becomes apparent that they need to collaborate to understand each other's perspectives in order to consider everybody's needs during their own decision-making.

Taking a sociotechnical approach not only provides a robust framework for addressing the complexities of GenAI in the workplace, but also opens new avenues for future research on responsible implementation. By considering both human and organizational needs, this perspective highlights critical areas for further investigation. Looking to the future, this approach encourages continued research to ensure that GenAI is developed and applied in ways that promote a balanced, responsible, and human-centered workplace.

Acknowledgements The authors would like to thank the Hessian Centre Responsible Digitality for funding the interdisciplinary project group on “Responsible Algorithmic Decision-Making in the Workplace.” The authors would furthermore like to thank Martina Block, Lucie Flek, Iryna Gurevych, Allison Lahnalala, Marc Pinski, Didem Sedefoglu, Bernd Skiera, and Pablo Weingart for their valuable input to the project groups work, as well as Klaus Angerer for his support in coordinating the activities of the project group.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as

long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adam M, Roethke K, Benlian A (2023) Human vs. automated sales agents: how and why customer responses shift across sales stages. *Inf Syst Res* 34(3):1148–1168. <https://doi.org/10.1287/isre.2022.1171>
- Adorno T (1998) Notes on philosophical thinking (originally “Anmerkungen zum philosophischen Denken”; lecture for Deutschlandfunk on October 9, 1964). Critical models: interventions and catchwords. Columbia University Press, New York, pp 127–134
- Akata Z, Balliet D, De Rijke M, Dignum F, Dignum V, Eiben G, Fokkens A, Grossi D, Hindriks K, Hoos H et al (2020) A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer* 53(8):18–28
- Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, Guidotti R, Del Ser J, Díaz-Rodríguez N, Herrera F (2023) Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf Fusion* 99:101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- Baird A, Maruping LM (2021) The next generation of research on IS use: a theoretical framework of delegation to and from agentic IS artifacts. *MIS Q* 45(1):315–341
- Banh L, Strobel G (2023) Generative artificial intelligence. *Electron Mark* 33(1):63. <https://doi.org/10.1007/s12525-023-00680-1>
- Barratt S, Sharma R (2018) A note on the inception score. <https://doi.org/10.48550/arXiv.1801.01973>
- Beck U (1988) *Gegengifte*. Suhrkamp, Frankfurt
- Benlian A, Wiener M, Cram WA, Krasnova H, Maedche A, Möhlmann M, Recker J, Remus U (2022) Algorithmic management. *Bus Inf Syst Eng* 64(6):825–839. <https://doi.org/10.1007/s12599-022-00764-w>
- Berger B, Adam M, Rühr A, Benlian A (2021) Watch me improve – algorithm aversion and demonstrating the ability to learn. *Bus Inf Syst Eng* 63(1):55–68. <https://doi.org/10.1007/s12599-020-00678-5>
- Bhaimiya S (2023) ChatGPT may be the fastest-growing consumer app in internet history, reaching 100 million users in just over 2 months, UBS report says. In: business insider. <https://www.businessinsider.com/chatgpt-may-be-fastest-growing-app-in-history-ubs-study-2023-2>. Accessed 6 May 2024
- Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE (2023) High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus* 15(5):e39238. <https://doi.org/10.7759/cureus.39238>
- Birhane A, Prabhu VU, Kahembwe E (2021) Multimodal datasets: misogyny, pornography, and malignant stereotypes. <https://doi.org/10.48550/arXiv.2110.01963>
- Block M, Jung C, Wendt D (2023) Requirements for “Legal tech AI systems” – reflections on the negotiated AI act with regard to legal technology using AI. *Comput Law Rev Int* 24(4):97–105. <https://doi.org/10.9785/cri-2023-240402>
- Bostrom RP, Gupta S, Thomas D (2009) A meta-theory for understanding information systems within sociotechnical systems. *J Manag Inf Syst* 26(1):17–48
- Bouschery SG, Blazevic V, Piller FT (2023) Augmenting human innovation teams with artificial intelligence: EXPLORING transformer-based language models. *J Prod Innov Manag* 40(2):139–153
- Buber M (1947) Between man and man, trans. RG Smith, Kegan Paul, London
- Buçinca Z, Malaya MB, Gajos KZ (2021) To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proc ACM Hum-Comput Interact* 5(CSCW1):1–21. <https://doi.org/10.1145/3449287>
- Chefer H, Gur S, Wolf L (2021) Generic attention-model explainability for interpreting bi-modal and encoder-decoder transformers. In: 2021 IEEE/CVF international conference on computer vision. IEEE, pp 387–396
- Cheng B, Lin H, Kong Y (2023) Challenge or hindrance? How and when organizational artificial intelligence adoption influences employee job crafting. *J Bus Res* 164:113987. <https://doi.org/10.1016/j.jbusres.2023.113987>
- Chowdhury S, Budhwar P, Dey PK, Joel-Edgar S, Abadie A (2022) AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *J Bus Res* 144:31–49. <https://doi.org/10.1016/j.jbusres.2022.01.069>
- Chui M, Hazan E, Roberts R, Singla A, Smaje K, Sukharevsky A, Yee L, Zemmel R (2023) The economic potential of generative AI. In: McKinsey. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>. Accessed 7 May 2024
- Cram W, Arcy J, Benlian A (2024) Time will tell: the case for an idiographic approach to behavioral cybersecurity research. *MIS Q* 48:95–136. <https://doi.org/10.25300/MISQ/2023.17707>
- Ebert A, Busch P, Spiecker genannt Dohmann I, Wendt J (2023) Roboter im supermarket - Eine Betrachtung haftungs- und datenschutzrechtlicher Probleme beim Einsatz von Robotern im Einzelhandel. *Z Prod Compliance (ZPC)* 1:16–29
- Elsbach KD, Hargadon AB (2006) Enhancing creativity through “mindless” work: a framework of workday design. *Organ Sci* 17(4):470–483. <https://doi.org/10.1287/orsc.1060.0193>
- Evas T (2024) The EU artificial intelligence act: advancing innovation for trustworthy AI. *J AI Law Regul* 1(1):98–101. <https://doi.org/10.21552/aire/2024/1/1>
- Fabri L, Häckel B, Oberländer AM, Rieg M, Stohr A (2023) Disentangling human-AI hybrids. *Bus Inf Syst Eng* 65(6):623–641. <https://doi.org/10.1007/s12599-023-00810-1>
- Ferrario A, Loi M (2022) How explainability contributes to trust in AI. In: 2022 ACM conference on fairness, accountability, and transparency. ACM, Seoul, pp. 1457–1466
- Fuerriegel S, Hartmann J, Janiesch C, Zschech P (2024) Generative AI. *Bus Inf Syst Eng* 66(1):111–126. <https://doi.org/10.1007/s12599-023-00834-7>
- Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. *Harv Data Sci Rev*. <https://doi.org/10.1162/99608f92.8cd550d1>
- Fügener A, Grahl J, Gupta A, Ketter W (2021) Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS Q* 45(3):1527–1556. <https://doi.org/10.25300/MISQ/2021.16553>
- Gao M, Hu X, Ruan J, Pu X, Wan X (2024) LLM-based NLG evaluation: current status and challenges. <https://doi.org/10.48550/arXiv.2402.01383>
- Gehring P, Augsberg S (eds.) (2022) Datensouveränität. Positionen zur Debatte. Campus, Frankfurt

- Georganta E, Ulfert A-S (2024) My colleague is an AI! Trust differences between AI and human teammates. *TPM* 30(1/2):23–37. <https://doi.org/10.1108/TPM-07-2023-0053>
- German Ethics Council (2023) Mensch und maschine – Herausforderungen durch künstliche Intelligenz [humans and machines – challenges of artificial intelligence]. Deutscher Ethikrat, Berlin
- Gownder JP, O’Grady M (2023) Generative AI will reshape far more jobs than it eliminates. <https://www.forrester.com/what-it-means/ep346-generative-ai-jobs-impact/>. Accessed 7 May 2024
- Grant AM, Parker SK (2009) 7 Redesigning work design theories: the rise of relational and proactive perspectives. *Annals* 3(1):317–375. <https://doi.org/10.5465/19416520903047327>
- Grynaum MM, Mac R (2023) The Times sues OpenAI and Microsoft over A.I. use of copyrighted work. In: The New York Times. <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>. Accessed 7 May 2024
- Hartmann J, Schwenzow J, Witte M (2023) The political ideology of conversational AI: converging evidence on ChatGPT’s pro-environmental, left-libertarian orientation. <https://doi.org/10.48550/arXiv.2301.01768>
- Hill NS, Axtell C, Raghuram S, Nurmi N (2022) Unpacking virtual work’s dual effects on employee well-being: an integrative review and future research agenda. *J Manag* 50(2):752–792. <https://doi.org/10.1177/01492063221131535>
- Humphrey SE, Nahrgang JD, Morgeson FP (2007) Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *J Appl Psychol* 92(5):1332–1356. <https://doi.org/10.1037/0021-9010.92.5.1332>
- Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P (2023) Survey of hallucination in natural language generation. *ACM Comput Surv* 55(12):1–38. <https://doi.org/10.1145/3571730>
- Jia N, Luo X, Fang Z, Liao C (2024) When and how artificial intelligence augments employee creativity. *Acad Manag J* 67(1):5–32. <https://doi.org/10.5465/amj.2022.0426>
- Kant I (1991) Answer the question: What is enlightenment? In: Reiss H (ed) Kant’s political writings, 2nd edn. Cambridge University Press, New York, pp 54–60
- Kecht C, Egger A, Kratsch W, Röglinger M (2023) Quantifying chatbots’ ability to learn business processes. *Inf Syst* 113(102):176. <https://doi.org/10.1016/j.is.2023.102176>
- Kellogg KC, Orlitzki WJ, Yates J (2006) Life in the trading zone: structuring coordination across boundaries in postbureaucratic organizations. *Organ Sci* 17(1):22–44. <https://doi.org/10.1287/orsc.1050.0157>
- Kelly J (2023) Goldman Sachs predicts 300 million jobs will be lost or degraded by artificial intelligence. <https://www.forbes.com/sites/jackkelly/2023/03/31/goldman-sachs-predicts-300-million-jobs-will-be-lost-or-degraded-by-artificial-intelligence/>. Accessed 6 May 2024
- Knight C, Olaru D, Lee JA, Parker SK (2022) The loneliness of the hybrid worker. In: MIT Sloan Management Reviews. <https://sloanreview.mit.edu/article/the-loneliness-of-the-hybrid-worker/>. Accessed 7 May 2024
- Lee AS, Thomas M, Baskerville RL (2015) Going back to basics in design science: from the information technology artifact to the information systems artifact. *Inf Syst J* 25(1):5–21. <https://doi.org/10.1111/isj.12054>
- Markus ML (2004) Technochange management: using IT to drive organizational change. *J Inf Technol* 19(1):4–20. <https://doi.org/10.1057/palgrave.jit.2000002>
- Mikalef P, Conboy K, Lundström JE, Popović A (2022) Thinking responsibly about responsible AI and ‘the dark side’ of AI. *Eur J Inf Syst* 31(3):257–268. <https://doi.org/10.1080/0960085X.2022.2026621>
- Mumford E (2006) The story of socio-technical design: reflections on its successes, failures and potential. *Inf Syst J* 16(4):317–342. <https://doi.org/10.1111/j.1365-2575.2006.00221.x>
- Mütze-Niewöhner S, Nitsch V (2019) Arbeitswelt 4.0. In: Frenz W (ed.) Handbuch Industrie 4.0: Recht, Technik, Gesellschaft. Springer, Heidelberg, pp 1187–1217
- Nathan G (2022) Meaningfulness at work: well-being, employeeship and dignity in the workplace. *SSRN J*. <https://doi.org/10.2139/ssrn.4101517>
- Newman DB, Tay L, Diener E (2014) Leisure and subjective well-being: a model of psychological mechanisms as mediating factors. *J Happiness Stud* 15:555–578. <https://doi.org/10.1007/s10902-013-9435-x>
- Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381(6654):187–192
- Ohly S, Sonnentag S, Pluntke F (2006) Routinization, work characteristics and their relationships with creative and proactive behaviors. *J Organ Behav* 27(3):257–279. <https://doi.org/10.1002/job.376>
- Ohly S, Göritz AS, Schmitt A (2017) The power of routinized task behavior for energy at work. *J Vocat Behav* 103:132–142. <https://doi.org/10.1016/j.jvb.2017.08.008>
- Oprea BT, Barzin L, Virgă D, Iliescu D, Rusu A (2019) Effectiveness of job crafting interventions: a meta-analysis and utility analysis. *Eur J Work Organ Psychol* 28(6):723–741. <https://doi.org/10.1080/1359432X.2019.1646728>
- Parker SK, Grote G (2022) Automation, algorithms, and beyond: why work design matters more than ever in a digital world. *Appl Psychol* 71(4):1171–1204. <https://doi.org/10.1111/apps.12241>
- Parker SK, Van Den Broeck A, Holman D (2017) Work design influences: a synthesis of multilevel factors that affect the design of jobs. *Annals* 11(1):267–308. <https://doi.org/10.5465/annals.2014.0054>
- Peng S, Kalliamvakou E, Cihon P, Demirer M (2023) The impact of AI on developer productivity: evidence from GitHub Copilot. <https://doi.org/10.48550/arXiv.2302.06590>
- Perez CC (2020) Invisible women: exposing data bias in a world designed for men. Vintage, London
- Perles FS (1971) Four lectures. In: Fagan J, Shepherd IL (eds) Gestalt therapy now: theory, techniques, applications. Harper & Row, New York, pp 14–38
- Pinski M, Benlian A (2024) AI literacy for users – a comprehensive review and future research directions of learning methods, components, and effects. *Comput Hum Behav Artif Hum* 2(1):100062. <https://doi.org/10.1016/j.chbah.2024.100062>
- Pinski M, Hofmann T, Benlian A (2024) AI Literacy for the top management: An upper echelons perspective on corporate AI orientation and implementation ability. *Electron Mark* 34(1):24. <https://doi.org/10.1007/s12525-024-00707-1>
- Reinhard P, Li MM, Peters C, Leimeister JM (2024) Generative AI in customer support services: a framework for augmenting the routines of frontline service employees. In: proceedings of the 57th Hawaii international conference on system sciences
- Rudolph CW, Katz IM, Lavigne KN, Zacher H (2017) Job crafting: A meta-analysis of relationships with individual differences, job characteristics, and work outcomes. *J Vocat Behav* 102:112–138. <https://doi.org/10.1016/j.jvb.2017.05.008>
- Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol* 55(1):68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan SD, Harrison DA, Schkade LL (2002) Information-technology investment decisions: when do costs and benefits in the social

- subsystem matter? J Manag Inf Syst 19(2):85–127. <https://doi.org/10.1080/07421222.2002.11045725>
- Sarker S, Chatterjee S, Xiao X, Elbanna A (2019) Says who? The effects of presentation format the sociotechnical axis of cohesion for the discipline: its historical legacy and its continued relevance. MIS Q 43(3):695–A5. <https://doi.org/10.25300/MISQ/2019/13747>
- Schreiber G (2022) Datentoxikalität: Eine technikethische Herausforderung. In: Augsberg S, Gehring P (eds) Datensouveränität: Positionen zur Debatte. Campus, Frankfurt a.M, pp 199–217
- Schreiber G (2024) Reconsidering agency in the age of AI. Filozofia 79(5):529–537. <https://doi.org/10.31577/filozofia.2024.79.5.5>
- Schreiber G, Ohly L (eds.) (2024) KI:Text. Diskurse über KI-Textgeneratoren. Berlin, Boston. <https://library.oapen.org/handle/20.500.12657/87894>
- Seeber I, Bittner E, Briggs RO, de Vreede T, de Vreede G-J, Elkins A, Maier R, Merz A, Oeste-Reiß S, Randrup N, Schwabe G, Söllner M (2020) Machines as teammates: a research agenda on AI in team collaboration. Inf Manag 57(2):103174. <https://doi.org/10.1016/j.im.2019.103174>
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128(2):336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Shetty HN, Asampalli S, Bonam PVP (2022) Data bias detection in machine learning. Int J Sci Res Publ 12(10):557–562. <https://doi.org/10.29322/IJSRP.12.10.2022.p13071>
- Shneiderman B (2022) Human-centered AI. Oxford University Press, Oxford
- Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):1–48. <https://doi.org/10.1186/s40537-019-0197-0>
- Sloane M (2019) Inequality is the name of the game: thoughts on the emerging field of technology, ethics and social justice. In: proceeding Weizenbaum Conference 2019 “challenges of digital inequality - digital education, digital work, digital life”. <https://doi.org/10.34669/wi.cp/2.9>
- Smits J, Borghuis T (2022) Generative AI and intellectual property rights. In: law and artificial intelligence: regulating AI and applying AI in legal practice. Springer Nature, pp. 323–344. https://doi.org/10.1007/978-94-6265-523-2_17
- Spiekermann S, Krasnova H, Hinz O, Baumann A, Benlian A, Gimpel H, Heimbach I, Köster A, Maedche A, Niehaves B et al (2022) Values and ethics in information systems: a state-of-the-art analysis and avenues for future research. Bus Inf Syst Eng 64(2):247–264
- Terwiesch C (2023) Would Chat GPT get a Wharton MBA? A prediction based on its performance in the operations management course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania
- Theis S, Jentzsch S, Deligiannaki F, Berro C, Rauff AP, Bruder C (2023) Requirements for explainability and acceptance of artificial intelligence in collaborative work. In: Degen H, Ntoua S (eds) Artificial intelligence in HCI. Springer Nature, Cham, pp 355–380
- UBS (2023) Let's chat about ChatGPT. In: UBS wealth management USA. <https://www.ubs.com/us/en/wealth-management/insights/article.1585717.html>. Accessed 8 May 2024
- van Dun C, Moder L, Kratsch W, Röglinger M (2023) ProcessGAN: Supporting the creation of business process improvement ideas through generative machine learning. Decis Support Syst 165:113880. <https://doi.org/10.1016/j.dss.2022.113880>
- Verma S, Singh V (2022) Impact of artificial intelligence-enabled job characteristics and perceived substitution crisis on innovative work behavior of employees from high-tech firms. Comput Hum Behav 131:107215. <https://doi.org/10.1016/j.chb.2022.107215>
- Vodrahalli K, Daneshjou R, Gerstenberg T, Zou J (2022) Do humans trust advice more if it comes from AI? An analysis of human-AI Interactions. In: proceedings of the 2022 AAAI/ACM conference on AI, ethics, and society. ACM, Oxford, pp 763–777
- Vynck GD (2023) ChatGPT ‘hallucinates.’ Some researchers worry it isn’t fixable. In: Washington Post. <https://www.washingtonpost.com/technology/2023/05/30/ai-chatbots-chatgpt-bard-trust-worthy/>. Accessed 7 May 2024
- Wachter S, Mittelstadt B, Russell C (2017) Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv J L Tech 31:841
- Wallace L, Keil M, Rai A (2004) How software project risk affects project performance: an investigation of the dimensions of risk and an exploratory model. Decis Sci 35(2):289–321. <https://doi.org/10.1111/j.00117315.2004.02059.x>
- Wardlaw JM, Mair G, Von Kummer R, Williams MC, Li W, Storkey AJ, Trucco E, Liebeskind DS, Farrall A, Bath PM, White P (2022) Accuracy of automated computer-aided diagnosis for stroke imaging: a critical evaluation of current evidence. Stroke 53(7):2393–2403. <https://doi.org/10.1161/STROKEAHA.121.036204>
- WEF (2024) Global risks report 2024. In: World Economic Forum. <https://www.weforum.org/publications/global-risks-report-2024/>. Accessed 6 May 2024
- Wendt J, Wendt D (2024) Einigung auf Rechtsrahmen für Künstliche Intelligenz in der EU: AI Act. Z Prod Compliance 2(2024):86
- Werner MH (2021) Verantwortung. In: Grunwald A, Hillerbrand R (eds) Handbuch technikethik. Metzler, Stuttgart, pp 44–48
- Wiener M, Cram WA, Benlian A (2023) Algorithmic control and gig workers: a legitimacy perspective of Uber drivers. Eur J Inf Syst 32(3):485–507. <https://doi.org/10.1080/0960085X.2021.1977729>
- Winner L (1983) Technologies as forms of life. In: Cohen RS, Wartofsky MW (eds) Epistemology, methodology, and the social sciences. Springer, Dordrecht, pp 249–263
- Wolfe R, Banaji MR, Caliskan A (2022) Evidence for hypodescent in visual semantic AI. In: 2022 ACM conference on fairness, accountability, and transparency. ACM, Seoul, pp 1293–1304
- Yang L, Holtz D, Jaffe S, Suri S, Sinha S, Weston J, Joyce C, Shah N, Sherman K, Hecht B, Teevan J (2022) The effects of remote work on collaboration among information workers. Nat Hum Behav 6(1):43–54. <https://doi.org/10.1038/s41562-021-01196-4>
- Zhao H, Chen H, Yang F, Liu N, Deng H, Cai H, Wang S, Yin D, Du M (2024) Explainability for large language models: a survey. ACM Trans Intel Syst Technol 15(2):1–38. <https://doi.org/10.1145/3639372>