Towards Responsible Augmentation: Identifying Characteristics of AI-based Technology with Ethical Implications for Knowledge Workers

Sünje Clausen  
*University of Potsdam, Germany*, suenje.clausen@uni-potsdam.de

Felix Brünk er  
*University of Potsdam, Germany*, felix.bruenker@uni-potsdam.de

Stefan Stieglitz  
*University of Potsdam, Germany*, stefan.stieglitz@uni-potsdam.de

Follow this and additional works at: https://aisel.aisnet.org/acis2023

Recommended Citation  
Clausen, Sünje; Brünk er, Felix; and Stieglitz, Stefan, "Towards Responsible Augmentation: Identifying Characteristics of AI-based Technology with Ethical Implications for Knowledge Workers" (2023). *ACIS 2023 Proceedings*. 123.  
https://aisel.aisnet.org/acis2023/123

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Towards Responsible Augmentation: Identifying Characteristics of AI-based Technology with Ethical Implications for Knowledge Workers

Research-in-Progress

Sünje Clausen
Business Information Systems and Digital Transformation
University of Potsdam
Potsdam, Germany
Email: suenje.clausen@uni-potsdam.de

Felix Brünker
Business Information Systems and Digital Transformation
University of Potsdam
Potsdam, Germany
Email: felix.bruenker@uni-potsdam.de

Stefan Stieglitz
Business Information Systems and Digital Transformation
University of Potsdam
Potsdam, Germany
Email: stefan.stieglitz@uni-potsdam.de

Abstract

Artificial intelligence (AI)-based technologies can increasingly perform knowledge work tasks, such as medical diagnosis. Thereby, it is expected that humans will not be replaced by AI but work closely with AI-based technology (“augmentation”). Augmentation has ethical implications for humans (e.g., impact on autonomy, opportunities to flourish through work), thus, developers and managers of AI-based technology have a responsibility to anticipate and mitigate risks to human workers. However, doing so can be difficult as AI encompasses a wide range of technologies, some of which enable fundamentally new forms of interaction. In this research-in-progress paper, we propose the development of a taxonomy to categorize unique characteristics of AI-based technology that influence the interaction and have ethical implications for human workers. The completed taxonomy will support researchers in forming cumulative knowledge on the ethical implications of augmentation and assist practitioners in the ethical design and management of AI-based technology in knowledge work.

Keywords: Artificial intelligence, augmentation, taxonomy, human-AI interaction, ethics
1 Introduction

In the future of knowledge work (KW), artificial intelligence (AI)-based technology will automate certain tasks and augment human workers in others (Dellermann et al. 2019; Raisch and Krakowski 2021). The integration of AI in KW promises, among other things, higher performance outcomes (Anthony et al. 2023; Enholm et al. 2021). While augmentation can benefit workers, for example, by reducing cognitive load and technostress (Brachten et al. 2020; Ulfert et al. 2022), it can also have negative implications. For example, by making workers feel threatened (Strich et al. 2021) or requiring them to make impactful decisions based on opaque AI advice (Lebovitz et al. 2022). Understanding how augmentation will affect human workers is crucial for making informed decisions about the design and management of AI-based technology in the future of KW (i.e., for “responsible augmentation”).

Changes to the nature of work have ethical implications for human workers as “work is one means by which we might pursue and achieve a good life” (Michaelson et al. 2014). Accordingly, business ethicists have argued that society and organizations have an obligation to support opportunities for work that contribute to human flourishing (Bowie 1998; Schwartz 1982). While there is no universally accepted definition of what constitutes so-called “meaningful work” (Michaelson et al. 2014), having autonomy and being able to develop and use one’s skills and cognitive abilities are considered to be important elements of meaningful work (Bowie 1998; Schwartz 1982). The introduction of AI-based technology will affect workers in a variety of ways (Mayer et al. 2020), including their autonomy in their work and ability to use and develop certain skills. Therefore, in addition to the previously mentioned aspects, AI has ethical implications for human workers in the sense that the introduction of AI will influence humans’ prospects to flourish (Bankins and Formosa 2023).

Anticipating how AI-based technology will change KW and affect human workers can be challenging. AI is not a single or static technology but “the frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems” (Berente et al. 2021, p. 1435), and is therefore not only an umbrella term that has been used to describe a variety of technologies with different characteristics, but also a moving target. Different AI-based technologies are associated with different risks (Asatiani et al. 2021). Some AI-based technologies differ significantly from previously available technologies (e.g., in terms of agency, adaptability, inscrutability), and thus, previous assumptions about human–technology interaction may not apply to AI-based technology in the same way (Baird and Maruping 2021; Berente et al. 2021; Schuetz and Venkatesh 2020).

The characteristics of AI-based technology and how they shape human–technology interaction matter for the (potential) impact of augmentation on human workers. For example, designing technology in a way that makes its inner workings more intelligible for humans (e.g., by displaying information about goals, proposed actions, reasoning processes, or uncertainty) influences how users perceive (or challenge) AI recommendations (Hofeditz et al. 2022; Lebovitz et al. 2022) and develop trust in the technology (Meske et al. 2022; Vössing et al. 2022). The distribution of agency between the human and the technology (e.g., who delegates to whom; hierarchy) relates to task performance (Fügener et al. 2021a) and how the workers experience the interaction with the AI and the meaningfulness of their work (e.g., effects on cognitive load, technostress, professional role identity) (Mirbabaie et al. 2022; Strich et al. 2021; Ulfert et al. 2022). How AI-based technology is designed and deployed in work environments influences potential positive or negative implications for human workers. Thus, the decision makers (i.e., the developers of the technology and the organizational actors that manage or implement it) have a responsibility to design and deploy technology in a way that is mindful of its ethical implications (Berente et al. 2021; Elliott et al. 2021; Lobchak et al. 2021; Martin et al. 2019).

In summary, augmenting KW with AI-based technology will have ethical implications for human workers, and decision makers have a responsibility to protect workers from potential harm. To evaluate the ethical implications, one must consider the specific characteristics of the technology and how they shape the interaction with human workers. Due to a vague definition of what constitutes AI and a high variability of possible configurations, current research lacks the specificity that would be required to build cumulative knowledge for responsible augmentation. Against this background, we formulate the following research question: How do characteristics of AI-based technology relate to ethical implications for human knowledge workers?

By answering this research question, we contribute to the cumulative knowledge base for both the ethical design and management of AI-based technology in the context of KW. As Information Systems (IS) research operates at “the intersection of knowledge of the properties of physical objects (machines) and knowledge of human behavior” (Gregor 2006, p. 613), we are in a unique position to address this question. In this research-in-progress paper, we propose to develop a taxonomy for AI-based technology in KW (Kundisch et al. 2022; Nickerson et al. 2013). A taxonomy allows classifying AI-based
technologies in KW based on those characteristics of the technology that shape the interaction with (and ethical implications for) human workers. This would enable researchers and practitioners to better describe, understand, analyze, and predict the impact of augmentation in KW (c.f. Nickerson et al. 2013) and pave the way towards responsible augmentation.

2 Literature Background

As the integration of AI promises to increase business value (Enholm et al. 2021), it is expected that AI-based technologies will play an important role in the future of KW (Coombs et al. 2020). We refer to KW as work that involves “the creation, distribution or application of knowledge” (Davenport 2005, p. 9) and focus on employed knowledge workers. Research suggests that knowledge workers will not be replaced by AI but rather “collaborate closely with machines to perform a task” (Raisch and Krakowski 2021, p. 192); also referred to as augmentation. Similarly, Seeber et al. (2020) describe a scenario in which machines become teammates working side by side with humans. Delligmann et al. (2019) argue that in the future, human intelligence and AI will be combined to complement each other’s strengths and form a superior hybrid intelligence. If humans and AI-based technologies will interact and collaborate more closely in the future of work, questions arise on how this interaction should be designed and how it might affect human workers (Mirbabaie et al. 2022)

While there is a wide range of knowledge work professions (Coombs et al. 2020), the types of AI-based systems that are used for KW augmentation often have similar characteristics and, correspondingly, have similar implications for human workers. For example, previous research found that medical professionals who use AI-based technology for diagnosing medical images (e.g., detecting lung cancer in computer tomography scans) experience uncertainty due to the opacity of the AI (i.e., not being able to understand how or why the AI makes a suggestion), especially if the AI’s judgement differs from their own professional opinion (Lebovitz et al. 2022). Similarly, in a study on an AI-based revenue forecasting technology to support personnel planners in restaurants, Vössing et al. (2022) found that making the AI’s reasoning more transparent increased the personal planners’ confidence in the technology. In both studies, domain experts are supported by an AI-based technology to make decisions under uncertainty. While the KW professions and tasks are different, the AI’s characteristic of (partial) inscrutability is similar and, in both cases, this characteristic affected the human workers’ experience of and response to working with AI-based technology. Therefore, for forming a cumulative knowledge base on how augmentation will affect human workers, it is a valuable starting point to identify and analyze characteristics of AI-based technologies that are unique to AI and shape the interaction with humans.

Several recent IS papers have discussed how AI-based technologies differ from previously available technologies and correspondingly, how these differences shape human–technology interaction and require new approaches for IS research and management (Baird and Maruping 2021; Berente et al. 2021; Schuetz and Venkatesh 2020). Schuetz and Venkatesh (2020) argue that cognitive computing systems (a form of an AI-based technology) enable new forms of user-system-environment interaction that challenge previous assumptions in IS research. Unlike previous types of technology, AI-based technologies are no longer just tools, but can act as agents of their own interests (i.e., creating a bilateral relationship), they are aware of their environment and can respond to unspecific inputs, they are adaptive, opaque, and humans may be unaware of interacting with an AI. Baird and Maruping (2021) also discuss the role of increasingly agentic IS artifacts which challenge the “human agency primacy” assumption that has been prevalent in IS research. Instead of being a passive tool, agentic IS artifacts might delegate tasks to human agents (Baird and Maruping 2021) which will affect human workers and work outcomes (Fügener et al. 2021a, 2021b). Lastly, Berente et al. (2021) discuss three interrelated facets of AI that set it apart from previous technologies and pose challenges for managing AI. These facets are autonomy (i.e., acting without human intervention), learning (i.e., improving through data and experience), and inscrutability (i.e., being unintelligible to specific audiences). Each of these facets pose challenges for making decisions with or about the technology (Berente et al. 2021). Thereby, it is important to recognize that not all AI-based technologies possess all of these characteristics. For example, different AI-based technologies have different degrees of inscrutability (Asatiani et al. 2021).

Augmentation with AI will bring different changes to KW but not all of them are equally relevant from an ethical point of view. One lens for evaluating changes to the nature of work is to focus on those aspects that contribute to (or hinder) meaningful work. Based on Kantian ethics, Bowie (1998) proposes six characteristics of meaningful work, including, work that is i) freely entered into, ii) allows the worker to exercise her autonomy and independence, iii) enables the worker to develop her rational capacities, iv) provides a wage sufficient for physical welfare, v) supports the moral development of employees, and vi) is not paternalistic in the sense of interfering with the worker’s conception of how she wishes to obtain happiness. In the context of AI’s impact on meaningful work, Bankins and Formosa (2023) focus on five
categories, including for example the cultivation and use of skills and autonomy in how work is conducted. From this perspective, AI-based technologies that reduce worker’s autonomy and prevent them from cultivating and using their skills have negative ethical implications for workers because they reduce opportunities for conducting meaningful work.

3 Method

In this research, we follow the taxonomy development method for IS research introduced by Nickerson et al. (2013) and refined by Kundisch et al. (2022). Kundisch et al. (2022) classify taxonomies as nascent design theories that provide prescriptions for design and action. An advantage of classifying objects or phenomena of interest in a taxonomy is that it allows the identification of common properties and provides a means for comparing and contrasting them (Gregor, 2006). By shifting the focus to characteristics of AI-based technology that shape the interaction with and experience of humans, a taxonomy allows both existing and future AI-based technologies to be compared and contrasted in terms of their impact on human workers.

The extended taxonomy design process by Kundisch et al. (2022) involves six steps: i) identify problem and motivate, ii) define objectives of a solution, iii) design and development, iv) demonstration, v) evaluation, and vi) communication. In this research-in-progress paper, we focus primarily on the first three steps of the process. We develop and describe a preliminary taxonomy in Section 4. We have already specified the observed phenomenon (augmenting knowledge workers with AI-based technology) in the previous chapters. In the following, we determine the meta-characteristic of the taxonomy based on the target user groups and intended purposes. The meta-characteristic describes the objective or purpose that the taxonomy should serve.

The first group of potential users are researchers in IS and closely related disciplines such as human-computer interaction, computer science, cognitive psychology, and organization science, who are conducting research on AI-based technology in the context of KW. We expect researchers to use the taxonomy to categorize their research and thereby systematically abstract from a specific AI-based technology, using case, or experimental result to more general implications for the design and management of AI-based technology. Systematically structuring the investigated use cases and experimental results could also help to identify research gaps and blind spots in current research on AI-based technology in KW.

A second group of potential users are managers who introduce or oversee the introduction of AI-based technology in knowledge work. Here, the taxonomy should help to anticipate the possible ethical implications of introducing a particular type of AI-based technology. Based on the taxonomy, managers should be able to make more informed decisions about a suitable type of AI-based technology and the process of introducing such a system into a work environment.

Based on the target users and purpose, we formulate the following meta characteristic for the taxonomy: Distinguishing features that characterize the interaction between a user and an AI-based technology and have ethical implications for the human worker. Following Nickerson et al. (2013), the taxonomy should satisfy both objective (“is it a taxonomy?”) and subjective (“is it an applicable taxonomy?”) ending conditions. The evaluation goal of the taxonomy is to enable researchers and practitioners to better describe and classify an AI-based technology with respect to the features that have ethical implications for human workers, and to support researchers and practitioners in analyzing the potential impact of implementing an AI-based technology in KW.

We follow a conceptual-to-empirical approach for the taxonomy development. This approach is recommended when the empirical data on the subject is limited and a theoretical understanding of the phenomenon in focus already exists (Kundisch et al. 2022; Nickerson et al. 2013). While there are initial studies on human-AI interaction and collaboration (Fügener et al. 2021a, 2021b; Lebovitz et al. 2022; Vössing et al. 2022), these might not cover all relevant characteristics of AI-based technology that could have ethical implications for workers. Therefore, as a basis for the first conceptualization, we draw on selected papers that discuss the interaction of humans and AI-based technology (e.g., Baird and Maruping 2021; Berente et al. 2021; Schuetz and Venkatesh 2020) as well as prior knowledge of the researchers. In a next step, we will examine both research papers on human-AI interaction and collaboration in KW and exemplary AI-based technologies for KW to evaluate and revise the proposed characteristics.
4 Preliminary Taxonomy

In the following, we describe four unique characteristics of AI-based technology and why it is relevant to consider them in the augmentation of knowledge workers. One characteristic in which AI-based technologies differ from previous technologies is their autonomy, that is, their ability to act on their own accord, sometimes even without human intervention or awareness (Baird and Maruping 2021; Berente et al. 2021). Schuetz and Venkatesh (2020) note that the relationship between human and technology is becoming bidirectional as the technology is no longer just a tool that responds to user instructions, but might pursue its own interests and objectives. Baird and Maruping (2021) distinguish four agentic archetypes (i.e., reflexive, supervisory, anticipatory, and prescriptive) on a continuum of agency. The degree of autonomy of an AI-based system could take the form of the power to execute decisions or to act without human awareness (e.g., if interactions can be initiated by the AI), and is related to a variety of implications for human workers (e.g., their professional role identity, Strich et al. (2021)). Thus, different dimensions of autonomy were included in the taxonomy (see Table 1).

Further, Berente et al. (2021) identify the inscrutability of (some) AI-based technologies, that is, their unintelligibility to specific audiences, as a facet of AI that differs from previous technologies and makes its management more challenging. While the desire to “unpack” the technology’s inner workings differs between user groups and application scenarios, being able to compare one’s own judgement with the technology’s output, challenge assumptions, and to negotiate differences is particularly important for experts, especially in the context of accountability for decisions and actions (Anthony et al. 2023; Lebovitz et al. 2022). Not all AI-based technologies are equally inscrutable, instead, there tends to be an accuracy-explainability trade-off (Asatiani et al. 2021). To address this issue, the field of explainable AI proposes a variety of methods for making AI-based technologies less inscrutable (Adadi and Berrada 2018). Thereby, global explanations convey how the model or technology works in general while local explanations provide insights on specific actions or suggestions of a system (Meske et al. 2022). The “degree of inscrutability” has ethical implications for workers and is influenced by technology design (e.g., provision of global and/or local explanations). Thus, different dimensions of inscrutability were included in the taxonomy (see Table 1).

<table>
<thead>
<tr>
<th>Characteristics of the AI-based technology</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Autonomy</strong></td>
<td></td>
</tr>
<tr>
<td>Interaction can be initiated by...</td>
<td></td>
</tr>
<tr>
<td>Actions be executed by...</td>
<td></td>
</tr>
<tr>
<td>Delegation decisions are made by...</td>
<td></td>
</tr>
<tr>
<td>Actions can be intercepted by...</td>
<td></td>
</tr>
<tr>
<td>• Only the human</td>
<td></td>
</tr>
<tr>
<td>• Both the human and the system</td>
<td></td>
</tr>
<tr>
<td>• Only the system</td>
<td></td>
</tr>
<tr>
<td><strong>Inscrutability</strong></td>
<td></td>
</tr>
<tr>
<td>The system is designed to...</td>
<td></td>
</tr>
<tr>
<td>• Make the user aware of interacting with AI</td>
<td></td>
</tr>
<tr>
<td>• Conceal or omit information on its nature</td>
<td></td>
</tr>
<tr>
<td>The system provides...</td>
<td></td>
</tr>
<tr>
<td>• Global explanations</td>
<td></td>
</tr>
<tr>
<td>• Local explanations</td>
<td></td>
</tr>
<tr>
<td>• Global and local explanations</td>
<td></td>
</tr>
<tr>
<td>• No explanations</td>
<td></td>
</tr>
<tr>
<td><strong>Learning</strong></td>
<td></td>
</tr>
<tr>
<td>The system’s functionality is...</td>
<td></td>
</tr>
<tr>
<td>• Adaptive</td>
<td></td>
</tr>
<tr>
<td>• Consistent until intentionally updated</td>
<td></td>
</tr>
<tr>
<td><strong>Human likeness</strong></td>
<td></td>
</tr>
<tr>
<td>The system’s appearance is...</td>
<td></td>
</tr>
<tr>
<td>• Human-like</td>
<td></td>
</tr>
<tr>
<td>• Artificial</td>
<td></td>
</tr>
<tr>
<td>The system’s behavior...</td>
<td></td>
</tr>
<tr>
<td>• Resembles human behaviour</td>
<td></td>
</tr>
<tr>
<td>• Does not resemble human behaviour</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1. Preliminary taxonomy*

A third facet, that has implications for human-AI interaction and sets it apart from previous technologies, is learning (Berente et al. 2021), also referred to as constant change (Anthony et al. 2023).
or adaptivity (Schuetz and Venkatesh 2020). While previous information systems typically had a consistent functionality, meaning that consistent use lead to consistent outcomes, this is not always true for AI-based systems as their functionality might change adaptively, potentially without the awareness of the user (Schuetz and Venkatesh 2020). However, not all AI-based technologies have this functionality. Accordingly, the taxonomy distinguishes between adaptive and consistent systems (see Table 1).

Lastly, another aspect that sets AI-based technologies apart from previous technology is that they can appear human-like to users. For example, most conversational agents are designed with human-like abilities and some even aim to form human-like relationships with the user (e.g., Pi by Inflection AI). The belief of interacting with a human instead of an AI positively affects the perception of the interaction (Ashtktorab et al. 2020). However, perceiving AI as human might also lead to unintended negative consequences for the human. In 2022, the Google engineer Blake Lemoine concluded that the “Language Model for Dialogue Applications” (LaMDA) was sentient and, after not being taken seriously, published internal documents demonstrate the model’s capabilities to the public (Tiku 2022). Google laid off the employee and denied the claims. It can be expected that future knowledge workers might ask themselves similar questions when interacting with AI-based technology that appears to be human-like. Deciding on the level of human characteristics of AI-based technology might also influence the perception of how much autonomy (and rights) the technology decision deserves. In addition, human workers might transfer potential biases (e.g., gender, race) to an AI that resembles a human with those characteristics. Therefore, the taxonomy should distinguish between AI-based technologies that are designed to build a human-like relationship and those that do not (Table 1).

5 Next Steps and Conclusion

In this research-in-progress paper, we propose the development of a taxonomy that distinguishes characteristics of AI-based technology which shape the interaction with and have ethical implications for human workers. By doing so, we aim to stimulate a more structured and specific debate on the impact of augmenting KW with AI-based technologies. We envision researchers to use the taxonomy to systematically abstract from an AI method, use case, or experimental result to the characteristics of AI-based technology that are included in the taxonomy. This will contribute to forming a cumulative knowledge base on how human workers will be affected by the introduction of different types of AI as well as uncover blind spots in existing research. In addition, we aim to support developers and managers of AI-based technology on a path towards responsible augmentation. Thereby, we do not claim that the characteristics of the AI-based technology are the only factor shaping the ethical implications of augmentation for knowledge workers. The organizational context in which the technology is deployed as well as characteristics of the user are likely to play a role as well. A taxonomy will enable decision makers to better anticipate and mitigate potential harm associated with different AI-based technologies in KW. In the next steps of this research, we will examine empirical research papers that investigate scenarios in which humans interact with AI-based technology for KW tasks. We will classify the investigated AIs with the taxonomy and potentially add and/or revise its dimensions. We will also classify exemplary AI-based technologies for KW that are already used in practice to evaluate and revise the taxonomy. Once the design and development of the taxonomy are complete, we aim to conduct workshops with researchers and practitioners who are working with AI-based technologies in the context of KW to demonstrate and evaluate the taxonomy.

6 References


Copyright © 2023 Sünje Clausen, Felix Brünker, Stefan Stieglitz. This is an open-access article licensed under a Creative Commons Attribution-Non-Commercial 3.0 Australia License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and ACIS are credited.