

# Explainability Challenges in Continuous Invisible AI for Self-Augmentation

Dinara Talypova<sup>1,2,\*</sup>, Philipp Wintersberger<sup>1,2</sup>

<sup>1</sup>University of Applied Sciences Upper Austria, Hagenberg, Austria

<sup>2</sup>TU Wien, Vienna, Austria

## Abstract

Despite the substantial progress in Machine Learning in recent years, its advanced models have often been considered opaque, offering no insight into the precise mechanisms behind their predictions. Consequently, engineers today try to implement the explainability factors into the developed models, essential for trust and adoptancy of the system. Still, there are several blocks in Explainable Artificial Intelligence (XAI) research that cannot follow the standard design methods and guidelines for providing transparency and ensuring maintaining human objectives. In this position paper, we attempt to chart various AI blocks from the perspective of Human-Computer Interaction field and identify potential gaps requiring further exploration. We suggest three-level dimension classification: relations with humans (replacing vs augmenting), interaction complexity (discrete vs. continuous), and the object of application (external world or users themselves).

## Keywords

XAI, HCI, Continuous AI, Seamless Technology, Attention Management System

## 1. Introduction

Until these days, Artificial intelligence (AI) maintains a nuanced relationships with humans, oscillating between a potential replacement of them to a technology that can augment and expand human capabilities. As smart technologies become more prevalent in our daily routines, the necessity for a smooth integration between humans and AI becomes evident ([1, 2, 3, 4, 5, 6] just to name a few). This involves achieving a balanced incorporation of technology in our day-to-day activities, recognizing its advantages while also being mindful of potential challenges and unintended consequences.

The common believe is that some of the issues could be uplifted by focusing on AI explainability [7, 8, 9]. Explainable Artificial Intelligence (XAI), a field that is focused on interpreting machine decisions, takes its roots with the advent of AI systems already half a century ago [10, 11, 12]. The shift from rule-based systems to modern deep learning has increased the difficulty of interpreting AI decisions, moving beyond the simpler applied rules scenarios. Current efforts in XAI involve the attempts to open AI's "black box" by enhancing transparency and providing clear explanations for AI predictions, like justifying flagged fraudulent transactions or selection processes in recruitment.

---

*MuM'23 Workshop on Interruptions and Attention Management: Exploring the Potential of Generative AI, 2023, Vienna, Austria*

✉ [dinara.talypova@fh-hagenberg.at](mailto:dinara.talypova@fh-hagenberg.at) (D. Talypova); [philipp.wintersberger@fh-hagenberg.at](mailto:philipp.wintersberger@fh-hagenberg.at) (P. Wintersberger)

🆔 0000-0002-6612-8061 (D. Talypova); 0000-0001-9287-3770 (P. Wintersberger)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

**Table 1**

AI research blocks from an HCI research perspective

	Replacing Humans	Augmenting Human Capabilities	
		User as an agent	User as an object
Discrete AI	<ul style="list-style-type: none"> <li>•Automated Customer Support (e.g., chatbots);</li> <li>•Automated Stock Trading</li> </ul>	<ul style="list-style-type: none"> <li>• Image Recognition apps (e.g., user apps for plant/bird detection);</li> <li>• Photoshop Generative Fill;</li> <li>• Clinical Diagnostics support;</li> <li>• Google Translate</li> </ul>	<ul style="list-style-type: none"> <li>•Physical Activity Trackers (with immediate reminders)</li> </ul>
Continuous AI	<ul style="list-style-type: none"> <li>•Self-Driving Cars;</li> <li>•Industrial robots</li> </ul>	<ul style="list-style-type: none"> <li>• (Advanced) Google Assistant;</li> <li>• Industrial robots;</li> <li>• Traffic Management Systems;</li> <li>• AI-driven Investment Portfolios</li> </ul>	<ul style="list-style-type: none"> <li>•Attention Management Systems;</li> <li>• AI-enhanced sleep app;</li> <li>•Diabetes App;</li> <li>•Personalized News Feed;</li> <li>•“Smart” Fridge (i.e., monitoring diet &amp; ordering food)</li> </ul>

Moreover, the XAI challenge extends beyond revealing the internal inference processes of AI algorithms — this is also important how the insights and explanations are communicated to end-users. Human-Computer Interactions (HCI) researchers are concerned with ensuring that the explanations are not only technically accurate but also usable, effective, and satisfying for the people who interact with AI systems [13, 14, 15, 16]. This entails considering different approaches based on user profiles, be they professionals utilizing the system (e.g., medical practitioners), those affected by the system (e.g., patients), or developers themselves seeking algorithm improvement. In summary, XAI’s role in HCI is twofold, aiming both to foster user trust (that is, how well the insights could be delivered) and reinforce the reliability of AI systems (that is, how reliable the conclusions are). As nicely formulated by Diefenbach et al., “*we must be convinced that the other is capable of handling the task and that the other means well.*” [17]

In this position paper, we attempt to chart various AI blocks from the perspective of HCI research in explainability and identify potential gaps requiring further exploration. We believe that despite long XAI research history, the field is unbalanced: while some categories of AI got more attention from XAI perspective, the others remain in shadow. Therefore, to highlight this skew, there is a need to categorize AI systems along the dimensions affecting their utilization, continuity, as well as the nature of interaction with the world and humans. We came up with three-level dimensions: relations with humans (replacing vs augmenting), interaction complexity (discrete vs. continuous), and the object of application (external world or users themselves). The blocks with examples of AI technology implementation from the industry are presented in Table 1.

## 2. Replacing Humans vs. Augmenting Human Capabilities

One perspective in the interaction debate revolves around whether smart technologies should replace humans [18, 19], thereby freeing up time for more meaningful or enjoyable activities, or whether AI technologies should operate in tandem with humans [20, 21, 22], offering support and enhancement. The former angle is exemplified by applications like automated customer support (e.g., chatbots), industrial automation, and self-driving vehicles, all aimed at automating tasks to potentially allow individuals to engage in more creative or self-developmental pursuits. This still assumes a clear level of system transparency and interpretability of the outcome, so

humans monitoring the smart machines could trace back the solution process [23]. From the XAI point of view, this approach may involve specific reports of what have been done and how [24]. However, under this perspective, the XAI goal is not just explainability of the results but rather reliability of the system, i.e., the trust that system can exhibit the same and expected behavior over time [25].

In contrast, proponents of the augmentation approach contend that machines should collaborate with people to reinforce human capabilities. Thus, Autor et al. [26] argue that typical job responsibilities are composed of task bundles; some of the tasks within these bundles are more amenable to automation than others. This view inclines that AI will not displace humans and their jobs entirely; instead, it will be used to enhance particular facets of tasks, promoting effective collaboration between humans and AI (HAIC) [27].

It has already been demonstrated that HAIC is more efficient in diverse types of tasks compared to humans' or AI performance in isolation [20, 21, 28, 29]. In such tandem, each party is responsible for the task in the bundle that it does better. In response to this belief, some academic conversations are shifting towards enhancing human capacities through AI integration, redefining job roles for effective partnership, and transforming business processes to align with these new collaborative models [30, 6, 31, 32]. Moreover, this concept does not stop in efficiency. We envision that AI systems should not only automate tasks in order to increase productivity, but also support, amplify, and extend human skills, creativity, and potential. That is, some tasks will not be fully passed to AI also because of their enjoyable nature for humans and feeling of autonomy.

Applications embodying skill augmentation approach include clinical diagnostic support, writing assistance tools (e.g., Grammarly or ChatGPT), image recognition apps, and traffic management systems. Advocates of this perspective believe that such cooperative relationships will empower individuals, effectively elevating them to the status of "superhumans." As Dhiman et al. reasonably claimed, "*the assistants of the future will not only have to be trustworthy, respect our privacy, be accountable and fair, but also help us flourish as human beings*" [33]. However, this approach requires additional emphasis on transparency and explainability of human-AI interaction. Since humans stay in the process loop, understandable and meaningful cooperation should be the key to success.

### 3. Discrete AI vs. Continuous AI

Within an amplifying humans vision of AI, we can distinguish AI as a tool (that enhances human natural capabilities) or as a collaborator (that interacts with human and create fruitful synergy). AI as a tool assumes that human explicitly uses the system to get (most probably) immediate output, for instance a decision or a product. Consequently, interaction involving user input and AI output occurs coherently in distinct stages, as is the case with, for example, conversational agents, image recognition apps, AI writing support tools, etc. In these instances, AI systems follow a sequential pattern: users initiate the process by issuing a command, which is then interpreted by the system to generate a response. We call this type of interaction - Discrete AI.

While human-centered XAI has been extensively studied in the HCI field in the last decades [34, 8, 35, 14, 9], the big scope of the research has concentrated in the context of discrete AI

[16], i.e., decision-making in situ. At the same time, contemporary research must adapt to increasingly diverse and dynamic usage scenarios, where, for instance, user input and machine output seamlessly unfold in parallel [36]. That is, AI as an agent that collaborates (explicitly or implicitly) with the human being for their needs. These scenarios encompass applications like Traffic Management Systems, smart homes, AI-driven Investment Portfolios, Self-Driving Cars, and Industrial robots. They are "*intermittent, continuous, and proactive*" [36]. We call this type of interaction - Continuous AI.

Continuous AI is defined as '*the extended and seamless synergy between intelligent systems and human users over extended periods of time*' [37]. This category of AI underscores the potential of smart technology to provide support, assistance, and enhancement across various real-life contexts, provided that we can ensure user interactions remain non-disruptive and do not overburden cognitive resources. However, due to its proactiveness and continuous nature, very soon this technology becomes unnoticeable, or in the terminology of [38], invisible. That is, critical decision points are challenging to pinpoint, as the consequences of AI decisions accumulate gradually, making it difficult to identify a clear 'point of no return.' Therefore, on the one hand, seamless continuous technology offers benefits by simplifying tasks and alleviating the user from cognitive load. Yet, on the other hand, it introduces a degree of opaqueness, potentially leading to feelings of uncertainty and diminished autonomy. In this context, the importance of XAI becomes even more pronounced.

An in-depth examination reveals that current methods and guidelines, initially designed to facilitate interactions with AI systems, often prove inadequate and occasionally counterproductive when applied to continuous usage scenarios [37]. For instance, conventional evaluation methods often rely on assessing compliance and reliance rates to gauge user interactions with AI or automated systems [39, 40, 41]. However, this approach is most effective when interactions can be modeled as a series of isolated, subsequent trials, allowing users to retrace AI decisions to specific events. It is less suitable for continuous and parallel interactions. Furthermore, designed feedback mechanisms sometimes conflict with the seamless nature of continuous AI. In continuous AI, the technology aims to operate on the background to achieve its objectives, as in the example of Attention Management Systems (AMS) - technology that intelligently postpones user's notifications based on human's cognitive load to cope with disruptions [42, 43]. If the AMS always reaches the user with information explaining delayed notification, there will be no meaning in this system as it interrupts the user by trying to intelligently delay message. As a result, there is a pressing need to develop fresh guidelines and evaluation methods tailored to the description and assessment of Continuous AI scenarios.

#### **4. "User as an agent" vs. "User as an object"**

Guerrero et al. define the concept of *augmented humanity* as "*a human-computer integration technology that proposes to improve capacity and productivity by changing or increasing the normal ranges of human function through the restoration or extension of human physical, intellectual and social capabilities*" [44]. Unfortunately, in this definition, the grade of intrusion, i.e., whether the restoration/extension of human capabilities realised within one's body or through human-computer synergy, is not defined. Still, within the realm of our discussion, it is more relevant to

differentiate the level of effect on human agency rather than embodied vs. extended nature in the expression *human-computer integration*.

Numerous discussions on human-machine interactions revolve around a balanced and explicit partnership between an agent and a human. That is, technology serves a specific function to assist in reaching the user's goal, predominantly concentrated on data detached from the actor (e.g., [45]). In other words, a human represents an actor, (in most cases) an equal contributor to the interaction process, whether AI-system is a writing assistant that shapes the typed text for the industry standard, or an industrial robot that partners with the human in the manufacture.

When the user becomes the object of AI work in the context of self-augmentation, additional challenges come to the fore. Notably, psychological nuances related to self-manipulation and its acceptance factors become a central issue (see studies [46, 47]). For instance, the level of intrusiveness is higher in using exoskeleton than cooperating with the robot in industrial environment. When passing the decision about one's own life to AI, concerns related to ensuring AI's transparency and alignment with desired needs and objectives rise up. Moreover, even if the tasks of AI are aligned with our initial goals, are these goals still in place after our "modification"? We probably do not have that strong need to understand how or why the smart house changes the temperature in rooms and starts self-cleaning at a specific time. As long as it works and we do not freeze at home, it still makes our life easier, and that is what appears to matter most. However, when it comes to manipulation of our own state, bigger questions arise. Within the Discrete AI block, i.e., when the process of user input and AI output is sequential and each request is initiated by human, individual has at least some control on technological influence (that is, can trace one step back of where the result came from). While with Continuous AI, the complexity of the technological integration make it difficult to assess the impact. According to De Greef et al., the combination of increased complexity and undesirable machine behavior should be considered a major risk [48]. If the AI wrongly concludes which support or when is needed (or not needed), the potential benefits of adaptive automation turn into risks, elevated by the hidden nature of this augmentation (see more on the topic of AI alignment [49]). The examples here are a personalized news feed in social media that shows us specific news [3], an intelligent sleep app that sends us to bed at a particular time and adjusts our sleeping environment [50], or a smart habit-tracker that strategically manipulates what and how we have to do [51].

Additionally, some scholars divide human augmentation by facets of human capabilities, whether it is augmented senses (aka enhanced senses, extended senses), augmented action (e.g., motor augmentation, amplified force, and movement, speech input, etc.) or augmented cognition (by detecting human cognitive state, using analytical tools to make a correct interpretation) [52]. Within this division, the biggest shift in human agency is the last one — augmented cognition. The principle of augmented cognition describes the symbolic human-AI integration in a closed-loop system, where the system is designed to recognize the individual's mental state and the environmental context [53]. Thus, AI assistants could operate on behalf of humans, based on their habits and preferences, proactively handling different types of tasks.

An instance of this nuanced interaction is found in Attention Management Systems (AMS), which seamlessly operate in the background, collecting data from diverse sensors and gradually influencing user behavior and interactions over time. The question of human awareness of constant intentional nudging is extremely relevant here. In this context, transparency involves

communicating the simple presence or activation of technology, questioning whether it has already initiated a process or if we are still in an unaltered environment.

It brings to the table the clear tension between smart and invisible versus transparent and understandable technology. How can we find a balance between the seamless nature of the continuous AI technology and keeping human in the loop of decision-making? How can we control that the user still wants this technology and measure whether the system's goals continue to align with the user's needs? What should be the explainability form of continuous human-AI interaction for supporting trust and reliance in human augmentation? These are the open questions that the modern research in XAI should aim to solve.

## 5. Conclusion

In this position paper, we made an attempt to chart AI systems through the lens of HCI explainability research. By identifying three dimensions — relations with humans, interaction complexity, and the object of application — we underscored the inadequacy of universal explainability guidelines, particularly for seamless, continuous, and proactive AI systems oriented towards self-augmentation. Consequently, we send out a call for the development of new guidelines and evaluation methods specifically crafted to suit the characteristics of different AI blocks.

Based on our exploration and the statements above, we provide a list of the most pressing research questions around Explainable AI research that must be addressed to advance the field effectively:

- **Addressing the Gap in Continuous AI for Self-Augmentation:** Highlighting the need for focused research on the specific challenges posed by Continuous AI technologies designed for self-augmentation, such as ensuring transparency and user control in highly integrated and proactive systems, to balance the benefits of seamless interaction with the necessity of maintaining individuals' autonomy.
- **Development of XAI Techniques for Continuous AI Systems:** Establishing new methods group to enhance transparency and human control in continuous AI presence, ensuring explanations are comprehensible without abusing limited users' attention.
- **Enhancing Human-AI Collaboration & AI Alignment:** Identifying ways to improve human-AI collaboration to augment human capabilities effectively, focusing on empowering individuals, preserving autonomy and ensuring the system still works in line with human needs.
- **Development of New Guidelines and Evaluation Methods:** Creating flexible guidelines and evaluation methods that can accurately assess the effectiveness as well as potential biases in explainability across varied and evolving usage scenarios in Continuous AI for self-augmentation.
- **Ethical Implications of Human-Augmentation AI Systems:** Addressing the ethical considerations of AI systems that act as agents of human augmentation, with particular attention to the questions of free will, autonomy and consent, ensuring these technologies, at minimum, avoid misleading and are developed responsibly.



## References

- [1] S. Wei, P. Huang, R. Li, Z. Liu, Y. Zou, Exploring the application of artificial intelligence in sports training: a case study approach, *Complexity* 2021 (2021) 1–8.
- [2] B. Stoel, Use of artificial intelligence in imaging in rheumatology–current status and future perspectives, *RMD open* 6 (2020).
- [3] R. Rathi, Effect of cambridge analytica’s facebook ads on the 2016 us presidential election, *Towards Data Science* (2019).
- [4] Q. André, Z. Carmon, K. Wertenbroch, A. Crum, D. Frank, W. Goldstein, J. Huber, L. Van Boven, B. Weber, H. Yang, Consumer choice and autonomy in the age of artificial intelligence and big data, *Customer needs and solutions* 5 (2018) 28–37.
- [5] E. Stamboliev, Proposing a postcritical ai literacy: Why we should worry less about algorithmic transparency and more about citizen empowerment, *Media Theory* 7 (2023) 202–232.
- [6] I. Rudko, A. Bashirpour Bonab, F. Bellini, Organizational structure and artificial intelligence. modeling the intraorganizational response to the ai contingency, *Journal of Theoretical and Applied Electronic Commerce Research* 16 (2021) 2341–2364.
- [7] G. Vilone, L. Longo, Notions of explainability and evaluation approaches for explainable artificial intelligence, *Information Fusion* 76 (2021) 89–106.
- [8] D. Gunning, D. Aha, Darpa’s explainable artificial intelligence (xai) program, *AI magazine* 40 (2019) 44–58.
- [9] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, M. Kankanhalli, Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda, in: *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–18.
- [10] D. W. Hasling, W. J. Clancey, G. D. Rennels, Strategic explanations for a diagnostic consultation system, *Int. J. Man Mach. Stud.* 20 (1983) 3–19. URL: <https://api.semanticscholar.org/CorpusID:15924245>.
- [11] W. R. Swartout, Explaining and justifying expert consulting programs, in: *Computer-assisted medical decision making*, Springer, 1985, pp. 254–271.
- [12] A. C. Scott, W. J. Clancey, R. Davis, E. H. Shortliffe, Explanation capabilities of production-based consultation systems, *American Journal of Computational Linguistics* (1977) 1–50.
- [13] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, J. D. Weisz, Expanding explainability: Towards social transparency in ai systems, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–19.
- [14] U. Ehsan, P. Wintersberger, Q. V. Liao, E. A. Watkins, C. Manger, H. Daumé III, A. Riener, M. O. Riedl, Human-centered explainable ai (hcxai): beyond opening the black-box of ai, in: *CHI conference on human factors in computing systems extended abstracts*, 2022, pp. 1–7.
- [15] D. Long, B. Magerko, What is ai literacy? competencies and design considerations, in: *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–16.
- [16] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, *Information*

fusion 58 (2020) 82–115.

- [17] S. Diefenbach, L. Christoforakos, D. Ullrich, A. Butz, Invisible but understandable: In search of the sweet spot between technology invisibility and transparency in smart spaces and beyond, *Multimodal Technologies and Interaction* 6 (2022) 95.
- [18] T. B. Sheridan, W. L. Verplank, T. Brooks, Human/computer control of undersea teleoperators, in: NASA. Ames Res. Center The 14th Ann. Conf. on Manual Control, 1978.
- [19] M. R. Endsley, From here to autonomy: lessons learned from human–automation research, *Human factors* 59 (2017) 5–27.
- [20] B. Shneiderman, Human-centered artificial intelligence: Reliable, safe & trustworthy, *International Journal of Human–Computer Interaction* 36 (2020) 495–504.
- [21] A. Fügener, J. Grahl, A. Gupta, W. Ketter, Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation, *Information Systems Research* 33 (2022) 678–696.
- [22] V. Trianni, A. G. Nuzzolese, J. Porciello, R. H. Kurvers, S. M. Herzog, G. Barabucci, A. Berditchevskaia, F. Fung, Hybrid collective intelligence for decision support in complex open-ended domains, in: *HHAI 2023: Augmenting Human Intellect*, IOS Press, 2023, pp. 124–137.
- [23] N. Emaminejad, R. Akhavian, Trustworthy ai and robotics: Implications for the aec industry, *Automation in Construction* 139 (2022) 104298.
- [24] T. B. Sheridan, T. B. Sheridan, K. Maschinenbauingenieur, T. B. Sheridan, T. B. Sheridan, *Humans and automation: System design and research issues*, volume 280, Human Factors and Ergonomics Society Santa Monica, CA, 2002.
- [25] K. A. Hoff, M. Bashir, Trust in automation: Integrating empirical evidence on factors that influence trust, *Human factors* 57 (2015) 407–434.
- [26] D. H. Autor, F. Levy, R. J. Murnane, The skill content of recent technological change: An empirical exploration, *The Quarterly journal of economics* 118 (2003) 1279–1333.
- [27] E. Brynjolfsson, T. Mitchell, D. Rock, What can machines learn and what does it mean for occupations and the economy?, in: *AEA papers and proceedings*, volume 108, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018, pp. 43–47.
- [28] I. Roll, E. S. Wiese, Y. Long, V. Aleven, K. R. Koedinger, Tutoring self-and co-regulation with intelligent tutoring systems to help students acquire better learning skills, *Design recommendations for intelligent tutoring systems* 2 (2014) 169–182.
- [29] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. H. Beck, Deep learning for identifying metastatic breast cancer, *arXiv preprint arXiv:1606.05718* (2016).
- [30] A. Jaiswal, C. J. Arun, A. Varma, Rebooting employees: Upskilling for artificial intelligence in multinational corporations, *The International Journal of Human Resource Management* 33 (2022) 1179–1208.
- [31] M. Xia, Co-working with ai is a double-sword in technostress? an integrative review of human-ai collaboration from a holistic process of technostress, in: *SHS Web of Conferences*, volume 155, EDP Sciences, 2023, p. 03022.
- [32] A. Zirar, Can artificial intelligence’s limitations drive innovative work behaviour?, *Review of Managerial Science* (2023) 1–30.
- [33] H. Dhiman, C. Wächter, M. Fellmann, C. Röcker, Intelligent assistants: Conceptual dimensions, contextual model, and design trends, *Business & Information Systems Engineering*



64 (2022) 645–665.

- [34] M. Nazar, M. M. Alam, E. Yafi, M. M. Su'ud, A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques, *IEEE Access* 9 (2021) 153316–153348.
- [35] S. Brdnik, Gui design patterns for improving the hci in explainable artificial intelligence, in: *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, 2023, pp. 240–242.
- [36] N. Van Berkel, M. B. Skov, J. Kjeldskov, Human-ai interaction: intermittent, continuous, and proactive, *Interactions* 28 (2021) 67–71.
- [37] P. Wintersberger, N. Van Berkel, N. Fereydooni, B. Tag, E. L. Glassman, D. Buschek, A. Blandford, F. Michahelles, Designing for continuous interaction with artificial intelligence systems, in: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–4.
- [38] C. O. Alm, A. Alvarez, J. Font, A. Liapis, T. Pederson, J. Salo, Invisible ai-driven hci systems—when, why and how, in: *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 2020, pp. 1–3.
- [39] B. E. Holthausen, P. Wintersberger, B. N. Walker, A. Riener, Situational trust scale for automated driving (sts-ad): Development and initial validation, in: *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2020, pp. 40–47.
- [40] Z. Lu, M. Yin, Human reliance on machine learning models when performance feedback is limited: Heuristics and risks, in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–16.
- [41] D. Keller, S. Rice, System-wide versus component-specific trust using multiple aids, *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology* 137 (2009) 114–128.
- [42] R. Vertegaal, et al., Attentive user interfaces, *Communications of the ACM* 46 (2003) 30–33.
- [43] C. Anderson, I. Hübener, A.-K. Seipp, S. Ohly, K. David, V. Pejovic, A survey of attention management systems in ubiquitous computing environments, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2 (2018). URL: <https://doi.org/10.1145/3214261>. doi:10.1145/3214261.
- [44] G. Guerrero, F. J. M. da Silva, A. Fernández-Caballero, A. Pereira, Augmented humanity: a systematic mapping review, *Sensors* 22 (2022) 514.
- [45] S. S. Kim, E. A. Watkins, O. Russakovsky, R. Fong, A. Monroy-Hernández, ” help me help the ai”: Understanding how explainability can support human-ai interaction, in: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–17.
- [46] S. Villa, J. Niess, T. Nakao, J. Lazar, A. Schmidt, T.-K. Machulla, Understanding perception of human augmentation: A mixed-method study, in: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023, pp. 1–16.
- [47] D. Talypova, A. Lingler, P. Wintersberger, User-centered investigation of features for attention management systems in an online vignette study, in: *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*, 2023, pp. 108–121.
- [48] T. De Greef, K. van Dongen, M. Grootjen, J. Lindenberg, Augmenting cognition: reviewing

- the symbiotic relation between man and machine, in: Foundations of Augmented Cognition: Third International Conference, FAC 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27, 2007. Proceedings 3, Springer, 2007, pp. 439–448.
- [49] I. Gabriel, Artificial intelligence, values, and alignment, *Minds and machines* 30 (2020) 411–437.
  - [50] N. F. Watson, C. R. Fernandez, Artificial intelligence and sleep: Advancing sleep medicine, *Sleep medicine reviews* 59 (2021) 101512.
  - [51] W. C. Xuan, P. Keikhosrokiani, Habitpad: A habit-change person-centric healthcare mobile application with machine learning and gamification features for obesity, in: Enabling Person-Centric Healthcare Using Ambient Assistive Technology: Personalized and Patient-Centric Healthcare Services in AAT, Springer, 2023, pp. 27–56.
  - [52] R. Raisamo, I. Rakkolainen, P. Majaranta, K. Salminen, J. Rantala, A. Farooq, Human augmentation: Past, present and future, *International Journal of Human-Computer Studies* 131 (2019) 131–143.
  - [53] A. A. Kruse, D. D. Schmorow, Session overview: Foundations of augmented cognition, *Foundations of augmented cognition* (2005) 441–445.