Human Factors Engineering in Explainable AI: Putting People First

Calvin Nobles
University Maryland Global Campus, Adelphi, USA

Calvin.nobles@umgc.edu

Abstract: This paper examines the integration of human factors engineering into Explainable Artificial Intelligence (XAI) to develop AI systems that are both human-centered and technically robust. The increasing use of AI technologies in high-stakes domains, such as healthcare, finance, and emergency response, underscores the urgent need for explainability, trust, and transparency. However, the field of XAI faces critical challenges, including the absence of standardized definitions and evaluation frameworks, which hinder the assessment and effectiveness of explainability techniques. Human factors engineering, an interdisciplinary field focused on optimizing human-system interactions, offers a comprehensive framework to address these challenges. By applying principles such as user-centered design, error management, and system adaptability, human factors engineering ensures AI systems align with human cognitive abilities and behavioral patterns. This alignment enhances usability, fosters trust, and reduces blind reliance on AI by ensuring explanations are clear, actionable, and tailored to diverse user needs. Additionally, human factors engineering emphasizes inclusivity and accessibility, promoting equitable AI systems that serve varied populations effectively. This paper explores the intersection of HFE and XAI, highlighting their complementary roles in bridging algorithmic complexity with actionable understanding. It further investigates how human factors engineering principles address sociotechnical challenges, including fairness, accountability, and inclusivity, in AI deployment. The findings demonstrate that the integration of human factors engineering and XAI advances the creation of AI systems that are not only technologically sophisticated but also ethically aligned and user-focused. This interdisciplinary synergy is a pathway to develop equitable, effective, and trustworthy AI solutions, fostering informed decision-making and enhancing user confidence across diverse applications.

Keywords: Artificial intelligence (AI), Explainable artificial intelligence (XAI), Human-Centered Artificial Intelligence (ACAI), Human factors, Human factors engineering (HFE)

1. Introduction

The rapid advancements in artificial intelligence (AI) have heightened concerns surrounding explainable artificial intelligence (XAI). Existing literature highlights the lack of standardized definitions and evaluation frameworks within the field of XAI, creating challenges in assessing the effectiveness of explainability techniques (Adadi & Berrada, 2018; Karimi et al., 2020; Rudin et al., 2021). This variability complicates the ability to draw meaningful conclusions about the efficacy of different approaches to explainability (Rudin, 2019). The Defense Advanced Research Projects Agency defines XAI as systems capable of providing human users with clear explanations of their decision-making processes, outlining system strengths and limitations, and offering insights into future behavior (Gunning & Aha, 2019). Such a definition underscores the critical role of explainability in fostering trust, transparency, and usability in AI systems.

The intersection of human factors engineering and XAI is pivotal in enabling users to understand how AI systems function and their decision-making capabilities. Human factors engineering is an interdisciplinary field that examines human capabilities and limitations to inform the design of devices, systems, and processes that optimize performance, safety, and usability (Anderson et al., 2010). By analyzing the intricate interactions between humans and their operational environments—including tools, technologies, and work systems—human factors engineering provides valuable insights into designing user-centered AI systems (Nobles & Robinson, 2024). Within the context of AI, these interactions are inherently complex, influencing human behavior and decision-making. Human factors engineering encompasses diverse subfields such as cognitive engineering, physical ergonomics, and human-computer interaction, each addressing specific aspects of human-system interactions (Anderson et al., 2010; Nobles & Robinson, 2024).

Despite its critical importance, human factors are often relegated to an afterthought in system design rather than treated as a central objective or a key enabler of effective solutions (Neumann et al., 2021). Though frequently framed as a technical challenge, explainability is equally human-centered (Ehsan et al., 2022). The predominantly algorithm-focused discourse surrounding XAI has overlooked the human dimensions of creating explainable systems (Ehsan & Riedl, 2020). Leveraging human factors engineering as a foundational framework in designing and evaluating XAI can address this gap by integrating principles that align AI systems with human needs and behaviors, as depicted in Figure 1 and Table 1. Incorporating human factors engineering principles can mitigate blind trust in AI systems, fostering a deeper understanding and more informed use of AI technologies among users.

The structure of this article is organized to provide a comprehensive examination of XAI and its intersection with human factors engineering. Section 2 establishes the foundational background and provides an in-depth discussion of XAI. Section 3 introduces human factors engineering and its associated principles, focusing on their application to XAI. Section 4 elaborates on Human-Centered AI (HCAI) and examines the nexus between human factors and XAI. Section 5 delves into the intersectionality of XAI and human factors engineering, highlighting their complementary roles in advancing AI systems. Finally, Section 6 presents the conclusion, summarizing key insights and implications.

2. Explainable Artificial Intelligence

According to Gunning (2016), XAI emphasizes the development of techniques that enhance user understanding, foster trust, and enable efficient management of modern AI systems. The growing prominence of AI and the use of complex machine learning algorithms to address intricate problems has highlighted a critical need for XAI. The lack of transparency in these systems makes it essential for end-users to understand clearly, as limited comprehensibility could hinder adoption in critical domains such as healthcare and cybersecurity (Saeed & Omlin, 2023). The ubiquitous use of AI and the increasing number of decisions made autonomously for end-users require greater explainability and transparency.

Some researchers argue that not all black-box AI systems require explanations for their decisions, as providing such explanations can increase development costs and reduce efficiency (Adadi & Berrada, 2018; Doshi-Velez & Kim, 2017). Explainability is typically unnecessary in two key scenarios: (1) when the outcomes have minimal impact and do not carry significant consequences, and (2) when the problem is well-understood, and the system's decisions are considered reliable, such as in applications like advertisement systems and postal code sorting (Adadi & Berrada, 2018; Doshi-Velez & Kim, 2017). Therefore, evaluating contexts where explanations and interpretations offer meaningful value (Adadi & Berrada, 2018; Doshi-Velez & Kim, 2017) is essential.

Existing literature highlights that explainability involves delivering insights tailored to a specific audience to address their needs. In contrast, interpretability measures how well those insights align with the audience's existing domain knowledge (Saeed & Omlin, 2023). Explainability is defined by three key components: (a) the insights provided, (b) the target audience, and (c) the specific needs those insights aim to address (Saeed & Omlin, 2023). Insights can be generated through various techniques, such as text explanations or feature relevance analyses, and are directed toward audiences, including domain experts, end-users, and modeling specialists (Saeed & Omlin, 2023). These insights include justifying decisions, uncovering new knowledge, refining AI models, and promoting fairness (Saeed & Omlin, 2023). Ultimately, explainability aims to enable audiences to effectively meet their objectives through the insights delivered (Saeed & Omlin, 2023). Explainability means providing clear, tailored information about how AI systems work to meet the needs of specific audiences, such as users or experts, for purposes like decision-making, learning, or ensuring fairness.

The National Institute of Standards and Technology outlines four foundational principles that underpin the concept of XAI. First, the principle of explanation mandates that an AI system should consistently provide accompanying evidence or rationale for its outputs and operational processes, ensuring transparency in decision-making (Phillips et al., 2021). Second, the meaningful principle emphasizes that explanations must be accessible and comprehensible to the system's intended audience, thereby bridging technical complexity with user understanding (Phillips et al., 2021). Third, explanation accuracy requires correctly representing the processes leading to the system's outputs and maintaining fidelity to the AI model's operations (Phillips et al., 2021). Finally, the knowledge limits principle asserts that the system should recognize and signal when functioning beyond its design parameters or lacks sufficient confidence in its output, safeguarding against inappropriate or unreliable responses in uncertain conditions (Phillips et al., 2021). These principles collectively aim to establish a foundation for trustworthy AI, particularly where the clarity and reliability of AI-driven decisions bear significant societal implications.

According to Gade et al. (2020), the need for XAI is approached through various perspectives, each highlighting different challenges and motivations for implementing XAI in black-box AI systems. Here is a breakdown of these perspectives (Gade et al., 2020):

- Regulatory Perspective: Black-box AI systems can lead to decisions with significant legal implications, necessitating a regulatory framework to ensure accountability. The EU's GDPR exemplifies this need by establishing a "right to explanation," allowing users to request justifications for algorithm decisions, such as loan rejections. However, the regulatory framework may be ineffective without effective technologies to provide these explanations.

- Scientific Perspective: Black-box AI models often approximate complex functions, representing knowledge derived from data rather than the data itself. XAI can uncover the scientific insights embedded within these models, potentially leading to discoveries across various scientific fields.
- Industrial Perspective: The complexity and opacity of black-box AI systems can lead to user distrust and industry regulatory challenges. As a result, industries may favor less accurate but more interpretable models. XAI can help bridge the gap between model performance and interpretability, although it may also increase development and deployment costs.
- Model Development Perspective: Understanding the decisions black-box AI systems make is crucial, significantly when they impact human lives. XAI can assist in diagnosing and improving these systems, enhancing their robustness and safety while minimizing biases and unfairness. It can also aid in model selection by revealing the features influencing decisions.
- End-User and Social Perspectives: Concerns about trust arise when black-box AI models produce incorrect classifications based on imperceptible changes to input data. Additionally, biases in training data can lead to unfair decisions. By providing explanations and enhancing interpretability, XAI can help build trust in these systems, ensuring they operate as intended and are fair in their decision-making.

Overall, these perspectives highlight the critical importance of XAI in overcoming the challenges of black-box AI systems by emphasizing transparency, accountability, and trust. Adopting a human-centered approach to XAI is essential to ensure that end-users actively engage with and contribute to improving these systems rather than relying on blind trust, which could pose significant risks.

Ehsan and Riedl (2024) recently coined the term explainability pitfalls (EPS), defined as unintended adverse effects arising from AI explanations that, without intention to deceive, may lead users to act against their own best interests or form incorrect perceptions about the capabilities and limitations of AI systems. Unlike dark patterns, which are carefully designed to mislead and manipulate, EPs are an inadvertent byproduct of adding explainability to AI systems, often due to a limited understanding of user interpretations (Ehsan & Riedl, 2024). Ehsan and Riedl (2024) argue that a deeper examination of EPs is critical as they can severely impact users' trust and reliance on AI systems, which are increasingly embedded in high-stakes fields like healthcare and criminal justice. Introducing this EP concept expands the current discourse in XAI by focusing on the harm that can arise even from well-intentioned design efforts, emphasizing the need for a more nuanced approach to designing explainable AI. Human factors engineering principles coupled with EP can potentially decrease the unintended consequences of AI by increasingly focusing on the design aspects.

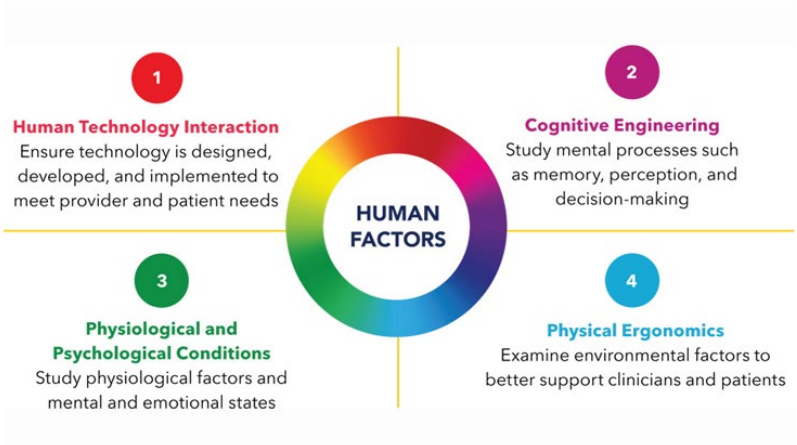
Applied research in human factors has developed a robust understanding of human traits, capabilities, and limitations, which is critical for designing systems that align with human characteristics (Anderson et al., 2010). Human factors engineers apply this knowledge to create systems that enhance usability, safety, and efficiency, enabling users to perform tasks accurately and effectively while minimizing errors (Anderson et al., 2010).

In the context of XAI, integrating human factors is essential. XAI aims to provide clear and understandable information about how AI systems work, ensuring that explanations are intuitive and aligned with the cognitive capabilities of the target audience, whether experts, general users, or developers. By leveraging human factors principles, XAI can design explanations that are accessible, actionable, and meaningful, enabling users to trust, understand, and effectively interact with AI systems. This alignment not only promotes optimal performance but also reduces cognitive overload and the likelihood of error, ensuring that AI systems meet the diverse needs of their users.

3. Human Factors Engineering

Human Factors Engineering is a multidisciplinary field dedicated to understanding human capabilities and limitations to design systems, devices, and environments that enhance overall performance and usability (Anderson et al., 2010). By harmonizing the interaction between individuals and their operational contexts, human factors engineering leverages a profound understanding of human strengths and constraints to optimize system functionality (Nobles & Robinson, 2024; PSN, 2019; Rogers & McGlynn, 2018). This discipline comprehensively analyzes task requirements, cognitive and physical demands, team dynamics, and environmental influences to promote safety, improve efficiency, and foster seamless user experiences (PSN, 2019). As illustrated in Figure 1, human factors serve as the scientific foundation for designing technologies that accommodate human needs while addressing inherent limitations. These principles ensure that technology aligns with human characteristics, facilitating effective and intuitive interactions.

Integrating human factors engineering principles into XAI necessitates a systematic and scientific approach that balances the alignment of AI systems with human users. The essence of this integration lies in designing systems that prioritize (a) actionable insights, (b) the targeted audience, and (c) the specific needs of users. Human factors engineering principles enhance XAI by creating AI systems that are technically robust and human-centered, fostering trust, transparency, and usability. By applying cognitive and ergonomic principles, human factors engineering ensures that XAI systems provide clear explanations, accommodate diverse user capabilities, and promote effective human-AI collaboration. This holistic approach underscores the critical role of human factors engineering in advancing XAI to support human decision-making and interaction in increasingly complex technological environments.



Source: MedStar Health (n.d.) https://www.medstarhealth.org/innovation-and-research/national-center-for-human-factors-in-healthcare

Figure 1: Human Factors

Human factors thrive as a multidisciplinary field, blending behavioral sciences, engineering, and physical sciences to foster a seamless integration between humans and the ever-evolving technological landscape. Human factors engineering is undoubtedly underrepresented in existing literature regarding AI and XAI. Chignell et al. (2023) point out the difference between human factors and human-computer interaction (HCI) and detail why human factors as a discipline are narrow in scope and could potentially underdeliver in addressing research demands in AI. Chignell et al. (2023) opine that human factors focus on theory-driven analysis, leveraging cognitive science and computational theory to study cognitive and performance aspects and designing interfaces that align with human capabilities.

In contrast, HCI takes an empirical approach, using contextual inquiry and surveys to develop and test user-centered interface designs (Chignell et al., 2023). While human factors explore theoretical aspects like trust and cognitive compatibility in AI, HCI emphasizes practical user experience, such as guiding users in creating strong passwords with AI (Chignell et al., 2023). Chignell underscores the synergy between the two fields, suggesting that human factors provide theoretical and cognitive insights. At the same time, HCI contributes practical tools and methods, both essential for developing human-centered AI systems. In truth, HCI and human factors are necessary to enhance XAI. HCI is vital for developing intuitive and explainable user interfaces; human factors could be essential in moving beyond blind trust, transparency, and interpretability.

Human factors principles systematically enhance the interaction between humans and systems by applying interdisciplinary methodologies that account for human capabilities, limitations, and behaviors. Rooted in the fields of psychology, engineering, and ergonomics, some of these principles are listed in Table 1.

Table 1: Human Factors Principles

Principles	Description	Reference
User-Centered Design	Systems and products should be designed with the user in mind, involving users throughout the design process to address their needs, abilities, and limitations.	Norman, D. A. (2013). The Design of Everyday Things. Basic Books.

Minimizing Cognitive Load	Interfaces and systems should be designed to minimize mental workload by simplifying tasks, organizing information logically, and providing clear instructions to prevent cognitive overload.	Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. E. (2004). An Introduction to Human Factors Engineering. Pearson Prentice Hall.
Consistency and Predictability	Systems should behave in predictable and consistent ways so that users can anticipate outcomes. This helps reduce errors and enhances user confidence.	Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. E. (2004). An Introduction to Human Factors Engineering. Pearson Prentice Hall.
Accessibility and Inclusivity	Systems should accommodate a wide range of users, including those with disabilities, by adhering to accessibility guidelines and designing for inclusivity.	Sanders, M. S., & McCormick, E. J. (1993). Human Factors in Engineering and Design. McGraw-Hill.
Error Prevention and Recovery	Design should minimize the possibility of user errors through clear labeling, intuitive controls, and feedback. Systems should also provide easy recovery options when errors occur.	Shneiderman, B., & Plaisant, C. (2010). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Addison-Wesley.
Psychosocial	Psychosocial needs are essential for effective sociotechnical systems, and the workplace's social environment—encompassing job demands, control, support, and satisfaction—significantly impacts physical and mental well-being.	Neumann et al. (2021). International journal of production economics, 233, 107992.
Simplicity and Clarity	Avoid unnecessary complexity by making information and interfaces straightforward, promoting user efficiency and satisfaction.	Norman (2013). The Design of Everyday Things.
Flexibility and Efficiency	Systems should allow users multiple ways to complete tasks and provide shortcuts for experienced users to improve efficiency and adaptability.	Shneiderman & Plaisant (2010). Designing the User Interface.
Feedback	Providing users with immediate and clear feedback ensures they know the results of their actions, helping them maintain control and understanding of the system's state.	Sanders & McCormick (1993). Human Factors in Engineering and Design.

Human factors principles are pivotal in optimizing XAI by ensuring systems are designed with a human-centered approach. While the list above of principles is not exhaustive, it underscores foundational tenets for fostering trust, enhancing transparency, and improving explainability, essential for aligning AI technologies with user expectations and workflows. Human factors prioritize user-centered design, requiring the active involvement of users throughout the design process to address their needs, abilities, and limitations (Norman, 2013). In XAI, human factors engineering can potentially reduce cognitive workload (Wickens et al., 2004) by presenting explanations in clear, logical, and concise formats, such as intuitive visualizations or natural language summaries, simplifying complex AI decision-making processes. Predictability and consistency (Wickens et al., 2024) in system behavior are equally critical, as they enable users to anticipate AI actions, build confidence, and minimize errors. Additionally, inclusive design (Sanders & McCormick, 1993) ensures that XAI systems are accessible to diverse users, including those with disabilities, by adhering to accessibility standards and fostering equitable usability. These considerations are integral to fostering fairness and trust in XAI systems, ensuring usability across varied demographics.

Another critical aspect of human factors in XAI is minimizing user errors and supporting effective error recovery. Intuitive controls, clear labeling, and actionable feedback are essential for guiding users and maintaining trust in the system, mainly when misunderstandings or discrepancies occur. For example, XAI systems should provide clear explanations and recovery options to mitigate errors and rebuild user confidence. Addressing psychosocial needs (Neumann et al., 2021) is also paramount, as workplace factors such as job demands, autonomy, and support (Norman, 2013) influence how users interact with XAI. Avoiding unnecessary complexity in system interfaces enhances adaptability and user satisfaction while offering flexible pathways and shortcuts for

experienced users fosters efficiency without compromising accessibility (Shneiderman & Plaisant, 2010). Immediate and clear feedback ensures users understand the outcomes of their actions and the system's state, reinforcing transparency and control (Sanders & McCormick, 1993). These principles not only enhance the comprehensibility and usability of XAI but also align the technology with human needs and cognitive capacities, promoting intuitive, trustworthy, and ethical interactions.

Moreover, human factors principles contribute to developing adaptive XAI systems that accommodate evolving user needs and operational contexts. Transparency is achieved by designing explanations that reveal how decisions are made and the system's limitations and assumptions, fostering trust through predictability and clear communication (Morrison et al., 2023). Ethical oversight and consistency further ensure that XAI systems provide uniform and unbiased explanations, mitigating risks of misinterpretation or bias. Human factors engineering also offers frameworks for optimal task allocation between humans and AI, enhancing collaboration and situation awareness while addressing the opacity often associated with AI decision-making (Grote, 2023). By aligning with users' mental models, human factors principles improve adoption and usability, particularly in applications requiring high levels of trust and transparency.

Finally, the diverse needs of XAI stakeholders underscore the importance of integrating human factors into system design. Businesses, regulators, consumers, and end-users require explainability for distinct purposes, including decision-making, fairness, regulatory oversight, trustworthiness, and reliability (Dwivedi et al., 2023). Addressing these multifaceted requirements demands deliberate integration of human factors principles, which provide a foundation for bridging the gap between humans and AI technologies. By fostering effective human-AI teaming, these principles enable creating systems that align with stakeholder objectives and promote seamless, trustworthy, and efficient interactions. The participatory ergonomics approach engages stakeholders throughout the design process and ensures that XAI systems are informed by diverse perspectives, resulting in functional and human-centered technologies (Grote, 2023). Through these efforts, human factors principles contribute to developing XAI systems that are robust, comprehensible, and aligned with their users' ethical and practical needs.

4. Human-Centered Artificial Intelligence (HCAI)

Many artificial intelligence (AI) systems were initially developed for low-risk applications, such as recommending movies or interpreting voice commands, where errors have minimal consequences (Barmer et al., 2021). However, when deployed in high-stakes contexts, such as determining mortgage approvals or guiding emergency responders during crises, the implications of errors can be profound (Barmer et al., 2021). Effective human-AI collaboration becomes essential in such critical scenarios, requiring trust, ethical accountability, and alignment with shared objectives to ensure reliable and safe outcomes.

Human-Centered Artificial Intelligence (HCAI) systems are designed to operate alongside humans in complex and dynamic environments where human decision-making plays a central role (Barmer et al., 2021). Despite their potential, the development of these systems is challenged by the rapid pace of AI advancements, the need for transparency in human-machine interactions, and the unpredictability of real-world operational contexts (Barmer et al., 2021). These issues are exacerbated by insufficient ethical guidelines and oversight, leading to risks such as bias, misuse, and unintended consequences (Barmer et al., 2021).

Barmer et al. (2023) highlight three critical focus areas for advancing HCAI to address these challenges. First, designers must deeply understand the intended use context and ensure that systems remain adaptable over time, maintaining clarity in their purpose and functionality (Barmer et al., 2021). Second, AI systems must foster effective human-machine teaming by building trust and transparency, enabling users to confidently interact with and understand system operations (Barmer et al., 2021). Third, continuous ethical oversight is necessary to mitigate risks and ensure AI systems' responsible development and deployment (Barmer et al., 2021). These considerations illustrate the transformative potential of HCAI to enhance technological capabilities and improve the human experience through ethical, transparent, and user-centered design (Maathuis, 2024; Riedl, 2019).

Human-centered approaches are pivotal in shaping XAI technologies, ensuring that technical solutions align with users' needs for explainability, thereby fostering trust and understanding in AI systems. These approaches advocate for designing and implementing XAI systems that prioritize the user's perspective, recognizing that effective explanations must be tailored to the knowledge and goals of their audience (Ehsan & Riedl, 2020; Ehsan et al., 2021; Liao & Varshney, 2021; Wang et al., 2021). Empirical studies involving real users are crucial for identifying limitations in existing XAI methods and prompting design interventions that challenge techno-centric paradigms (Ehsan & Riedl, 2020; Ehsan et al., 2021; Liao & Varshney, 2021; Wang et al., 2021). Moreover,

integrating theories from human cognition and behavior can inspire innovative computational and design frameworks, ultimately enhancing the usability and effectiveness of XAI applications (Liao & Varshney, 2021).

By emphasizing user interactions and experiences, human-centered approaches address the technical and cognitive challenges of explainability while contributing to a nuanced understanding of how individuals process and utilize explanatory information. This alignment bridges the gap between algorithmic explanations and actionable insights, enabling users to understand better and trust AI systems (Ehsan & Riedl, 2020; Ehsan et al., 2021; Wang et al., 2021). The human factors principles outlined in Table 1 serve as a foundational framework for moving beyond basic concepts of trust and transparency, fostering a more holistic approach to XAI that integrates human factors engineering to advance usability, reliability, and ethical accountability in AI systems.

5. Intersection of Explainable Artificial Intelligence and Human Factors Engineering

The intersectionality of human factors engineering and XAI represents a crucial nexus for advancing AI technologies that are both functional and human-centered. Human factors engineering emphasizes understanding human capabilities and limitations to design systems that optimize usability, performance, and safety. These principles directly apply to XAI, which aims to provide clear, intuitive, and actionable explanations of AI decision-making processes. By integrating human factors engineering principles, XAI systems can better align with human cognitive abilities, reducing mental workload and fostering trust and transparency. For instance, designing interfaces that use concise visualizations or natural language explanations ensures that users can comprehend complex AI mechanisms regardless of expertise. Furthermore, predictable and consistent system behavior, as emphasized in human factors engineering, enables users to anticipate AI actions and outcomes, thereby minimizing errors and reinforcing confidence in AI systems. This synergy ensures that XAI addresses technical explainability and creates a seamless, accessible experience for diverse user groups, promoting equitable and inclusive AI applications.

Incorporating human factors engineering into XAI also addresses critical ethical and operational challenges associated with deploying AI in high-stakes environments. HCAI systems, which prioritize human decision-making and collaboration, exemplify this intersectionality by integrating human factors engineering principles such as user-centered design and participatory ergonomics. These approaches ensure that stakeholders' needs—whether businesses, regulators, or end-users—are considered throughout the AI development lifecycle, aligning system functionality with real-world requirements. Additionally, human factors principles such as error management and adaptive system design are pivotal in mitigating risks such as bias, misuse, and operational failures in XAI. Continuous ethical oversight, another fundamental tenet of human factors engineering, reinforces the transparency and accountability of AI systems, ensuring that explanations are meaningful and reliable. By bridging the technical aspects of AI with human-centric considerations, the intersection of human factors engineering and XAI fosters the development of AI systems that meet technological standards and resonate with users, driving trust, adoption, and ethical alignment across diverse applications.

6. Theoretical and Practical Implications of Human Factors Engineering in XAI

Human factors engineering is essential for advancing XAI by ensuring AI systems are user-centric, adaptive, and ethically transparent. Theoretical contributions include user-centered design, which aligns AI with human cognitive processes to enhance usability (Hoffman et al., 2018), and error management strategies that mitigate user challenges, fostering trust and efficiency (Wang et al., 2019). HFE also promotes system adaptability, making AI accessible across diverse user contexts (Amershi et al., 2019). Ethical considerations such as transparency, accountability, and fairness remain critical in ensuring responsible AI deployment (Miller, 2019). These principles reinforce the necessity of integrating HFE into XAI to build reliable, user-focused AI systems.

Practically, human factors engineering enhances AI interaction by improving user interfaces for clarity, facilitating training to support informed usage, and implementing feedback mechanisms that refine AI explanations (Raees, 2024). Ensuring predictable system behavior reinforces user trust, while robust error mitigation strategies minimize misinterpretations (Raji et al., 2020). Adaptive explanations further personalize AI interactions, making systems more effective in real-world applications (Endsley, 2021). By integrating these principles, HFE ensures that XAI technologies remain transparent, intuitive, and aligned with human needs.

7. Future Research Direction

Future research in human factors engineering within XAI should emphasize empirical validation and real-world applicability to improve AI system usability and trust. Empirical evaluations of user-centered design methodologies across high-stakes domains such as healthcare and finance are essential to understanding the

impact of design principles on user trust, comprehension, and decision-making (Hoffman et al., 2018). Longitudinal studies on trust dynamics in AI can provide insights into how user reliance evolves over time, particularly in decision-critical environments (Polyportis, 2024). Additionally, inclusive design frameworks must be explored to ensure AI accessibility for diverse and marginalized populations, mitigating algorithmic biases in deployment (Raji et al., 2020).

Research on adaptive error recovery mechanisms in XAI could improve user confidence by refining feedback interventions that assist in understanding and correcting AI-induced errors (Wang et al., 2019). Case studies evaluating human factors engineering principles in real-world applications—such as emergency response and medical diagnostics—would offer practical insights into the tangible benefits of user-centered AI (Amershi et al., 2019). Furthermore, interdisciplinary collaboration among HFE, cognitive science, and AI engineering is necessary to develop frameworks that align AI decision-making with human cognitive and operational capabilities (Endsley, 2021). Finally, addressing ethical and accountability challenges in AI deployments is critical for ensuring meaningful and responsible explainability, reinforcing transparency and equitable use across applications (Miller, 2019). Advancing these research directions will facilitate the integration of HFE principles into XAI, fostering AI technologies that are both technically robust and deeply human-centered.

8. Conclusion

In conclusion, the integration of human factors engineering into XAI represents a critical advancement in the design and deployment of AI systems that are both functional and human-centered. The rapid growth of AI technologies has heightened the need for explainability, trust, and transparency, as emphasized by DARPA's definition of XAI. However, the lack of standardized definitions and evaluation frameworks poses significant challenges in assessing explainability techniques' efficacy. Human factors engineering provides a robust interdisciplinary framework to address these challenges by examining human capabilities, limitations, and behaviors to inform system design. By leveraging human factors principles—such as user-centered design, error management, adaptability, and others listed in Table 1—XAI systems can align with human cognitive processes, fostering more intuitive and trustworthy interactions. This alignment not only improves usability but also mitigates blind trust and encourages informed user engagement with AI technologies.

The intersectionality of human factors engineering and XAI underscores the importance of treating explainability as both a technical and human-centered issue. As highlighted in this discussion, human factors engineering's diverse subfields, including cognitive engineering and human-computer interaction, offer valuable insights into designing systems that prioritize user needs and ethical considerations. By integrating these principles, XAI systems can bridge the gap between algorithmic complexity and actionable understanding, ensuring that explanations are clear, relevant, and aligned with user contexts. Furthermore, embedding human factors engineering in the development of XAI addresses the broader sociotechnical challenges associated with AI, such as fairness, transparency, and accountability. Ultimately, the collaborative application of human factors engineering and XAI principles advances the creation of AI systems that enhance technological capabilities and uphold the ethical and practical demands of diverse stakeholders, contributing to a more equitable and effective AI landscape.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). *Guidelines for human-AI interaction*. Proceedings of the CHI Conference on Human Factors in Computing Systems.

Anderson, J., Gosbee, L., Bessesen, M., & Williams, L. (2010). Using human factors engineering to improve the effectiveness of infection prevention and control. *Critical Care Medicine*, 38, S269-S281. https://doi.org/10.1097/CCM.0b013e3181e6a058.

Barmer, H., Dzombak, R., Gaston, M., Palat, V., Redner, F., Smith, C., & Smith, T. (2021). Human-Centered AI. *IEEE Pervasive Comput.*, 22, 7-8. https://doi.org/10.1184/R1/16560183.V1.

Carroll, J. M. (1987). Interfacing thought: Cognitive aspects of human-computer interaction. The MIT Press.

Carroll, J. M., & Campbell, R. L. (1986). Softening up hard science: reply to Newell and Card. *Human-Computer Interaction*, *2*(3), 227-249.

Chignell, M., Wang, L., Zare, A., & Li, J. (2023). The evolution of HCI and human factors: Integrating human and artificial intelligence. *ACM Transactions on Computer-Human Interaction*, *30*(2), 1-30.

Defense Innovation Board. (2019). AI principles: Recommendations on the ethical use of artificial intelligence by the Department of Defense. Https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PD

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), 1–33. https://doi.org/10.1145/3561048

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Towards a reflective sociotechnical approach. In *HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22* (pp. 449-466). Springer International Publishing.

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., Wachter, S., ... & Riedl, M. O. (2021, May). Operationalizing human-centered perspectives in explainable AI. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1-6).

Ehsan, U., Wintersberger, P., Liao, Q. V., Watkins, E. A., Manger, C., Daumé III, H., ... & Riedl, M. O. (2022, April). Human-Centered Explainable AI (HCXAI): beyond opening the black-box of AI. In *CHI conference on human factors in computing systems extended abstracts* (pp. 1-7).

Ehsan, U., & Riedl, M. O. (2024). Explainability pitfalls: Beyond dark patterns in explainable AI. *Patterns*, *5*(6).

Endsley, M. R. (2021). Situation awareness. *Handbook of human factors and ergonomics*, 434-455.

Gade, K., Geyik, S., Kenthapadi, K., Mithal, V., & Taly, A. (2020, April). Explainable AI in industry: Practical challenges and lessons learned. In *Companion Proceedings of the Web Conference 2020* (pp. 303-304).

Grote, G. (2023). Shaping the development and use of artificial intelligence: How human factors and ergonomics expertise can become more pertinent. *Ergonomics*, 1-9. https://doi.org/10.1080/00140139.2023.2278408.

Gunning, D. (2016). Broad agency announcement explainable artificial intelligence (XAI). *Defense Advanced Research Projects Agency (DARPA), Tech. Rep.*

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for explainable AI: Challenges and prospects*. arXiv preprint arXiv:1812.04608.

Karimi, A. H., Barthe, G., Scholkop, B., & Valera, I. (2020). A survey of algorithmic recourse: Definitions, formulations, solutions, and prospects. arXiv preprint arXiv:2010.04050.

Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable AI (XAI): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790*.

Maathuis, C. (2024, March). Human-Centered AI in Military Cyber Operations. In *International Conference on Cyber Warfare and Security* (Vol. 19, No. 1, pp. 121-128).

Miller, T. (2019). *Explanation in artificial intelligence: Insights from the social sciences*. Artificial Intelligence, 267, 1-38.

Morrison, K., Shin, D., Holstein, K., & Perer, A. (2023). Evaluating the impact of human explanation strategies on Human-AI Visual Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, 7, 1 - 37. https://doi.org/10.1145/3579481.

Neumann, W. P., Winkelhaus, S., Grosse, E. H., & Glock, C. H. (2021). Industry 4.0 and the human factor–A systems framework and analysis methodology for successful development. *International journal of production economics*, *233*, 107992.

Nobles, C., & Robinson, N. (2024). The benefits of human factors engineering in cybersecurity. *Cybersecurity Risk Management: Enhancing Leadership and Expertise*, 53.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Phillips, P. J., Hahn, C. A., Fontana, P. C., Yates, A. N., Greene, K., ... & Przybocki, M. A. (2021). Four principles of explainable artificial intelligence. NIST Publication NISTIR 8312.

Patient Safety Network (PSN). (2019, September 07). Human factors engineering. https://psnet.ahrq.gov/primer/human-factors-engineering

Polyportis, A. (2024). A longitudinal study on artificial intelligence adoption: understanding the drivers of ChatGPT usage behavior change in higher education. *Frontiers in Artificial Intelligence*, *6*, 1324398.

Raees, M., Meijerink, I., Lykourentzou, I., Khan, V. J., & Papangelis, K. (2024). From explainable to interactive AI: A literature review on current trends in human-AI interaction. *International Journal of Human-Computer Studies*, 103301.

Raji, I. D., Smart, A., White, R. N., & Mitchell, M. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency.

Rogers, W. A., & McGlynn, S. A. (2018). Human factors and ergonomics: History, scope, and potential. In Human Factors and Ergonomics for the Gulf Cooperation Council (pp. 1–20). CRC Press.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2021). Interpretable machine learning: Fundamental principles and 10 grand challenges. arXiv preprint arXiv:2103.11251.

Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273.

Sanders, M. S., & McCormick, E. J. (1993). *Human Factors in Engineering and Design*. McGraw-Hill 7th Edition.

Salvendy, G., & Karwowski, W. (Eds.). (2021). Handbook of human factors and ergonomics. John Wiley & Sons.

Sanneman, L., & Shah, J. A. (2022). The situation awareness framework for explainable AI (SAFE-AI) and human factors considerations for XAI systems. *International Journal of Human–Computer Interaction*, *38*(18-20), 1772-1788.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India.

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon Becker, S. E. (2004). Decision making. *An introduction to human factors engineering*, 156-183.