

Guidelines for Human-AI Interaction

Michael Christopher – S224830467

A basic framework for understanding and improving human-AI interaction is offered by the 2019 study "Guidelines for Human-AI Interaction" by Amershi et al. Based on extensive empirical research and expert analysis, the authors offer 18 recommendations to help practitioners create transparent, helpful, and efficient AI systems. Four major categories that match the stages of human-AI interaction make up the guidelines' structure: (1) initial interaction to help users determine what the system can and cannot do; (2) system interaction that provide prompt, context-sensitive feedback; (3) error handling that helps users recover when the system malfunctions; and (4) system evolution to improve user efficiency and trust over time by learning from interactions. This method emphasizes the need to build AI systems for human-centred usability, reliability, and adaptability in addition to technical performance. The guidelines, for instance, emphasize the significance of establishing expectations early on, giving justifications for choices, and enabling users to challenge or overrule results (Amershi et al., 2019).

Building on this framework, Li, Vorvoreanu, DeBellis, and Amershi (2023) used factorial surveys to empirically evaluate the recommendations and analyse the impact of early design choices on user impressions of AI systems. Their results confirm the original guidelines' applicability while emphasizing that certain rules have a greater impact on satisfaction and trust than others. It was discovered that user confidence in AI systems was significantly influenced by criteria related to explainability and error mitigation. This implies that practitioners should place equal emphasis on communicating justification and ambiguity to users as well as technical accuracy. The idea that human-AI guidelines should be an essential part of development cycles rather than an afterthought is further supported by Li et al.'s (2023) demonstration that implementing the standards early in the design phase can avoid expensive usability concerns later.

Through a comparative investigation of industry recommendations for human-AI engagement, Wright et al. (2020) add to this conversation. Their work demonstrates a great deal of overlap among many companies, with many of them reiterating the fundamental ideas presented by Amershi et al. (2019), including error recovery, feedback, and transparency. Wright et al. (2020) also point out that different industries place different priorities. This comparative viewpoint emphasizes that although the principles provide a solid basis, their implementation needs to be designed for user scenarios and domains.

By investigating how practitioners use and apply guidelines in practical projects, specifically, through the People + AI Guidebook, Yildirim et al. (2023) elaborate on the

conversation. Their case study shows that although practitioners appreciate standards, they frequently struggle with transforming impersonal ideas into workable design choices. The need for more tangible tools, templates, and examples to enable various teams with different levels of competence to use guidelines is highlighted by the difference between theory and reality.

Lastly, Abedin et al. (2022) adopt a managerial viewpoint, stressing that managing and creating human-AI interactions is an organizational challenge as much as a technological or design one. They believe that finding a balance between the strategic, ethical, and operational aspects is necessary for successful implementation. For instance, companies need to think about bigger implications like accountability, governance, and long-term human-AI collaboration strategies in addition to user trust at the interaction level.

Taken together, these findings support the lasting value of Amershi et al.'s (2019) principles while demonstrating their adaptability and evolving relevance in industry, research, and practice. The standards continue to affect how practitioners and organizations envision trustworthy and successful AI, and their inclusion into both design and management practices remains important for guaranteeing AI systems that are useful, ethical, and aligned with human needs.

References:

Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers*, 24(3), 691-697.

Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).

Li, T., Vorvoreanu, M., DeBellis, D., & Amershi, S. (2023). Assessing human-AI interaction early through factorial surveys: a study on the guidelines for human-AI interaction. *ACM Transactions on Computer-Human Interaction*, 30(5), 1-45.

Wright, A. P., Wang, Z. J., Park, H., Guo, G., Sperrle, F., El-Assady, M., ... & Chau, D. H. (2020). A comparative analysis of industry human-AI interaction guidelines. *arXiv preprint arXiv:2010.11761*.

Yildirim, N., Pushkarna, M., Goyal, N., Wattenberg, M., & Viégas, F. (2023, April). Investigating how practitioners use human-ai guidelines: A case study on the people+ ai guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).