HHAI 2023: Augmenting Human Intellect
P. Lukowicz et al. (Eds.)

doi:10.3233/FAIA230119

A Multi-Dimensional Development and Deployment Framework for Hybrid Intelligence

Jacob SHERSON[a], Baptiste RABECQ[a], Dominik DELLERMANN[b], and Janet RAFNER[a]
[a]Center for Hybrid Intelligence, Dep. of Management, Aarhus University
[b]Vencortex

Abstract. This paper introduces a novel multi-dimensional framework for developing and deploying AI-human systems, incorporating both technical and managerial design principles. The paper then applies the framework to four standard human-AI interaction patterns, including Human Out Of the Loop (HOOTL), Human On the Loop (HOTL), Human In the Loop (HITL), and Hybrid Intelligence (HI). The dimensions are used to succinctly describe the essential characteristics of each pattern, highlighting potential risks and benefits, such as end-user resistance, employee deskilling, value-misalignment and employee upskilling and business model reengineering. The framework provides a valuable tool for AI developers and managers to characterize their current solutions and optimize the integration of humans and machines in complex systems.

Keywords. Hybrid Intelligence, Human-Centered AI, Human-in-the-loop AI, organizational learning

1. Introduction

The underwhelming results in up to 90% of corporate AI projects [1], along with high-profile failures of several autonomous AI systems [2], has prompted a greater focus on incorporating humans into the AI-systems loop. This has led to the development of detailed technical frameworks and design guidelines within related fields such as Human-Centered AI (HCAI) [3, 4] and Hybrid Intelligence (HI) [5, 6]. Recently, HI has gained traction as a holistic development and deployment approach in both consultancies [7] and academia [8], merging advancements in computer science with established management research practices. In management research, human-AI interaction has so far primarily been explored along selected pairs of dimensions such as the degree of openness of the process and the risk of severe failures [9] and parallel-sequential task allocation and degree of human-AI communication constraints [10]. In this paper, we introduce a novel multi-dimensional framework for AI-human systems development and deployment by integrating technical and managerial systems design principles. We first outline the framework, followed by a comparison of a selection of standard human AI interaction patterns within the proposed structure.

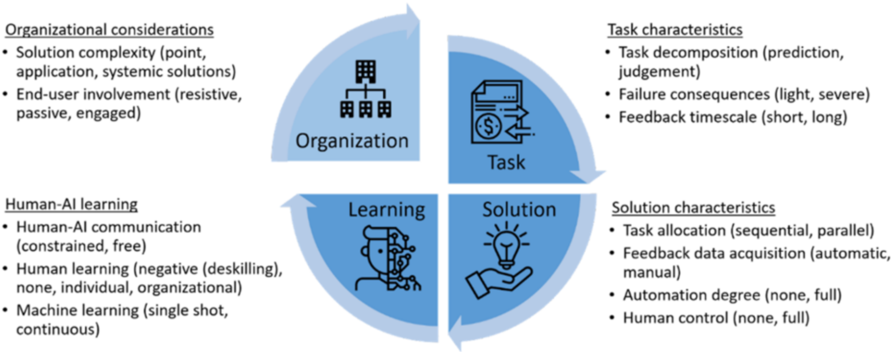

Figure 1. The 12 chosen dimensions of human-AI systems. Often the organizational perspective is left out of the AI-design considerations, which poses an increasing challenge for the design of real-world, highly integrated, human-centered systems. Dimensions are either explicitly defined in the text or through the examples below. Parentheses indicate the 2-4 defining points along each continuously defined dimension.

2. The AI Design and Deployment Dimensions (AI-DDD)

Most prior technical frameworks frequently focus on i) task specifications to generate ii) proposed solutions which then may lead to iii) varying levels of human and AI learning and integration. Although often neglected, ideally, negative learning outcomes (machine failures and human deskilling as well as positive ones (employee learning and upskilling) should prompt systemic (re)evaluations of organizational processes and business value streams or such considerations should be considered upfront in the design process [11]. Furthermore, Agrawal et al. argue [12] that many AI-deployment failures stem from a focus on point solutions in which the AI is inserted independently without changing systems and procedures. Instead, they argue for systems solutions which generate or utilize novel value streams by simultaneously revising a set of interdependent procedures and business units (see AI systems discovery canvas of Ref. [12]). To account for these considerations, we introduce a fourth category containing organizational dimensions.

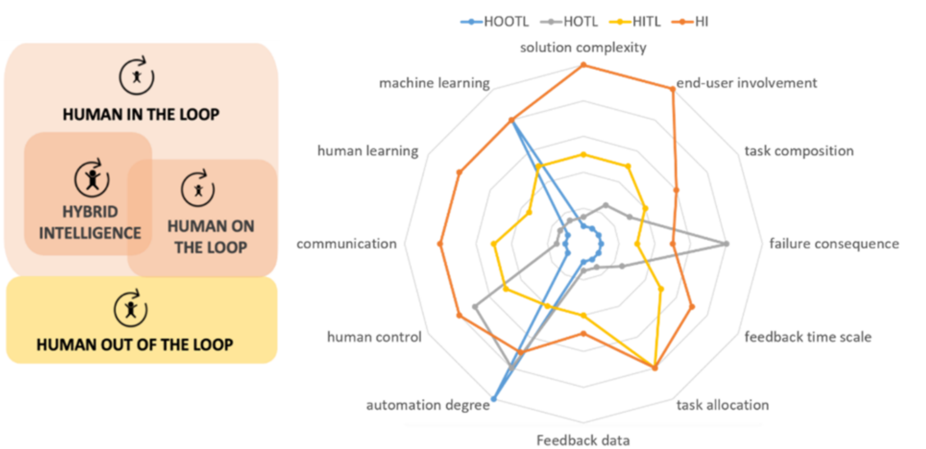

Figure 2. Left: illustration of the relation of four standard human-AI interactions patterns. **Right:** Radar plots along the 12 AI-DDD dimensions for the four interaction patterns

3. Applying AI-DDD to human-AI interaction patterns

In the following, we demonstrate how the dimensions can be applied to succinctly describe the essential characteristics (underlined) of four canonical human-AI interaction patterns. Dimensions that are not mentioned could take on various values depending on the concrete application.

Human Out Of the Loop (HOOTL) applications typically involve a point solution consisting of a pure prediction task with perceived low failure consequences has been fully automated. Machine learning may either consist of the application of a pre-trained model or *may*, if feedback time scales are low and feedback acquisition is automated, enable continuous training and evolution of the model resulting in increasingly efficient fulfillment of the optimization objectives. If the prediction is combined with automated decision-making, there is a substantial risk of end-user resistance and catastrophic unintended outcomes [13].

Human On the Loop (HOTL) applications involve a heavily automated process that has sufficiently severe perceived failure consequences that it is deemed necessary to have a human operator in control to perform human judgment of the output of the machine computation. This human involvement increases feedback time scale in operation compared to HOOTL applications but may lead to less risk of end-user resistance.

Human In the Loop (HITL) applications involve human judgment integrated continuously within the task allocation. Without an explicit attention to organizational aspects, the allocation of tasks from human to machine often risks resulting in substantial employee deskilling [11].

Hybrid Intelligence (HI) applications are a subset of HITL in which the technical human-AI interface is developed according to the principles of human-centered AI in which high degrees of sub-task automation and human control are pursued simultaneously. Both continuous human and machine learning are facilitated by transparent human-AI communication. Typically such highly integrated solutions can only be developed with a high degree of end-user involvement and with deliberate attention to acquiring feedback data (e.g. through deliberate experiments). Such integrated solutions often require systemic organizational deployment across multiple business units (e.g. AI system discovery canvas of Ref [12]) in order to optimally monetize the potential business process reengineering enabled by the employee upskilling (individual learning) as well as organizational learning.

References

[1] Boston Consulting Group. (2020, October 20). Study finds significant financial benefits with AI. BCG. https://www.bcg.com/press/20october2020-study-finds-significant-financial-benefits-with-ai

[2] Heaven, D. (2019). Deep trouble for deep learning. Nature, 574(7777), 163-166.

[3] Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. AIS Transactions on Human-Computer Interaction, 12(3), 109-124.

[4] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19) (pp. 1-13). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300233

[5] Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., ... & Welling, M. (2020). A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer, 53(08), 18-28.

[6] Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. Business & Information Systems Engineering, 61(5), 637-643.
[7] McKinsey & Company. (n.d.). Hybrid intelligence: The future of artificial intelligence. New at McKinsey Blog. Retrieved April 17, 2023, from https://www.mckinsey.com/about-us/new-at-mckinsey-blog/hybrid-intelligence-the-future-of-artificial-intelligence
[8] Rafner, J., Bantle, C., Dellermann, D., Söllner, M., Zaggl, M. A., & Sherson, J. (2022). Towards Hybrid Intelligence Workflows: Integrating Interface Design and Scalable Deployment. In HHAI2022: Augmenting Human Intellect (pp. 310-313). IOS Press.
[9] Saenz, M. J., Revilla, E., & Simón, C. (2020). Designing AI systems with human-machine teams. MIT Sloan Management Review, 61(3), 1-5
[10] Puranam, P. (2021). Human–AI collaborative decision-making as an organization design problem. Journal of Organization Design, 10(2), 75-80.
[11] Rafner, J., Dellermann, D., Hjorth, A., Verasztó, D., Kampf, C., Mackay, W., & Sherson, J. (2022). Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence. Morals & Machines, 1(2), 24-39.
[12] Agrawal, A., Gans, J., & Goldfarb, A. (2022). Power and Prediction: The Disruptive Economics of Artificial Intelligence. Harvard Business Review Press. ISBN-10:1647824192
[13] Soares, N. (2018). The value learning problem. In *Artificial intelligence safety and security* (pp. 89-97). Chapman and Hall/CRC.