

Enriching human–AI teaming based on risk envelopment and teammates’ inherent capabilities

Antti Salovaara^{1,*}, Virpi Roto¹

¹Department of Design, Aalto University, Finland

Abstract

With development of new technologies, humans have been able to continuously automate larger amounts of work. The recent developments of AI offer again new opportunities for such automation. This time, however, the situation is different, as it has become possible to consider technologies as teammates instead of only as tools. This demands for conceptualization of such a human–AI teaming and better understanding of the ways in which risks and efficiency can be balanced, as well as how inherent strengths and weaknesses of the involved parties can be orchestrated. In this paper, we build on our earlier considerations how this may be achieved in a way that would ensure work enrichment for the human teammates, and thus keep work meaningful and interesting for people even with highly skilled AI teammates.

Keywords

Automation, Human–AI teaming, Envelopment, Work enrichment

1. Introduction

While many models and proposals for human–AI interaction and teaming have been presented in the recent years (e.g., [1, 2]), and many older ones (e.g., [3, 4, 5]) also are still seen as valid, researchers are still seeking for a synthesizing picture into the best ways to describe successful teaming between humans, automation, robots, and AI. Instead of a synthesis, the recent development in AI have brought about a new influx of richer ways to consider these relationships.

In this paper, we wish to offer more views into this exciting research area that is continuously under development, with a wish for contributing to an emerging synthesis in the future. We are interested in task and responsibility allocations that on one hand would elevate human potential and meaningfulness in teams comprising humans and AIs, and in finding ways for using AIs for their best benefit while offsetting the related dangers and risks. We thus seek to present tentative answers to the following research question: What are the task/responsibility configurations in human–AI teaming that a) provide meaningful experiences at work for humans and b) offer balancing principles for AI’s strengths and weaknesses?

2. Background

In this paper, we consider AI as a new, more competent form of automation – one which has higher agency, representational capacity, and ability for interactive cooperation. In the following sub-sections, we first review what literature has suggested about task allocations between humans and automation, and then have a closer look at the recent works on human–AI interaction.

2.1. Humans’ and AI’s inherent weaknesses and strengths

One of the first recommendations for task division between humans and automation can be found from 1950s when Fitts [3] presented the “humans are better at – machines are better at” (HABA-MABA) lists

AutomationXP25: Hybrid Automation Experiences, April 27, 2025, Yokohama, Japan. In conjunction with ACM CHI’25.

*Corresponding author.

✉ antti.salovaara@aalto.fi (A. Salovaara); virpi.roto@aalto.fi (V. Roto)

ORCID 0000-0001-7260-8670 (A. Salovaara); 0000-0002-6880-5636 (V. Roto)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Table 1

Paul Fitt's list for tasks where humans and machines are better than the other party.

Humans are better at	Machines are better at
Detecting small amounts of visual, auditory, or chemical energy	Applying great force smoothly and precisely
Perceiving patterns of light or sound	Responding quickly to control signals
Improvising and using flexible procedures	Reasoning deductively
Storing information for long periods of time and recalling appropriate parts	Storing information briefly, erasing it completely
Reasoning inductively	
Exercising judgment	

in a report on air navigation and traffic control (see Table 1).

A much more recent presentation on human–automation task division is by Parasuraman et al. [5] whose proposal has four information processing steps (information acquisition, information analysis, decision/action selection and action implementation), each which can involve different levels of automation (e.g., in decision/action selection, from a stage where a human makes all decisions through automation-executed actions after human approval to a level where the automation performs all decisions and ignores humans). This work has been widely cited especially in human factors research.

The increasing capacity of machines to assume tasks that earlier have been humans' responsibility (especially with the last decade's breakthroughs with AI) have motivated researchers to suggest orchestrations by which the new forms of agency in AI's operation can be harnessed successfully but safely.

Salovaara et al. [6] presented a three-layered organisational model by which organisations in high-risk environments (malware protection in the case of their study) can automate virus protection processes in customers' computers while also making it amenable for rapid updates and reconfigurations should errors occur or new detection require additions. The model builds on the differences between mindful and mindless activities [7]. The lowest level in their model is fully automated and takes care of "mindless" activities: it analyses operations within the computer's communications and internal processes, and takes swift action if something doubtful is detected. The second layer is human-operated, mindful and thus slow. It is a reactive layer that responds to new virus families, and continuously updates the lowest layer's automations. The third layer is anticipatory and research oriented, and tries to identify new threats, so that the second layer can respond to new threats before they become epidemics. This way, automation's performance potential can be harnessed maximally yet its processes can be adapted based on changing needs.

Another way to maximise AI's power while curbing its risks is to establish "envelopes" around AI's training data, training method, permitted outputs, and tasks that it is recruited to solve [8]. The term "envelope" is adopted from shopfloor automation, where robot arms in assembly lines have 3D no-go areas around them, to prevent fatal injuries [9]. With such envelopes, organisations may be able to deploy powerful yet inscrutable black-box AI systems to tasks: they know that they will not be used in decisions that are too risky, and will not suggest actions that would be severely erroneous.

Also human–AI interactions have been an active topic of research. Microsoft researchers' guidelines for designing AI-infused systems [1], for example, are a widely cited source. It presents 18 guidelines for 4 different interaction situations ("initially", "during interaction", "when wrong" and "over time") in human–AI interactions. The guidelines suggest, for example, that AI should remember earlier interactions with its human teammates, so that it can adapt to it over time.

However, what has remained less studied is the needs that human workers will have for the future work with AI teammates. The focus has been heavily on the capabilities of AI, while the human teammates' psychological needs and best cognitive capabilities have been understudied. We therefore review some of the much fewer works in these topics in the following sub-section.

2.2. Requirements inherent to humans

As sentient teammates, humans need to be considered differently than machines. At minimum, they should have decent work conditions and fair income. According to International Labour Organization (<https://www.ilo.org/topics/decent-work>), every worker should have security in the workplace, social protection, better prospects for personal development and social integration, freedom to express their concerns, organize and participate in the decisions that affect their lives and equal opportunities and treatment for all women and men. AI colleagues influence many of these aspects, such as security, social protection, personal development, social integration, and participating in decision making. Thus, when designing ethical AI co-workers, these aspects should be addressed.

3. Enriched human–AI teaming

We find that much of what Fitts [3] wrote about relative strengths of humans and machines is still valid today, even if automation nowadays includes systems that are able to learn autonomously. What has admittedly changed, however, is AI's ability to demonstrate behaviours (e.g., pattern recognition and use of natural language) that previously was considered solely as a territory that humans master.

One of the underlying principles by which Fitts's two lists can be distinguished from each other seems to be the mindlessness of actions that machines can perform, and mindfulness that humans are capable of (occasionally, at least). This distinction still holds today: even if machines can perform even superhuman tasks cognitively, they are not capable of such self-reflection that mindful actions presuppose.

In the absence of self-reflection and mindfulness, the above-presented envelopment-oriented organization-level techniques [8] may be amenable also for team-level human–AI cooperation. Also on a team level, “boundary envelopment” can be applied to AIs' tasks so that they only perform tasks where errors are non-critical and where their use is socially acceptable.

In some cases, however organisational envelopment is or inadequate or insufficient for team-level cooperation. Not every team trains their own models, so training-data envelopment, for example, is unlikely to be applicable. Also, new cloud-based large language model AIs introduce data privacy and hallucination/fabrication risks, calling for consideration whether new forms of team-level envelopment could be identified and operationalized. Currently precautions for these two risks rest too much on the shoulders of human teammates.

Also the needs of human teammates need more attention. For example, an AI-induced robotic co-worker must be designed to address employees' safety. Traditionally, robotic tools are placed in isolated areas to avoid accidents, but when robots become colleagues, it would be nicer to have them safely among the other workers. Another requirement for decent work is the possibility for workers to participate in decision making, especially when it relates to their own work. While it may be difficult to interfere to the processes of deep neural networks, designers should allow people an opportunity to influence their own work task allocation. Following the decent work requirements in AI design makes AI a well-behaving co-worker.

Decent work is the minimum requirement for designing human–AI teamwork. It prevents harm at work but does not address aspects that make work intrinsically motivating. According to Roto [10], addressing the requirements of decent work is the first level of work enrichment. Work enrichment is the counteract of work simplification, which often happens with industrial AI. To avoid future work being passive monitoring of automated processes, work enrichment introduces activities that address employees' timely pragmatic and/or psychological needs and makes work worthwhile. People are motivated by their basic psychological needs, the fulfillment of which makes life and work worthwhile and interesting. This applies also at work, and we see it as the next level of work enrichment [10].

Self-Determination Theory [11] states that people are motivated at work when the work meets the needs of feeling competent when executing work tasks, being able to execute one's work tasks autonomously, and having good social relations with colleagues. For example, to foster the competence feeling, AI colleagues should allow people to complete tasks and keep up their competence at work, even

by teaching them how to complete new interesting tasks. AI teammates, if respecting the autonomy of the human colleagues, give them choice and ensure they are willingly doing their work. To address the third psychological need, relatedness, AI system could act socially like a human teammate. A social AI can act like a butler, foreseeing workers' needs and helping them complete uninteresting tasks.

Roto [10] further proposes that the third level of work enrichment is eudaimonic design of AI. While making work motivating for people is the goal of work enrichment on level two, the third level targets a future where highly autonomous AI systems are widely spread and people can choose how much they want to work. In this context, people may find lazy life meaningless and not improving their wellbeing in long term. Some people may prefer to go to work, where they can both develop themselves and make themselves useful. The Aristotelean philosophy of eudaimonia guides people to live a good life and to realize their true potential (e.g., [12]). When people come to work to improve meaningfulness of their life, an AI colleague can be designed to act like a coach that helps them flourish and live a good life. For example, AI can help workers in self-discovery and recognizing the most suitable work roles and tasks for them. When people are aware of their best potential, they can put it into use at work. AI can guide them to find personally meaningful objectives and make the best with their talents. AI can even guide people to workplaces that provide suitable tasks for the individual. Through this kind of AI teammates, we can move from work where humans are servants of AI to work where AI can make human employees flourish.

4. Discussion

In this position paper, we have drafted two future pathways for improved human–AI teaming. First, we have suggested that envelopment methods identified in organization-level AI deployments, where they balance opportunities of more powerful models with risks that inscrutability introduces, could be adapted to team-level concerns. Second, we have turned our attention to the human teammates' psychological needs – which is a topic that has been overshadowed with the surging interest that the expanding technological capabilities of AIs have created. We plan to investigate these two avenues in the future, and wish that these viewpoints raise interesting discussions in this CHI workshop.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools in the preparation of this article.

References

- [1] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, E. Horvitz, Guidelines for human–AI interaction, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019)*, ACM Press, New York, NY, 2019, pp. 3:1–3:13. doi:10.1145/3290605.3300233.
- [2] S. Berretta, A. Tausch, G. Ontrup, B. Gilles, C. Peifer, A. Kluge, Defining human–AI teaming the human-centered way: A scoping review and network analysis, *Frontiers in Artificial Intelligence* 6 (2023). doi:10.3389/frai.2023.1250725.
- [3] P. Fitts, *Human Engineering for an Effective Air Navigation and Traffic Control System*, Technical Report, National Research Council, 1951.
- [4] J. M. Bradshaw, V. Dignum, C. Jonker, M. Sierhuis, Human–agent–robot teamwork, *IEEE Intelligent Systems* 27 (2012) 8–13. doi:10.1109/MIS.2012.37.
- [5] R. Parasuraman, T. B. Sheridan, C. D. Wickens, A model for types and levels of human interaction with automation, *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 30 (2000) 286–297. doi:10.1109/3468.844354.

- [6] A. Salovaara, K. Lyytinen, E. Penttinen, High reliability in digital organizing: Mindlessness, the frame problem, and digital operations, *MIS Quarterly* 43 (2019) 555–578. doi:10.25300/MISQ/2019/14577.
- [7] B. S. Butler, P. H. Gray, Reliability, mindfulness, and information systems, *MIS Quarterly* 30 (2006) 211–224. doi:10.2307/25148728.
- [8] A. Asatiani, P. Malo, P. R. Nagbøl, E. Penttinen, T. Rinta-Kahila, A. Salovaara, Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems, *Journal of the Association for Information Systems* 22 (2021). doi:10.17705/1jais.00664.
- [9] S. Robbins, AI and the path to envelopment: Knowledge as a first step towards the responsible regulation and use of AI-powered machines, *AI & Society* (2019). doi:10.1007/s00146-019-00891-1.
- [10] V. Roto, Co-worker, butler, or coach? Designing automation for work enrichment, in: R. Rousi, C. von Koskull, V. Roto (Eds.), *Humane Autonomous Technology: Re-thinking Experience with and in Intelligent Systems*, Springer International Publishing, Cham, 2024, pp. 45–65. URL: https://doi.org/10.1007/978-3-031-66528-8_3. doi:10.1007/978-3-031-66528-8_3.
- [11] E. L. Deci, R. M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*, Springer, New York, NY, 1985. doi:10.1007/978-1-4899-2271-7.
- [12] C. D. Ryff, Happiness is everything, or is it? Explorations on the meaning of psychological well-being, *Journal of Personality and Social Psychology* 57 (1989) 1069–1081. doi:10.1037/0022-3514.57.6.1069.