Human Factors in Robots, Drones and Unmanned Systems, Vol. 93, 2023, 155–160
https://doi.org/10.54941/ahfe1003758

'Human-AI Teaming' – Review of the National Academies of Science Report

Ryan Phillip Quandt

Claremont Graduate University, Claremont, CA 91711, USA

ABSTRACT

Human-machine teams offer possibilities for conceptualization and action that could be achieved by neither alone. "Human-AI Teaming," a recent report by the National Academies of Sciences observed that teams are not reducible to their aggregation: their individual performance does not entail successful team performance. The present paper selectively reviews the report and argues that their observation supports the development of a mathematical, behavioural, and physical model of human-machine teaming as a first, essential step toward integrating AI. Joint trade-offs between structural fitness and performance underlies such a model.

Keywords: Human-machine systems, Models, Teams, Uncertainty, Bias

INTRODUCTION

The National Academies of Science, Engineering, and Medicine released a recent report, "Human-AI Teaming," in which they observe that the performance of humans or machines singly does not translate into their joint performance, nor can a team reduce to its parts (Endsley, 2022, pg. 11).[1] This observal has significant consequences for human-AI teaming: namely, a novel direction of research that draws from machine/deep-learning, natural language processing, and engineering to experimental psychology and philosophy.[2] If Human-AI teams cannot be engineered or understood apart from their teaming, how they integrate is assumed, developed, tested, deployed. A programmer must converse with a psychologist, an engineer with a philosopher, and so on. The object of study requires as much. The NAS report, then, is a welcome call for interdisciplinary projects on human-AI teaming.

There are conspicuous absences in the report, however: (i) a theory for what human-AI teams require over and above human-AI interaction; (ii) a model for their integration. These lacunae are conspicuous since the report suggests their need given the non-reducibility of teams. Or the report assumes these needs are met, but they have yet to be. The present paper selectively reviews the report to argue for the aforementioned claims. I examine the definition of team members, uncertainty and context, and the problem of bias. A theory and model are required to move from individual competencies to joint action—a central concept for teaming.

[1] On wholes not reducing to their parts, see (Bub, 2020).

[2] See the "Representative Multi-Disciplinary Team Competency Topics" (Endsley, 2022, pg. 73).

INTEGRATED TEAMS

The challenge and promise of human-AI teams is their integration. How they integrate enables possibilities that surpass humans or AI acting alone. But what integration means and requires is unclear—a theory and model will clarify. The report notes three senses of 'model,' which they leave ambiguous: (i) computational descriptions of performance, (ii) a theory of elements and processes, or (iii) best practices (Endsley, 2022, pg. 16). The first is strictly in terms of AI systems. Though needed, it will not describe their integration. The second sense overlaps with theory, save that it records parts instead of the whole (unless integration amounts to a process). The last sense concerns any rule of thumb in development or deployment, and so does not directly address integration. Nothing is said, then, of a theory or model for teaming itself.

After stating these senses, the committee makes two judgments. First, no past descriptive model "has progressed toward computational models or quantifications of the relevant importance of team characteristics, processes, or other factors" (ibid.). A new computational description is required when AI enters team—a demand resulting from the non-reducibility of teams. Second, these models "need to be informed by an understanding of the real-world demands" (ibid.). Context sensitivity is a topic I return to shortly. Suffice to say, there is scant research on the three senses of model as it applies to teams. One reason may be uncertainty about what teaming uniquely requires.

Teams and Their Members

A team is an interdependent group, each with roles that share a goal (Salas et al., 1992). Undefined is what interdependence is. The report describes the requirements of team membership functionally: based on doings, knowledge, and contribution. Presumably, then, a team is interdependent insofar as their separate doings and knowledge fulfil their assigned role to achieve an end. In this way, human-AI teams are "a step beyond" a human interacting with an AI (Endsley, 2022, pg. 19). The committee observes in a later chapter that human-AI interaction has "a significant effect" on team performance (ibid., 41), and so an account is expected for how the functionality in interaction differs from that of teams. An AI system may simply need an assigned role in a team, the execution of which contributes to a stated goal. But this account hardly suffices. A person seated at a station and told what to do, unaware they are contributing to a stated goal or collaborating with others toward that goal, is not part of a team.

The committee states that a team has heterogenous yet interdependent members (ibid., 20-21). Citing Johnson et al. (2014), the report pairs heterogeneity with structure, interdependence with process. Interdependence has two senses: assignment of responsibility for a team function sensitive to context and overlap in roles and responsibilities. Shared mental models, communication and coordination, and social intelligence are also required. A shared mental model suggests aligned goals yet, the committee remarks, "the true meaning of goal alignment is unclear" (Endsley, 2022, 21). One reason for this is layered and heterogenous goals:[3] for example, ground soldiers secure a position while pilots destroy a target, yet the position cannot be secured until the target is destroyed; both contribute to a successful mission unless the target was abandoned by enemies, in which case the respective aims of the soldiers and pilots alter. Communication and coordination require more than shared knowledge, but a certain type of interaction. As the committee observes, reporting information must be sensitive to the needs of team members given a situation (ibid., 22). And, similarly, social intelligence requires an AI's sensitivity to human beliefs, desires, and intentions (ibid., 22-23). These additional requirements support a more robust theory and model of teaming.

A functionalist account of teams may fail to appreciate three key elements of teaming: a shared intent, collaboration/coordination, and relative autonomy. These are concepts implied in joint action, which the report alludes to when it observes that teams "co-act" (ibid., 14). As a team member, AI must have common knowledge of rationality in the following sense: it must be self-aware of its task, aware that its teammate is aware of their own task, aware that the teammate is aware of the AI's awareness of its task, *ad infinitum*.[4] The report broaches this requirement in its chapter on situational awareness (Ch. 5) and transparency and explainability (Ch. 6).[5] Unsaid, though, is the simultaneity or circularity of this awareness. Some may argue that this requirement holds for human-AI interaction, too, so a theory and model of human-AI teams should account for this requirement as well as explain how it differs across interaction among role actors and teams (if at all).

Uncertainty & Context

The design and evaluation of AI must be sensitive to "context of use" (ibid., 69). Preparations, such as "field observations and interviews with domain practitioners," lessen uncertainty through insight on users, their activities, distribution of work, scope of possible situations, and the sociotechnical environ (ibid.). These preparations cannot remove uncertainty, as the committee recognizes. The report lists three sources of uncertainty: human behaviour, environment or context, and AI blind spots (ibid., 75-76). These sources are not unique to teams, except insofar as teams are unique contexts. For this reasons uncertainty may surface with teaming, such as aligning and coordinating tasks, that does not result from an uncertain situation. Parallel to bias (next section), there may be unique uncertainty introduced through teaming. This observal awaits future research since the report is silent on it.

Besides adopting more practices of Human Systems Integration (HSI), the committee suggests that HSI and human-AI interaction will evolve in unexpected ways as AI enters human teams. There are two areas of concern: (a) competencies with respect to uncertainty and context; (b) the evolution of teams when performing. The former consists in addressing AI brittleness or

[3] I am grateful to William Lawless for drawing out this point.

[4] This is not to say that one member of the team will know the state of another member's performance. Rather, each member assumes that other members are handling their task in a similarly rational manner. Otherwise, it would be difficult, if not impossible, to coordinate via anticipation.

[5] See, also, the committee's recommendations for training (ibid., 65).

edge cases, for example, and the question becomes how these limits influence teams. These are questions of AI as an agent among humans and so broach research on human-AI interaction unless these limits change in teams.

The second area of concern is unique to teaming. The committee writes that "changes in both software and environmental conditions occur almost continually" and that AI systems "may not always behave in a repeatable fashion" (ibid., 77). Calling these the result of AI blindspots conflicts with the stated aim of integrating AI into teams: namely, responsiveness to context and team members such that AI coordinates its actions relative to those of the group according to a shared goal. Teams deployed in open contexts with constraints may act unpredictably. Assuming such behaviour is not a malfunction, improvisation evinces autonomy. Team members must be able to separate ingenuity from error, just as humans discern mistakes among peers. If correct, "assured autonomy" is at least a vexed notion.[6]

Bias in Teaming

The committee reports that there are currently no standards for evaluating bias and mitigating it—a concern given the literature on bias in machine learning (West, Whittaker, & Crawford, 2019). Biases in limited or skewed training data may be hidden. Humans likewise can introduce bias through data curation, algorithm design, and interpretation (Cummings & Lee, 2021; Endsley, 2022, pg. 58). The report declares, "…the importance and impact of AI bias cannot be understated, especially for users of time-pressured systems" (Endsley, 2022, pg. 58). Potential bias ramifies in teams, creating "human-AI team bias" from interaction. This bias takes a form akin to confirmation bias or misleading representations.[7]

In Research Objective 8-1, the committee reports, referring to interactive bias, "This interconnectedness of heterogeneous and autonomous AI systems with humans who continuously learn and adapt their behaviors can generate emergent behaviors that are difficult to predict and may result in catastrophic effects" (ibid., 60). The problem is anticipating, perceiving, and correcting biases which emerge in teaming. Underlying the question of bias, then, is the uncertainty entailed in relative team member autonomy and interdependence. Since teams are likely employed in open contexts in which they must adapt over time, it cannot be known a priori how human-AI teams will evolve. The potential for interactive bias is inherent in teaming.

More than bias threatens catastrophic effects. A wider worry is that the report underappreciates differences between AI development and the integration of AI systems into teams. The evolution of teams falls strictly in the chapter on bias, whereas the open-endedness of team evolution informs team performance generally. There may be a trade-off in which greater team autonomy and interdependence means larger uncertainty, but less autonomy and interdependence limits team performance and may undermine their purpose. Uncertainty, like bias, may emerge through teaming. While there has been extensive research on human decision-making through interacting with AI

[6] See (Topcu et al., 2020).

[7] See (Singal, 2023).

systems, less has been done on team interdependence and autonomy. The recommendations through the NAS report are based on the state-of-the-art and, as a result, stress the absence of research in human-AI teaming.

CONCLUSION

A theory for human-AI teaming must account for the social, contextual, or purposive nature of concepts and action, the requirements for acting jointly, and so separate concurrent actions from coordinated and deliberated ones. Such a theory will enable researcher to specify the non-reducibility of teams, their interdependence and autonomy. A theory shapes how we place AI in teams, develop systems responsive to team members and context, identify among uncertainties in team performance in an uncertain world, and isolate bias that emerges through teaming. Theory then lends itself to a model. Note that, of the three definitions of model the report lists, none capture teaming as such. As Lawless has argued (2022), a science of individual agents based on independent and identically distributed data has been insufficient for team autonomy. Another physics and philosophy may be required.[8] These theoretical issues are left out of the NAS report despite its awareness that the nature of teams has unique challenges. Addressing these issues, however, is an essential step toward human-machine integration.

ACKNOWLEDGMENT

The author thanks Bill Lawless for inviting him to conduct this review and annotating an earlier draft.

REFERENCES

Bub, Jeffrey. (2022) "Quantum Entanglement and Information," in: Stanford Encyclopedia of Philosophy (Summer 2020 Ed.), Zalta, Edward N. (ed). Available: https://plato.stanford.edu/archives/sum2020/entries/qt.entangle/.

Cummings, M. L., Li S. (2021) Subjectivity in the creation of machine learning models, JOURNAL OF DATA AND INFORMATION QUALITY Volume 13 No. 2.

Endsley, M. R. (2022) Human-AI Teaming: State-of-the-Art and Research Needs. Washington, DC: The National Academies Press.

Johnson, M., Bradshow, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, M. B., Sierhuis, M. (2014) Coactive design: Designing support for interdependence in joint activity. JOURNAL OF HUMAN ROBOT INTERACTION Volume 3 No. 1.

Lawless, W. (2022) Toward a physics of interdependence for autonomous human-machine systems: The case of the Uber fatal accident, 2018. FRONTIERS IN PHYSICS 10:879171.

Salas, E., Dickinson, T. L., Converse, S. A., Tannenbaum, S. I. (1992) "Toward an understanding of team performance and training," in: Teams: Their training and performance, Swezey, R. W., Salas, E. (Eds.). pp. 3-29.

Singal, J. "What if diversity trainings do more harm than good?" New York Times. 1/17/23.

[8] E.g., since a self-driving car operates independently, it permits a rider to work on their iPad or engage in another distracting task. The AI system and its rider are viewed independently.

Topcu, U., Bliss, N., Cooke, N., Cummings, M., Llorens, A., Shrobe, H., Zuck, L. (2020) Assured autonomy: Path toward living with autonomous systems we can trust. arXiv: 2010.14443.

West, S. M., Whittaker, M., Crawford, K. (2019) Discriminating systems: Gender race and power in AI. Available: https://ainowinstitute.org/discriminatingsystems.pdf.