Human-Autonomy Teaming with Generative AI: How is it new and different?

January 13, 2025

Karen M. Feigh
Georgia Tech, Atlanta, GA, USA

The proliferation of Large Language Models, Generative AI, and Agentic AI has caused excitement within the human-autonomy teaming (HAT) community to explore HAT problems with these more advanced algorithms. The area builds from traditional HAT, and I refer to it as "Human-Autonomy Teaming with Generative AI" (HAT-GenAI). HAT built upon significant prior research on the ways humans and machines/systems/automation work together: Human-Automation Interaction (HAI), Human-Computer interaction (HCI), Human-Robot Interaction (HRI), Human Systems Integration (HSI), Human-Machine Teaming (HMT) and Computer-Supported Cooperative Work (CSCW). At a first glance it is debatable whether HAT-GenAI should be considered a new area or not, however, I have yet to find these differences properly articulated. What differentiates HAT-GenAI from other forms of HAT or Interaction?

I have been thinking deeply about this question lately and have concluded that HAT-GenAI is indeed different than existing HAT areas. But as HAT-GenAI builds directly upon prior HSI, HRI, HMT, and HAI work, it inherits a substantial amount from prior HAT and other human factors work. In the years since Schulte in 2015 and McNeese, Demir, Cooke, and Myers defined HAT in 2018, many studies that utilized any kind of teaming or sophisticated automation and even some classical AI, e.g. Machine Learning, Neural Networks, tended to describe themselves as studying HAT - making any distinction between prior forms of Interaction research and HAT difficult to find. And it is easy to understand how one might continue that line of thinking to the new HAT-GenAI. In many ways it remains similar:

Work Environments - The factors of a team's environment, algorithmic design, composition, and the automation's role are highly similar and in many cases identical to prior studies.

Importance of Transparency - The need for humans to understand what the automation is doing and why it is doing so were as strong then as it is now. The need for transparency and communication is the same. The interactions are largely the same at this point (although LLMs are making NLP interactions more common and this could be a distinction).

Figures of Merit - The end-goal performance metrics are largely the same (at this point). So far, most studies are measuring team performance with the same metrics we have used for decades – workload, taskload, performance measures, trust scales, situation awareness assessments, physiological measures.

Team Composition - Teams with generative AI have been similar in composition to those of other forms of automation. While the vast majority of work in HAT has been done with human-agent dyads, some HAT work was accomplished with larger teams (the work of Nancy

Cooke's Synthetic Teammate Project and Julie Adams' multi-robot teams work come to mind most prominently). To my understanding, very little work in the HAT-GenAI space has explored beyond human-agent dyads (notable exception of the work of McNeese).

Role of Humans - Beyond just the number of humans, what their role is and the use of human confederates in experiments seem to remain the same. These are common practices that have been used for years and have been exceedingly fruitful.

Simulated or Physical Environments - The benefits of utilizing digital task environments and the necessity of sometimes utilizing physical task environments are not unique to HAT-GenAI problems. Again I'll point to work by Nancy Cooke in simulated environments and to the wider HRI community for their collective work in prototyping in simulation and then transitioning to real hardware to illustrate proof of concepts in both domains,

Adaptations & Triggers - Years ago, Michael Dorneich, Caroline Hayes and I laid out how automation (both adaptive and adaptable) could adapt and what categories of triggers of those adaptations were available. The basic ideas of adaptation and triggers are well trodden by prior research - the works of Wickens, Parasuraman, Kaber, Seridan, Smith, and Endsley come to mind.

With such overlap between traditional Interaction and HAT areas and the new HAT-GenAI, what remains? The major distinctions seem to be:

Who Should Adapt and How - Automation was already capable of adapting behavior to the external world state, however agentic systems and generative AI enable semantic reasoning beyond what was previously practical. HAT-GenAI allows autonomy to consider the teammates' relative strengths, weaknesses, and preferences to a level exceeding frameworks considered by traditional HAT or Interaction. This is the fundamental basis for *teaming*, not simply interaction. The work of Scheutz 2017 lays out a thorough description of what information an agent may want to include to adapt its behavior to the team.

Range of Tasks Possible - The range and breadth of tasks that can be accomplished with generative AI is broader and the likelihood of emergent behaviors is stronger (see Brandizzi 2023 and Trusilo 2023), bringing in new and novel domains to study. Expanding the scope of HAT provides a key distinction for HAT-GenAI.

Novel Metrics - Generative AI presents an entirely different communication paradigm than is traditionally explored in HAT. As such, new teaming metrics are certainly needed. How will we "rate" the teaming capabilities of our new autonomous teammates? Not only their capabilities at task work, but soft skills such as effective bidirectional communication that can make or break a human-human team? Perhaps the human factors community can draw from information theory to objectively quantify communication while considering writing quality.

Elevated Capabilities - We are expecting (and relying on) more from generative AI than we have needed to with previous algorithms $\rightarrow$ ethical behavior, consideration of our preferences, a desire to not "fail" their human teammate (see Goodrich 2023 and Giolando 2024).

These are the key differences that I can think of based on my experiences with HAT-GenAI and much work prior to generative AI. There are likely a few others that I have not captured, and probably better dimensions along which to organize these thoughts - I wanted to quickly kick these out of my brain and possibly start a conversation.

So to conclude: Yes, I think HAT-GenAI warrants being labeled distinctly from other HAT fields, and that research in HAT-GenAI will blend the tried and true methods and metrics of prior eras, as well as incorporate novel elements as outlined above.