Human-Agent Teaming for Higher-Order Thinking Augmentation

Chung-Chi Chen
Artificial Intelligence Research Center,
National Institute of Advanced Industrial Science and Technology, Japan
c.c.chen@acm.org

Abstract

Human-agent teaming refers to humans and artificial agents working together toward shared goals, and recent advances in artificial intelligence, including large language models and autonomous robots, have intensified interest in using these agents not only for automation but also to augment higher-order cognition. Higher-order thinking involves complex mental processes such as critical thinking, creative problem solving, abstract reasoning, and metacognition, and intelligent agents hold the potential to act as genuine teammates that complement human strengths and address cognitive limitations. This tutorial[1] synthesizes emerging research on human-agent teaming for cognitive augmentation by outlining the foundations of higher-order thinking and the psychological frameworks that describe it, reviewing key concepts and interaction paradigms in human–AI collaboration, and examining applications across education, healthcare, military decision-making, scientific discovery, and creative industries, where systems such as language models, decision-support tools, multi-agent architectures, explainable AI, and hybrid human–AI methods are used to support complex reasoning and expert judgment. It also discusses the major challenges involved in achieving meaningful augmentation, including the calibration of trust, the need for transparency, the development of shared mental models, the role of human adaptability and training, and broader ethical concerns. The tutorial further identifies gaps such as limited evidence of long-term improvement in human cognitive abilities and insufficient co-adaptation between humans and agents. Finally, it outlines future directions involving real-time cognitive alignment, long-term studies of cognitive development, co-adaptive learning systems, ethics-aware AI teammates, and new benchmarks for evaluating collaborative cognition, offering a comprehensive overview of current progress and a roadmap for advancing human-agent teaming as a means of enhancing higher-order human thinking.

[1] https://nlpfin.github.io/sites/aacl2025.html

1 Human-Agent Teaming

Traditional AI or automation has often been viewed as a tool that a human uses – a passive instrument executing tasks under human direction. In contrast, human-agent teaming (HAT) envisions AI systems as active team members that collaborate with humans in a more symmetric, interdependent manner. In a HAT scenario, the human and AI share a common goal, and each contributes their distinct capabilities to jointly achieve outcomes that neither could alone as effectively. HAT is sometimes termed human–AI teaming (HAIT), human–autonomy teaming, or human–AI collaboration – reflecting overlapping concepts. Across these definitions, the emphasis is on leveraging the complementary strengths of humans (e.g. intuition, ethical judgment, creativity) and AI agents (e.g. speed, data processing, precision) in an integrated way. For example, an AI might generate options or analyze large datasets while the human makes contextual judgments and provides oversight, together making a better decision than either could alone. Crucially, HAT is seen as a human-centered approach to AI deployment: its aim is not just raw efficiency, but also to ensure human well-being, learning, and motivation by making the AI a supportive partner rather than a black-box replacer.

As AI technologies become more autonomous and "smart," people begin to perceive them as social agents rather than mere devices. This opens the door to designing AI that engages in teamwork behaviors – communicating, adapting, even exhibiting social qualities like encouragement or etiquette – thereby fitting more naturally into human teams. Make an AI feel like a teammate instead of a tool, including the agent's agency (ability to act independently), benevolence (being oriented to help the

Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics: Tutorial Abstract, pages 20–24

human), communicativeness, interdependence (its actions depend on human actions and vice versa), synchrony (coordination and timing in interaction), and a team focus (shared goals). When humans perceive these attributes in an AI – for instance, an AI that proactively updates a plan in response to a human's change in strategy (showing interdependence and initiative) – they are more likely to trust and "team up" with it rather than treat it as just an automated tool (Lyons et al., 2021).

An important aspect of HAT is the level of autonomy the agent has. Early framework (Sheridan and Parasuraman, 2005) defined levels ranging from complete human control to full machine control. In HAT contexts, instead of replacing the human at high autonomy, the goal is a balance where the AI has enough autonomy to act proactively as a teammate, but not so much that the human is out of the loop. Lyons et al. (2021) suggest that to qualify as a "teammate," an agent must possess a degree of autonomy (to sense, decide, and act) and adaptive behavior – it cannot be completely pre-scripted or it would be a tool, not a partner. Levels of Autonomy (LOA) in human–agent teams refer to how decision-making responsibility is allocated. For example, one common scale is: at low LOA the AI might only suggest options and the human decides; at medium LOA the AI makes a recommendation which the human can approve or veto; at high LOA the AI can decide and execute actions on its own unless the human intervenes (Rebensky et al., 2022). Each level has trade-offs in human workload, trust, and team effectiveness. A recent study on multi-agent teaming in a simulated drone surveillance task found that varying the LOA impacted the human operator's workload, stress, and trust in the agents. Higher autonomy reduced the operator's micro-management burden but also required the human to trust the agents' decisions – underscoring the importance of calibrating autonomy to human preferences and situational demands. In general, the literature suggests that an optimal HAT often involves a dynamic autonomy approach (sometimes called adjustable or adaptive autonomy), where the level of agent independence can shift as needed, maintaining an appropriate division of labor and authority between human and AI.

Human-agent interactions can be characterized along a spectrum from loosely coupled to tightly coupled collaboration. A useful taxonomy distinguishes between: co-existence, where humans and AIs work in parallel with minimal direct interaction; coordination/cooperation, where they share some information and resources but have mostly separate sub-tasks; collaboration, where they work more closely on shared tasks and must synchronize their actions; and teaming, which is the most interdependent form of collaboration often implying shared intentions, continuous mutual adjustment, and even social bonding akin to human teams. Teaming implies a high degree of interdependence – each agent (human or AI) relies on the other's actions – and often involves a sense of mutual commitment or cohesion. In concrete terms, consider a spectrum in a driving context: a self-driving car that simply drives while the human passenger does unrelated work is automation (co-existence at best); a driver-assist system that can take over in some situations if the human requests is coordination; whereas a car that actively converses with the driver about navigation choices, taking over routine control so the human can focus on situational strategy (and vice versa in complex scenarios), could be seen as teaming. The latter requires rich communication and each partner understanding the other's roles – hallmarks of teaming. Indeed, the form of interaction is a key part of HAT design: whether the interaction is through natural language dialogue, through a GUI with visualized AI reasoning, via implicit signals (e.g. the AI picking up on human behavior patterns), etc., all influence how effectively the human and AI can function as a team.

One well-known paradigm is the CASA (Computers as Social Actors) concept, which notes that people tend to apply social rules even to computers given minimal cues (Nass et al., 1996). Modern AI with human-like conversational ability or embodiment amplifies this effect. This means designers can leverage social interaction patterns – for example, having an AI explain its reasoning or acknowledge errors – to improve teamwork. Another paradigm is mixed-initiative systems, where both human and AI can initiate actions or changes in the task based on who is best suited at the moment. Effective HAT often calls for transparency (the AI reveals its intent and reasoning) and shared control, enabling fluent turn-taking or simultaneous contributions. For instance, in a writing assistant scenario, a mixed-initiative agent might not only generate text when asked, but also pose questions or highlight potential improvements unprompted, thus actively engaging the writer in a back-and-forth creative process.

Shared Mental Models and Teaming: In human

teams, a critical factor for success is a shared mental model – a common understanding among team members of the task, the goals, each other's roles, and the state of the environment. Similarly, for human–AI teams, researchers emphasize the need for the AI to form (or approximate) a model of the human's intentions and preferences, and for the human to develop an accurate mental model of what the AI can do, how it behaves, and when to rely on it. Without this, the human may be surprised by the AI's actions or not trust it appropriately. The National Academies (2022) identified conditions for successful human–AI teams, including: the human's ability to understand and anticipate the AI's behavior, to maintain appropriate trust, to use the AI's outputs effectively in decisions, and to effectively control or intervene in the AI's operations. These conditions allude to alignment of mental models – the human must grasp the AI's capabilities/limits and the AI ideally should adapt to the human's goals and provide information in a way the human can make sense of. Research on "teachable AI" or "partner AI" explores methods for agents to learn a user's preferences over time or to engage in dialogue to clarify goals, thereby improving the team's shared understanding. For example, an AI assistant might learn a particular scientist's experimental style and pre-filter results accordingly, or a planning AI might ask "Do you prefer a faster route or a more scenic one?" to ensure it models the driver's priorities correctly.

As noted, balancing AI autonomy with human control is a central design decision. If the AI is too unassertive (always waiting for explicit human commands), it might under-utilize its abilities and burden the human with micro-management. If it is too assertive (acting without human awareness or input), the human can become out-of-the-loop, leading to mistrust or misuse (e.g. over-relying on an autonomous system without monitoring it). A taxonomy by O'neill et al. (2022) discusses human-autonomy teaming where the human and AI continuously negotiate control – sometimes the AI leads, other times the human leads, depending on who has the advantage in that situation. In practical terms, many HAT systems implement adaptive autonomy: the AI might take over routine tasks autonomously but will defer to the human for critical or ambiguous decisions, or it might ask permission when it is unsure. Research in contexts like military UAV control has tested autonomous agents that handle low-level flying and surveillance tasks, freeing the human operator to focus on higher-level mission strategy. Initial results suggest that such agents can improve overall mission performance if the human can maintain situation awareness of what the agents are doing and intervene when necessary. Thus, interface design (alerts, explanations, etc.) that supports coordination is crucial. The concept of "continua of autonomy" has emerged – envisioning a slider or flexible assignment of function that adjusts as a mission or task evolves. Such flexibility is key to treating the AI as a teammate whose level of initiative can change rather than a fixed autopilot.

In summary, human-agent teaming is an evolution from simple "human-in-the-loop" automation to a partnership model. It demands careful consideration of roles, communication, autonomy, and trust. The AI must be designed not just for task performance but for teamwork performance, which includes being predictable, transparent, and adaptive to the human. The human, on the other hand, may take on new roles such as a supervisor, collaborator, or student in relation to the AI. With these concepts in mind, we will discuss how HAT is being applied in various domains where higher-order thinking is critical. Each domain illustrates different ways that intelligent agents can augment human cognition – and the unique challenges that arise.

2 Higher-Order Thinking

Higher-order thinking (HOT) broadly refers to cognitive processes that involve going beyond rote memorization or basic comprehension to engage in analysis, synthesis, evaluation, and creation. A classic definition by Lewis and Smith (1993) describes HOT as occurring "when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations." In other words, it entails transforming knowledge to solve novel or non-routine problems. Lewis and Smith note that HOT is used in tasks such as "deciding what to believe or do; creating a new idea or artistic expression; making a prediction; and solving a non-routine problem." Higher-order thinking skills are often contrasted with lower-order skills that involve recall or routine procedures.

Key Components of HOT: include the following cognitive skills under the HOT umbrella (Yatani et al., 2024):

Critical Thinking: The capacity for purposeful, reasoned, and goal-directed thinking in evaluating evidence, forming judgments, and solving problems. Halpern (2013) defines critical thinking as "thinking that is purposeful, reasoned, and goal-directed – the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions." It also implies a disposition of reflective skepticism, i.e., being willing to question assumptions. Critical thinking enables one to analyze arguments, identify biases, and avoid being misled – a skill increasingly essential in the information age.

Creative Thinking: The ability to generate novel and valuable ideas or solutions. Torrance (2018) describes creativity as "the process of sensing gaps or missing elements; forming ideas or hypotheses concerning them; testing these hypotheses; and communicating the results." Creative thinking is not limited to the arts; it is vital for innovation in sciences, engineering, business, and everyday life. It involves divergent thinking (exploring many possible solutions) as well as convergent thinking (synthesizing information into a workable idea). Notably, both critical and creative thinking can be cultivated through practice in reasoning, analysis, and open-ended problem solving. In addition to cognitive strategies, attitudes matter: effective critical thinkers tend to be willing to plan, persistent, self-correcting, and mindful of bias, and creative individuals benefit from confidence in their creativity and a willingness to take intellectual risks.

Problem Solving: The process of working through details of a challenge to reach a solution when the path is not immediately obvious. Mayer and Wittrock (1996) famously define problem solving as "cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver." This definition highlights that true problem solving requires more than applying a known formula; it involves dealing with uncertainty, devising or discovering methods, and often, iterative trial and error. Problem solving encompasses subskills like problem representation (understanding and framing the problem), strategy formulation, reasoning through possible actions, and evaluating outcomes. Complex, ill-defined problems (e.g., designing a new product or diagnosing an unfamiliar patient case) particularly demand higher-order reasoning, as opposed to well-defined problems that might be solved by routine application of learned rules.

Metacognition: Commonly described as "thinking about thinking," metacognition involves awareness and regulation of one's own cognitive processes. It includes metacognitive knowledge (knowing one's cognitive strengths, weaknesses, and the strategies available) and metacognitive regulation (planning, monitoring, and adjusting one's approach to a cognitive task). Metacognition plays a supporting role in higher-order thinking by helping individuals select appropriate strategies and reflect on the effectiveness of their thinking. For example, a person solving a complex problem uses metacognition to plan how to approach it, monitor their progress ("Have I considered all possible options?"), and revise strategy if stuck. Strong metacognitive skills are associated with better application of critical thinking and problem-solving skills. Notably, as AI systems take on more cognitive tasks, researchers point out that humans may face metacognitive demands in working with AI – e.g., checking AI outputs and understanding their limits – which in turn requires support. Tankelevitch et al. (2024) argue that generative AI can impose heavy metacognitive load on users and propose incorporating metacognitive support into AI tools to help users manage this load.

Abstract Reasoning: The ability to reason with concepts that are not tied to concrete experiences, often involving recognizing patterns, logical relationships, or general principles that can be applied in new contexts. Abstract reasoning is closely related to fluid intelligence – the capacity to solve novel problems independent of acquired knowledge. Examples include understanding metaphorical or symbolic representations, solving puzzles like analogies or matrix patterns, or constructing models of complex systems. Abstract reasoning allows one to think conceptually and handle complexity by mentally manipulating ideas. It enables "thinking about things that are not immediately present or tangible . . . using concepts, patterns, and relationships." This skill underpins higher-order tasks like theoretical reasoning in science or strategic planning, where one must infer general rules from specifics or envision possibilities beyond the here-and-now.

These components are interrelated and often used together. For instance, solving a real-world problem might require critical analysis of information, creatively brainstorming solutions, using abstract reasoning to model the problem, and mon-

itoring one's problem-solving approach metacognitively. Collectively, they enable "effective use of higher-order thinking skills like analysis, evaluation, and creation" to deal with unfamiliar, complex challenges. Developing higher-order thinking has long been an educational goal, as it equips individuals to adapt and learn in new situations – a need that is ever more pressing in the face of rapid technological change.

3 Scope of the Tutorial and Cross-Domain Synthesis

Building on the conceptual foundations of human–agent teaming and the cognitive frameworks underlying higher-order thinking, this tutorial proceeds to explore how these ideas manifest across multiple real-world domains. In the sections that follow, we examine diverse application settings—including education, healthcare, scientific discovery, creative industries, military and safety-critical operations, and knowledge-intensive professional work—where human–AI collaboration holds particular promise for augmenting complex reasoning, decision-making, and metacognitive processes. Each domain illustrates both the opportunities and constraints of treating AI systems as cognitive partners rather than passive tools, revealing how contextual factors such as expertise level, task structure, risk profile, and social expectations shape the dynamics of teaming.

Across these domains, common patterns begin to emerge. First, effective augmentation depends on an alignment of human and agent mental models, where the AI not only communicates its internal states, uncertainties, and intentions, but also adapts to human goals, preferences, and cognitive styles. Second, higher-order thinking augmentation is most successful when the AI supports—not replaces—core human reasoning processes: helping users reflect, plan, generate alternatives, explore conceptual space, and evaluate competing hypotheses. Third, challenges such as trust calibration, over-reliance, cognitive offloading, and the opacity of model reasoning recur regardless of domain, underscoring the need for interaction designs that balance autonomy with interpretability. Finally, the empirical evidence highlights substantial gaps: while short-term performance gains are often observed, there is limited understanding of whether AI teammates can foster long-term cognitive growth, transfer of reasoning strategies, or durable improvements in critical and creative thinking.

By synthesizing these cross-domain insights, the tutorial aims to provide a unifying perspective on how intelligent agents can be designed to meaningfully augment human higher-order cognition. We conclude by identifying open research directions that offer a roadmap for advancing the practice of human–agent teaming for cognitive augmentation.

Acknowledgments

This tutorial was supported in part by AIST policy-based budget project "R&D on Generative AI Foundation Models for the Physical Domain."

References

Diane F Halpern. 2013. *Thought and knowledge: An introduction to critical thinking*. Psychology press.

Arthur Lewis and David Smith. 1993. Defining higher order thinking. *Theory into practice*, 32(3):131–137.

Joseph B Lyons, Katia Sycara, Michael Lewis, and August Capiola. 2021. Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in psychology*, 12:589585.

Richard E Mayer and Merlin C Wittrock. 1996. Problem-solving transfer.

Clifford Nass, Brian Jeffrey Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6):669–678.

Thomas O'neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2022. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, 64(5):904–938.

Summer Rebensky, Kendall Carmody, Cherrise Ficke, Meredith Carroll, and Winston Bennett. 2022. Teammates instead of tools: The impacts of level of autonomy on mission performance and human–agent teaming dynamics in multi-agent distributed teams. *Frontiers in Robotics and AI*, 9:782134.

Thomas B Sheridan and Raja Parasuraman. 2005. Human-automation interaction. *Reviews of human factors and ergonomics*, 1(1):89–129.

Human-AI Teaming. 2022. State-of-the-art and research needs. *National Academies of Sciences, Engineering and Medicine, Washington DC*, 10:26355.

E Paul Torrance. 2018. *Guiding creative talent*. Muriwai Books.

Koji Yatani, Zefan Sramek, and Chi-Lan Yang. 2024. Ai as extraherics: Fostering higher-order thinking skills in human-ai interaction. *arXiv preprint arXiv:2409.09218*.