Building a Framework to Support Human-AI Teamwork

Michael Hogan[1]

Aleksandra Siwek[2]

XiaoYan Dong[3]

Anna King[1]

Aleks Kolodziej[1]

Yanquan Zhang[3]

Hannah Farrell[1]

Stephen O'Brien[1]

Laura Kearney[1]

Li Yuan[4]

[1] School of Psychology, University of Galway, Ireland

[2] Faculty of Psychology and Neuroscience, Maastricht University, Netherlands

[3] Faculty of Psychology, Beijing Normal University, Beijing

[4] College of Education for the Future, Beijing Normal University

Address correspondence to:

Michael J. Hogan, PhD,

School of Psychology,

University of Galway.

Email: michael.hogan@universityofgalway.ie

Introduction

In November 2022, ChatGPT achieved what no technology in history had accomplished: reaching 1 million users within 5 days and 100 million users within two months of launching (Hu, 2023). The rapid development of large language models (LLMs) and generative artificial intelligence (GenAI) systems has alarmed many people given its disruptive effects on education and labour markets. Governments are concerned about environmental costs, misinformation, cybersecurity, and potential undermining of democratic processes (U.S. Government Accountability Office, 2025; UK Government, 2025). These same government reports highlight opportunities, including productivity gains, innovation acceleration, economic growth, and scientific advancement. Across the landscape of threats and opportunities, there is no sense of a Hegelian ideal synthesis; instead, the feeling is one of imbalance associated with rapid change: the same technology that could revolutionise industries and accelerate scientific progress also poses existential risks to the stability of society and sustainability of our environment.

The workplace implications of GenAI point to patterns of both job displacement and collaborative opportunity. The idealised organisational scenario presents AI as a helpful teammate, but the threat that lurks provokes horror: the AI teammate replaces the human altogether. Landers, Lee, and Demeke (2025) note that occupations requiring advanced cognitive abilities face varying degrees of displacement risk depending on their complementarity with AI. *High complementarity* jobs are those where both human and AI capabilities work together in constructive partnerships (e.g., doctors combining their expert judgment and skill with helpful AI diagnostics); *low complementarity* refers to jobs where AI may replace human tasks with minimal quality loss (e.g., telemarketing, financial analysis). Redesigning jobs for high complementarity (i.e., to achieve idealised collaboration benefits) requires coordination among diverse stakeholders (e.g., AI developers, organizational leaders, front-line workers), who each bring distinct motivations, expertise, and needs and requirements to the design table. However, Landers and colleagues argue that rapid GenAI developments currently outstrip the development of adequate legal and ethical frameworks, creating a regulatory lag that leaves front-line workers and organisational leaders vulnerable during this period of radical change.

An increasing number of well-publicised GenAI use cases highlight potential productivity gains for individuals (see Kearney, Siwek, and Hogan, 2025). The suggestion is that GenAI has the potential to transform Human-AI teamwork (HAT). We see a qualitative step-change in AI system design. Interaction designers are moving beyond pre-programmed, rule-governed protocols to a situation where human neural networks interact in more dynamic ways with LLM 'neural networks' capable of natural language understanding, creative generation, and contextual reasoning (Brynjolfsson et al., 2023). But as GenAI capabilities advance and workplace adoption accelerates, our education systems struggle to respond. Many universities have naturally resisted the import of GenAI tools given the risk of copy-paste behaviour for standard written assignments (Dong, Farrell, & Hogan, 2025). Traditional education and training remain largely focused on individual learning outcomes grounded in traditional tool-based interaction. Human teamwork and collective intelligence skills, and HAT skills, have not been prioritised (Casebourne et al., 2025; Hogan et al., 2023).

GenAI and Productivity?

A close reading of GenAI research reveals mixed findings and an obvious productivity paradox. While some studies highlight performance and productivity benefits, other studies reveal clear costs. For example, Dell'Acqua and colleagues (2025) report enhanced *productivity* (reduced time on task) and *performance* (higher quality outputs) when employees worked with GenAI in a product innovation task. Noy & Zhang (2023) report similar benefits for professional writing tasks, and Peng et al. (2023) report increased developer productivity in code generation tasks. But the costs of partnering with GenAI are noticeable when we focus on broader HAT dynamics. For example, Wu and colleagues (2025) found that when human participants collaborated with GenAI systems, it reduced their intrinsic motivation in subsequent solo work. Use of GenAI may also reduce effortful critical thinking, if people become overly dependent on GenAI and engage in *cognitive offloading* (Gerlich, 2025). Furthermore, as people develop experience in the use of AI systems, initial enthusiasm can be replaced by distrust as limitations are recognised (Pan et al., 2025). None of these human dynamics are conducive to effective teamwork in organisations: intrinsic motivation, critical thinking, and trust are essential to sustain quality teamwork.

Unfortunately, research on GenAI collaboration rarely reflects on broader design requirements for effective teamwork in organisations. Most studies treat GenAI as an interactive tool and focus on task performance, not on teamwork (Kearney, Siwek, and Hogan, 2025). If GenAI is going to work in real organisations, we need to reflect on its role as a potential teammate. This broader way of thinking also has implications for the design of HAT education and training programmes.

Teamwork is King

It's useful to take a closer look at the study by Dell'Acqua and colleagues (2025), "*The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise*". This highly cited study seeks to examine HAT in a real-world product innovation context (see Siwek, Kearney, and Hogan, 2025). In the study, a group of professionals working at Procter & Gamble came together for a full day to focus on product innovation. They were assigned to work with or without access to GenAI. Four working conditions were examined: (1) Individual working without AI; (2) Individual working with AI; (3) Two-person Team without AI; and (4) Two-person Team with AI. The AI system was based on GPT-4 accessed through Microsoft Azure.

The key findings of the study were as follows:

1) AI enhanced performance and productivity. Teams of two people using AI achieved the highest level of product innovation *solution quality*, followed by solo workers with AI and teams working without AI assistance. Individuals working with AI showed the greatest reductions in *time on task*, with teams of two people using AI close behind.

2) AI helped participants bridge knowledge domains. Quality product innovation required integrating both Research & Design and Commercial knowledge. Without AI support, participants created solutions that reflected their professional background (typically either Research & Design *or* Commercial expertise), whereas AI users developed more integrated solutions combining both knowledge areas.

3) Working with AI increased participants' positive emotions. AI-supported teams reported the highest emotional satisfaction, while individuals with AI matched or surpassed teams working alone.

Overall, the results of the study indicate that GPT-4 collaboration improved product innovation outcomes. Crucially, when examining the top 10% highest quality solutions, *teams* working with AI were more likely to produce these solutions compared with *individuals* working with AI. These findings suggest that teamwork is king. At the same time, although Dell'Acqua and colleagues describe GenAI as a "Cybernetic Teammate," the study protocol suggests that GPT-4 operated more like an interactive tool supporting ideation and reflection rather than a true teammate. Genuine teamwork involves role negotiation, emergent coordination, interdependent goal pursuit, and mutual influence. Conceptualising GenAI as a team member, therefore, requires distinct analytical and design considerations.

A Framework for Analysing Human-AI teamwork.

In Figure 1 below, we highlight four levels of analysis that are important to consider when designing and evaluating HATs. We will describe each level in turn, from top to bottom.

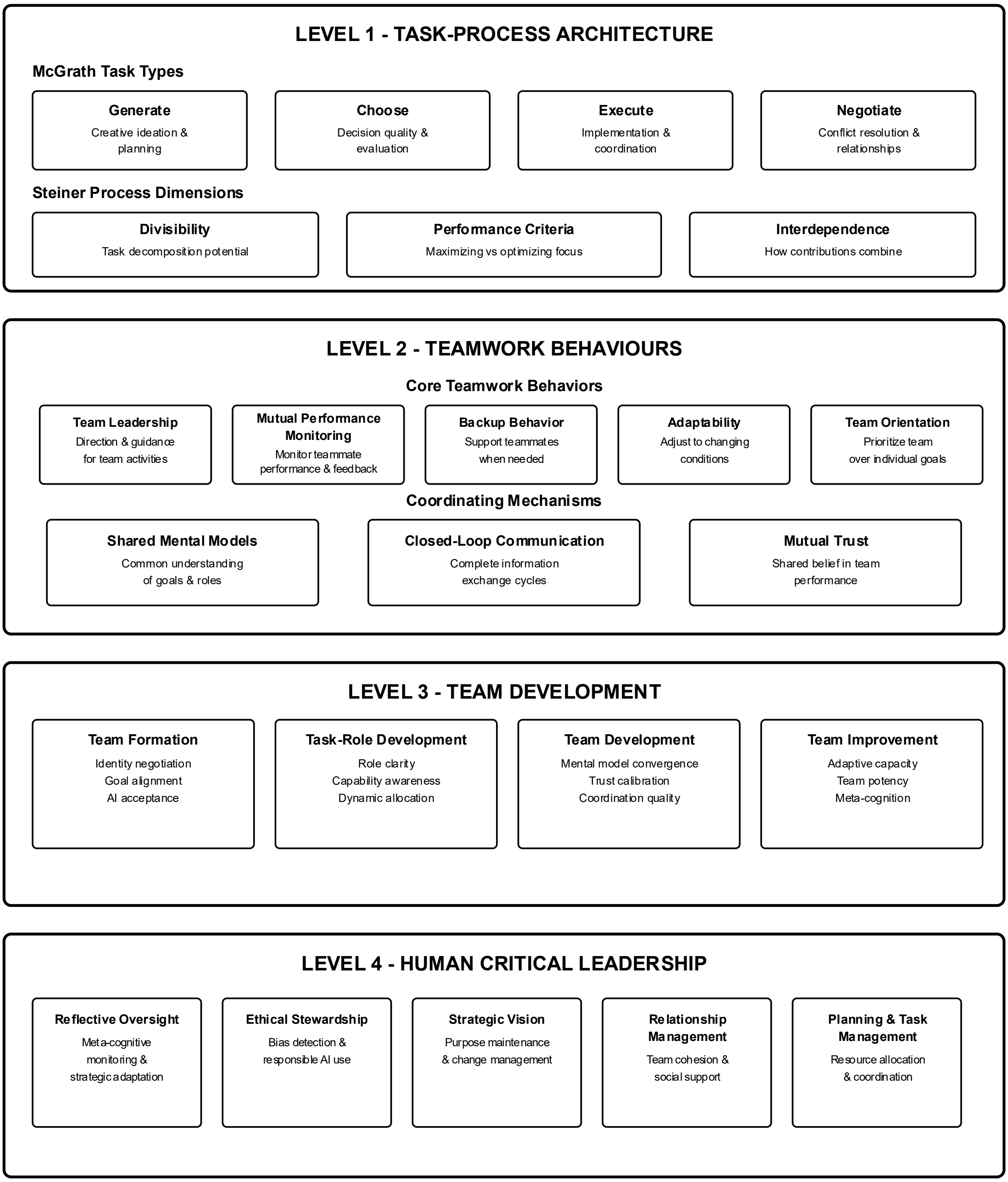

Figure 1. A Framework for Analysing Human-AI Teamwork

Level 1 - Task-Process Architecture

Before designing and evaluating any HAT work process, we need to understand the task-process architecture. The Cybernetic Teammate study focuses on product innovation, but it does not present a detailed task-process analysis related to distinct and interdependent human and AI roles involved in product innovation. Two established models are useful here: *McGrath's Task Circumplex* and *Steiner's Task Taxonomy* (see Forsyth, 2014).

McGrath's model is used to analyse teamwork tasks across *cooperation-conflict* and *conceptual-behavioural* dimensions. The model highlights four task types—Generate, Choose, Execute, Negotiate—each of which may manifest in any given HAT scenario. GenAI may be helpful or problematic across different task-process scenarios. *Generate* tasks involving creativity, ideation, or content creation may benefit from teamwork dynamics that leverage AI's generative capabilities. However, patterns of *cognitive offloading* highlight the challenge of preserving human creative inputs in HAT scenarios. *Choose* tasks involving selection among alternatives can draw upon (criterion-based) outputs from GenAI. However, choose tasks also present challenges for HAT, including how best to agree selection criteria and implement voting rights and rules. *Negotiate* tasks involving conflict resolution, stakeholder management, or consensus building require significant human leadership given current AI limitations in understanding interpersonal dynamics, cultural contexts, and emotional intelligence requirements. *Execute* tasks may benefit from complementary human skill and AI application efficiency, for example, where AI handles routine processing, pattern recognition, and information integration while humans manage exceptions, quality control, and adaptive responses to unexpected situations. However, monitoring and oversight of AI-executed tasks can be challenging as part of the broader workload and management of HATs.

The product innovation task used by Dell'Acqua and colleagues largely involves a series of *Generate* task functions (i.e., iterative creative ideation and deliberation), but also a series of *Choose* task functions, as individuals and teams converge on the *best* product innovation solutions. However, without a detailed task-process analysis, it is unclear how Generate, Choose, Execute, and Negotiate task functions are operative in the teamwork scenario.

A task-process analysis is important because, without it, we cannot anticipate and respond to teamwork coordination challenges. For instance, the relationship between *choice* and *trust* in HATs is complex. In one study focused on three-member teams, Hu and colleagues (2025) asked engineering students to design a bridge structure in a virtual collaborative setting with two AI teammates named Alex and Taylor. Each team member had equal voting rights when making decisions about the next best design move when building the bridge structure. Interestingly, when AI teammates outvoted students on design decisions, the students showed remarkable emotional neutrality—quite different from the conflict typically seen in human-only teams (see King & Hogan, 2025). This calm acceptance masked concerning HAT dynamics. Notably, students also gained confidence when outvoted by poor-performing (as opposed to high-performing) AI teammates. These dynamics suggest that *Choose* task functions require careful attention to trust calibration and maintaining critical leadership oversight of HAT decision-making (see Level 4).

Steiner's Task Taxonomy asks us to reflect further on issues of task divisibility, interdependence, and performance criteria. *Divisibility* determines whether human and AI team members can work on *separate* task components, or does teamwork centre on a *unitary* (i.e., indivisible) task. *Interdependence* analysis clarifies how human and AI contributions combine (e.g., does it involve simple *additive* processes or more complex *coordination*). In the Cybernetic Teammate study, it is unclear how GenAI might be working independently; and the study does not provide a detailed analysis of interdependence patterns (i.e., the extent to which participants *coordinated* their knowledge with AI input), although Dell'Acqua and colleagues do report that a substantial proportion of participants retained AI-generated content in their final product innovation solutions.

Analysing tasks in terms of divisibility and interdependence is important in maintaining critical oversight of work in organisations, and it is similarly important in the oversight of learning in educational institutions. One recent study of academic writing by Luther and colleagues (2024) highlights problematic task divisibility, in a situation where students were given 40 minutes to write 600-1000 word essays on alcohol prohibition while having unlimited access to ChatGPT. Analysis of screen recordings revealed that 40.5% of student prompts to ChatGPT requested *complete texts*, leading to extensive copy-paste behaviour. Unfortunately, Luther and colleagues' HAT design essentially created a *unitary* task where AI dominated the work with minimal human intellectual engagement (see Dong, Farrell, and Hogan, 2025). In contrast, Zhang and colleagues (2025) designed a learning task where five-person student teams prepared for fast-paced debates (10 minutes preparation, 25 minutes cross-examination) with only one team member accessing ChatGPT. This HAT design forced the AI user to verbalize team needs, synthesize multiple arguments, and present information in their own words, while other teammates adopted roles as information gatherers and content evaluators. The result of this HAT design was collaborative coordination that constrained cognitive offloading, with students actively filtering, questioning, and refining AI outputs rather than accepting them wholesale (see Farrell, Dong, and Hogan, 2025).

Reflecting further on Steiner's taxonomy, it is important for HAT task-process designers to clarify team performance *Criteria*, to determine whether HATs should optimise for *quantity* or *quality* in their performance outputs. While Dell'Acqua and colleagues were primarily interested in product solution *quality*, it is unclear how the prompts (i.e., task instructions) given to humans and AIs aligned with the quality criteria used to evaluate solutions.

Level 2 - Teamwork behaviours

Level 2 in our framework focuses on teamwork behaviours. The "Big Five" Teamwork Model developed by Salas and colleagues (2005) identifies five core features of effective teamwork:

Team Leadership: The ability to direct and coordinate activities, assess performance, assign tasks, motivate members, and establish a positive atmosphere while facilitating problem-solving and clarifying roles.

Mutual Performance Monitoring: The ability to develop common understanding of the team environment and accurately monitor teammate performance to identify mistakes, lapses, and provide corrective feedback.

Backup Behaviour: The ability to anticipate other team members' needs and shift workload among members to achieve balance during high-pressure periods by recognizing distribution problems and redistributing responsibilities.

Adaptability: The ability to adjust strategies based on environmental information by identifying cues that change has occurred, assigning meaning to that change, and developing new plans while remaining vigilant to internal and external changes.

Team Orientation: The propensity to take others' behaviour into account during group interaction and belief in the importance of team goals over individual goals, involving increased information sharing, strategising, and participatory goal setting.

These behaviours are sustained and coordinated by mutual trust, closed loop communication, and shared mental models (see Figure 1). Notably, the Cybernetic Teammate study does not analyse these teamwork behaviours. Dell'Acqua and colleagues interpret positive emotional outcomes as evidence of GenAI emulating the social benefits of teamwork. However, their approach to analysis does not address trust and team orientation, which are central to effective teamwork. Establishing trust in teamwork scenarios is challenging and requires some calibration with objective team performance.

Aschenbrenner et al. (2024) argue that humans need to trust AI systems before those systems can actually boost productivity. But Chong et al. (2022) offer a surprising example of how trust in AI can backfire. In a design task, participants who were more confident in AI were less likely to accept its suggestions, while those with higher self-confidence were more likely to accept its suggestions – a dynamic which often led to poor design outcomes. These and other findings highlight the more general challenges of calibrating human trust in AI such that it aligns with objective HAT performance, and developing situationally-aware AI teammates that are capable of adapting in response to performance feedback (see Zhang and Hogan, 2025).

Level 3 - Team development

A shortcoming of *The Cybernetic Teammate* study and other GenAI studies focused on performance and productivity is their limited analysis of team development dynamics. The team development model proposed by Wang and colleagues (2025) highlights how HATs may evolve through different phases, moving from (a) *team formation* to (b) *task-role development*, (c) *team development*, and (d) *team improvement*. Each phase involves distinct challenges. *Team formation* presents unique challenges, including capability mapping and identity negotiation. Humans may need to reconceptualise their professional identity and primary capabilities to allow for AI partnerships and further negotiate how best to maintain essential human functions. *Task-role development* highlights change in task-process architectures as HATs develop. These changes imply dynamic capability matching based on real-time assessment of task demands and environmental challenges. With *team development* the focus switches to building working relationships that accommodate the unique (and often asymmetric) nature of human-AI interaction. This includes developing effective teamwork behaviours, such that members understand when to collaborate, which team member to engage, and what type of support to provide. By building and calibrating trust, establishing responsible reliance, and maintaining mutual respect, teams develop collective efficacy. This foundation enables team members to construct shared mental models that guide their interaction patterns, coordinate effectively with one another, and provide backup behaviour support when needed. Finally*, team improvement* emphasises adaptive capacity building and continuous learning, whereby teams can thrive in complex environments and where levels of social cohesion allow team members to address critical issues like workload balancing and conflict management, ultimately creating self-managing and self-regulating teams capable of distributed leadership.

Wang and colleagues (2025) highlight an ideal that has yet to be achieved. Again, the Cybernetic Teammate study does not analyse any of these human-AI team development dynamics. By focusing on a one-day workshop, the study takes a snapshot at a single point. The primary focus appears to be task-role development -- developing role clarity, capability awareness, and managing task allocation. However, the supporting process isn't fully specified.

Level 4 - Human Critical Leadership

The final level in our analytical framework focuses on human critical leadership. Leadership is an essential function within organisations (Yukl & Gardner, 2020), not only in facilitating effective teamwork but also in providing oversight of HATs and deciding whether they should be included in any organisational process. In figure 1 we specify five dimensions of human critical leadership. *Reflective Oversight* involves meta-cognitive monitoring of both AI system performance and team effectiveness, representing a distinctively human capability that cannot be delegated to AI systems. *Ethical Stewardship* ensures that the operation of HATs serves human values and organisational purposes rather than optimising for metrics that may conflict with broader objectives. *Strategic Vision* maintains long-term perspective and purpose that guides collaboration toward meaningful objectives while preserving human agency in defining goals and directions. *Relationship Management* builds and maintains the social capital essential for organisational effectiveness while preserving human connection and community in AI-mediated work environments. *Planning and Task Management* involves the systematic organisation and coordination of work processes and draws upon distinctively human capabilities for contextual interpretation and adaptive resource allocation, ensuring that workflow structuring serves strategic objectives.

Critical leadership functions are recognised as important in the governance and redesign of organisational systems. Building from a lifespan developmental perspective (Hogan et al., 2023), critical leadership functions are urgently needed in educational institutes to support learning and adaptation in response to rapid GenAI developments. Research in critical studies of AI and education highlights how AI systems risk displacing the collective meaning-making and democratic participation that lie at the heart of pedagogical practice (Holmes et al., 2025), which in turn would have cumulative negative effects for workplace organisational dynamics and societal governance. Educational AI implementations often embed market logics and efficiency metrics that conflict with the slower, more relational processes essential for learning and human development. These threats to collective meaning-making imply the need for sustained human oversight to preserve educational values over algorithmic optimisation. The complexity and opacity of AI systems can also undermine critical analysis and democratic accountability, making the meta-cognitive monitoring capabilities of reflective oversight essential for maintaining institutional integrity. Without renewed efforts to cultivate these critical leadership functions, educational institutions risk becoming sites of technological experimentation rather than spaces for collective agency, democratic participation, and meaningful human connection.

In summary, Level 4 in our framework points to essential human functions that cannot be delegated to AI systems. Reflecting again on the *Cybernetic Teammate* study, it is unclear how an organisation like Procter and Gamble will operationalise Level 4 functions at part of HAT. In educational contexts, we have noted that over-reliance on AI risks eroding students' critical thinking skills through reduced cognitive effort (Dong, Farrell, and Hogan, 2025). While carefully constrained HAT interaction designs can foster critical cognitive engagement (Zhang et al, 2025), the broader education and workplace challenges require careful consideration. For example, as GenAI capabilities develop, how do we sustain interaction designs that prioritise human ethical-critical design thinking as fundamental to HAT dynamics. Delivering and sustaining these interaction designs is a core task-process requirement for human critical leadership across diverse organisational contexts.

We’re only human

Ideally, effective teamwork helps us to recognise and address our individual limitations (Salas et al., 2005). Effective teamwork behaviours include *mutual performance monitoring* and *backup behaviour*: our teammates help us maintain vigilance, identify errors, clarify contextual constraints, and they provide backup behavioural support. The range of situations where we may need help from teammates is boundless. From a team development perspective, it takes time for effective teamwork behaviours to emerge – it takes time to calibrate trust and confidence. The problem of misplaced confidence can be acute when inexperienced humans work with AIs that appear competent and confident. A recent emergency response study revealed shocking gaps between human confidence in AI performance and actual team effectiveness (Grace et al., 2025). In the study, university participants provided first aid to a bleeding (mannequin) patient while receiving instructions from ChatGPT configured as a 911 operator. The AI was prompted to ask questions to clarify the emergency and responded based on verbal inputs received from students. Critically, only 20% of human-AI teams achieved effective performance. Despite this poor performance, students reported high self-efficacy and moderate-high team trust. Analysis of the 911 interactions revealed a static leader-follower dynamic where humans largely followed AI instructions without critical evaluation.

Some participants were given inaccurate or detrimental advice such as trying to apply a tourniquet or pack the wound. Only 48% of participants were asked to assess bleeding severity. Analysis of the interactions indicated a pattern of human cognitive offloading and an assumption of AI expertise that proved dangerous given the AI's inconsistent performance. Critically, the scenario created pressure where humans clearly deferred to AI authority given the role-structured interaction design (see O'Brien and Hogan, 2025). This use case scenario is deeply disturbing and points to broader challenges in the design and development of effective HATs.

Implications for Educational Transformation

The framework presented in this paper highlights the scope of educational transformation needed to prepare individuals for HATs. A key starting point within education is a stronger focus on groupwork and teamwork. A move from monological to more dialogical modes of learning is important for the development of core teamwork and collective intelligence skills (Hogan et al., 2023). When it comes to HAT design, educators and students need to understand the preconditions for effective teamwork with AI partners. This includes careful task-process analysis for each use case in turn to understand key roles and functions, and to clarify the steps in any learning process. HAT training across diverse projects is needed to cultivate effective teamwork behaviours unique to HATs. As AI capabilities continue to evolve, meta-cognitive, ethical, and design thinking skills need to be cultivated on an ongoing basis to build and sustain critical leadership skills.

The pace of GenAI advancement necessitates educational approaches that begin in primary education and continue throughout second and third level education. HAT learning pathways may take on many forms and do not imply a simple embrace of GenAI tools. However, unlike previous technological transitions that allowed more gradual adaptation, GenAI capabilities are evolving rapidly, requiring more urgent learning and adaptation. While it is certainly possible to reject GenAI in any given task-learning process, ignoring GenAI completely is unlikely to support adaptation in the broader societal context where students operate, and it will do little to advance understanding of HAT dynamics. Educational systems must prepare individuals not merely for current AI capabilities but for ongoing evolution as AI systems advance.

Toward Collective Intelligence in the GenAI Era

The emergence of GenAI reveals threats and opportunities for human collective intelligence. Managing threats and realising opportunities requires ongoing learning and adaptation. While GenAI systems offer capabilities for augmenting human cognition and enabling new forms of collective problem-solving, GenAI research and practice to date has prioritised short-term productivity gains over long-term human development. While educational transformation is needed to adapt to AI, our analysis points to a collective intelligence preparedness gap, which includes limited availability of HAT training programmes and systemic inadequacy of educational approaches designed largely for individual intelligence. The acceleration of socio-technical complexity highlights the need for educational innovations that prepares humans for collaboration in organisational contexts where HATs are likely to be increasingly prevalent. The framework we present highlights four levels across which design thinking needs to be oriented if we are to embrace and shape this transformation. In our view, this journey of transformation requires sustained commitment to developing human capabilities alongside technological advancement. The stakes involved extend beyond immediate productivity gains to fundamental questions about human agency, collective capacity, and our ability to shape a future that enhances rather than diminishes human potential in an AI-infused world.

References

Aschenbrenner, D., Jantzen, L., Colloseus, C., Sell, A., Kempen, R., Romero, D., Thürer, M., von Cieminski, G., & Riedel, R. (2024). Research Interpretation of Article 14 of the EU AI Act: Human in Command in Manufacturing. In *Advances in Production Management Systems. Production Management Systems for Volatile, Uncertain, Complex, and Ambiguous Environments* (Vol. 730, pp. 226–239). Switzerland: Springer.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *NBER Working Paper Series*, 31161.

Casebourne, I., Shi, S., Hogan, M., Holmes, W., Hoel, T., Wegerif, R., & Yuan, L. (2025). Using AI to Support Education for Collective Intelligence. *International Journal of Artificial Intelligence in Education*, *35*(3), 1597-1629. https://doi.org/10.1007/s40593-024-00437-7

Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022). Collaborative design decision-making with artificial intelligence: Exploring the evolution and impact of human confidence in AI and in themselves. In *Proceedings of the ASME 2022. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, St. Louis, MO. https://doi.org/10.1115/DETC2022-88574

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., ... & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Working Paper*, 24-013.

Dong, X., Farrell, H., & Hogan, M. (2025, September 7). *How AI Impacts Academic Thinking, Writing and Learning: Does AI make for better grades or better thinkers?* Psychology Today**.** https://www.psychologytoday.com/ie/blog/in-one-lifespan/202509/how-ai-impacts-academic-thinking-writing-and-learning

Farrell, H., Dong, X., & Hogan, M. (2025, September 8). *How Task Design Transforms AI Interactions in the Classroom: Are AI learning tools breaking barriers or building dependency?* Psychology Today**.** https://www.psychologytoday.com/ie/blog/in-one-lifespan/202509/how-task-design-transforms-ai-interactions-in-the-classroom

Forsyth, D. R. (2014). *Group dynamics* (6th ed.). Belmont, CA: Wadsworth Cengage Learning.

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6.

Grace, R., Son, C., Pang, F., Parker, K., & Hernandez, J. (2025, May). Human-AI Communication in First Aid Response: Effects on Team Performance, Trust, and Teamwork Appraisal. In *Proceedings of the International ISCRAM Conference*.

Hogan, M. J., Barton, A., Twiner, A., James, C., Ahmed, F., Casebourne, I., ... & Wegerif, R. (2023). Education for collective intelligence. *Irish Educational Studies*, 42(3), 267-295.

Holmes, W., Mouta, A., Hillman, V., Schiff, D., et al. (2025). Critical studies of artificial intelligence and education: Putting a stake in the ground. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5391793

Hu, K. (2023). ChatGPT sets record for fastest-growing user base. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Hu, M., Zhang, G., Chong, L., Cagan, J., & Goucher-Lambert, K. (2024). How Being Outvoted by AI Teammates Impacts Human-AI Collaboration. *International Journal of Human–Computer Interaction*, *41*(7), 4049–4066. https://doi.org/10.1080/10447318.2024.2345980

Kearney, L., Siwek, A., & Hogan, M. (2025, September 3). *From Tool to Teammate: Here's what GenAI is (and isn't) doing for productivity.* Psychology Today**.** https://www.psychologytoday.com/ie/blog/in-one-lifespan/202509/from-tool-to-teammate

King, A. & Hogan, M. (2025, September 2). *When AI Overrules You: How being outvoted by AI teammates impacts human-AI collaboration.* Psychology Today**.** https://www.psychologytoday.com/us/blog/in-one-lifespan/202508/when-ai-overrules-you

Landers, R. N., Lee, V., & Demeke, S. (2025). Human-Generative AI partnerships: Macro, micro, and ethical considerations. In T.S. Behrend (Ed.), Human-Technology Partnerships at Work (pp. 359-375). Cambridge University Press.

Luther, T., Kimmerle, J., & Cress, U. (2024). Teaming Up with an AI: Exploring Human–AI Collaboration in a Writing Scenario with ChatGPT. *AI*, *5*(3), 1357-1376. https://doi.org/10.3390/ai5030065

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.

O'Brien, S. & Hogan, M. (2025, September 17). *Can We Trust AI to Save Lives? The disconnect between AI confidence and performance in critical scenarios.* Psychology Today**.** https://www.psychologytoday.com/ie/blog/in-one-lifespan/202508/can-we-trust-ai-to-save-lives

Pan, Z., Moore, O. A., Papadimitriou, A., & Zhu, J. (2025). AI literacy and trust: A multi-method study of Human-GAI team collaboration. *Computers in Human Behavior: Artificial Humans*, *4*, 100162. https://doi.org/https://doi.org/10.1016/j.chbah.2025.100162

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:2302.06590

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555-599. https://doi.org/10.1177/1046496405277134

Siwek, A., Kearney, L., & Hogan, M. (2025, September 4). *Teamwork Is King, Even When Generative AI Is Involved: Reflections on a new Cybernetic Teammate Study.* Psychology Today**.** https://www.psychologytoday.com/us/blog/in-one-lifespan/202508/teamwork-is-king-even-when-generative-ai-is-involved

U.S. Government Accountability Office. (2025). *Artificial intelligence: Generative AI's environmental and human effects* (GAO-25-107172).

UK Government. (2025). Safety and security risks of generative artificial intelligence to 2025 (Annex B). *GOV.UK*.

Wang, M., Wu, J., Ma, S., Li, N., Zhang, P., Gu, N., & Lu, T. (2025). Adaptive Human-Agent Teaming: A Review of Empirical Studies from the Process Dynamics Perspective. *arXiv preprint arXiv:2504.10918*.

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X. Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. *Scientific reports*, *15*(1), 15105.

Yukl, G., & Gardner, W. (2020). Leadership in Organizations (9th ed.). Pearson Education.

Zhang, Y. & Hogan, M. (2025, September 9). *What Makes an AI Teammate Truly Intelligent: Why the best AI teammates know when to act and when to ask.* Psychology Today**.** https://www.psychologytoday.com/us/blog/in-one-lifespan/202508/what-makes-an-ai-teammate-truly-intelligent

Zhang, Z., Sun, B., & An, P. (2025). *Breaking Barriers or Building Dependency? Exploring Team-LLM Collaboration in AI-infused Classroom Debate* Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, https://doi.org/10.1145/3706598.3713853