The Promise and Peril of Collaboration: Fostering Appropriate Reliance When Problem-Solving with GenAI

Janet G. Johnson
jgjanet@umich.edu
University of Michigan
Ann Arbor, MI, USA

Steven R. Rick
srick@mit.edu
Massachusetts Institute of Technology
Cambridge, MA, USA

ACM Reference Format:
Janet G. Johnson and Steven R. Rick. 2025. The Promise and Peril of Collaboration: Fostering Appropriate Reliance When Problem-Solving with GenAI. In *Proceedings of ACM Conference (Conference'25).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnn.nnn

1 INTRODUCTION

Generative AI (GenAI) has exploded into the public eye and become increasingly ubiquitous when people work with computers. This technology has changed the way we separate human work from computer work, with GenAI now presenting the ability to produce human-like multimedia artifacts of text, images, audio, and more. These advances have been shown to offer significant opportunities to augment human capabilities [1, 8, 9, 26]. Furthermore, GenAI systems are evolving to the point where they can be used to somewhat autonomously complete complex tasks with minimal human guidance [3, 21, 25, 39].

While integrating GenAI's autonomous and seemingly human-level capabilities into high-stakes and complex environments could improve efficiency, increase access to expert knowledge, and introduce innovative problem-solving methods, there is a growing concern that people may increasingly offload critical intellectual processes and cognitive engagement to these systems. The fear that GenAI might lead humanity at large to think less is not purely speculative – while the benefits have been extolled widely, there is emerging work showing the side effects of these technologies at large [15, 30, 36]. For example, there is evidence that the use of GenAI for problem-solving frequently leads to over-reliance and reduces critical engagement with the task [4, 16, 24].

In some of our own work, we've witnessed that scaffolding a person's use of a Large Language Model (LLM) can improve the innovativeness of solutions [14] and that using LLMs to generate new ideas can improve creative inspiration, enjoyment, and utility [5]. But in those same works, we found that experimental participants did not really use the interventional systems any differently than off-the-shelf GenAI solutions. Users engaged in the minimal amount of work needed to get their task done, as the GenAI system handled much of the burden of knowledge synthesis, and people often trusted output from the systems to be "good enough" rather than refining those outputs. While this may be relatively benign in creative domains that have no definitive right or wrong answers, it increasingly presents challenges and harms not yet realized when applied to domains like governance or education where misinformation and cognitive disengagement pose real risks [38].

Conference'25, April 2025, Washington, DC, USA
2025. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnn.nnn

Designing GenAI systems therefore raises a fundamental challenge: How do we design these systems to maintain, and ideally amplify, human cognitive abilities without leading to deskilling, cognitive disengagement, or detrimental over-reliance? As more capable GenAI technologies are developed, they need to be designed to support complex and creative problem-solving and ensure that humans remain actively engaged in the intellectual process.

Recent work in HCI has increasingly explored approaches to foster appropriate reliance on technology [6, 29, 33]. However, one approach to reducing cognitive disengagement and over-reliance that has received little attention is enabling multi-user interactions with GenAI to leverage the collective intelligence of groups during collaborative problem-solving. In this paper, **we posit that collaborative use of GenAI offers an underexplored yet promising framing to foster critical thinking and appropriate reliance**, and discuss the core focus areas and considerations involved in developing this research agenda.

2 COGNITIVE ENGAGEMENT THROUGH GROUPWORK

The case for collaborative GenAI systems

Today, many GenAI-powered systems and tools are designed with the goal of augmenting a single individual's capabilities and solving problems as directly as possible. However, this does not reflect the reality of how we tackle complex and wicked problems in the world – which is often through a collaborative process where multiple individuals contribute ideas to a larger pool of thought, and problem-solving unfolds across a mix of individual and group work phases [12, 27]. Research has also consistently demonstrated that groupwork encourages individuals to stay actively engaged and that collaborative problem-solving enhances critical thinking [11, 12, 37].

Collaborative workflows could serve as a natural counterbalance to some of the cognitive pitfalls of individual interactions with GenAI. Group interactions could allow individuals to challenge, scrutinize, and refine generated ideas or content (whether by a human or through computational means), reducing the likelihood that AI generated content is accepted uncritically. For example, only one member needs to question, evaluate, or be skeptical of contributions to prompt the group as a whole to reconsider them.

Groups also naturally engage in collaborative boundary regulation [34] which is a process of establishing limits and ground rules to maintain a healthy group environment. This mechanism would enable them to dynamically regulate when and how GenAI contributes to manage the level of influence machine generated output has on both the process and the outcomes of the group.

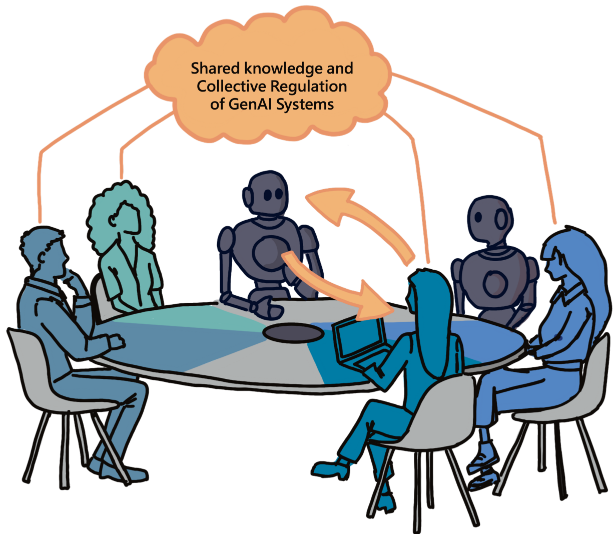

Figure 1: An artistic depiction of collaborative problem-solving including individual and group interactions with GenAI where the shared knowledge and collective regulation mechanisms of the group can foster appropriate reliance.

This ability to regulate GenAI's role could prevent it from dominating creative and problem-solving processes while still leveraging the strengths of the technology. Social dynamics and pressures (like not wanting to be seen as overly-reliant on technology) could further reinforce this balance and safeguard groups from cognitive disengagement.

The challenge with group interactions

While collaborative problem-solving with GenAI offers some potentially significant advantages, groups also have inherent limitations. Our current research and ongoing explorations around collaborative GenAI agents [17] suggests that while the collective can critically evaluate agent contributions, the complexities of group interaction make them susceptible to cognitive biases and social dynamics that could result in over-reliance or undue influence of GenAI-content on group outcomes.

Introducing one or more GenAI agents into creative and collaborative problem-solving processes therefore adds another layer of complexity as some limitations of group work like groupthink, premature fixation, and social loafing [22, 23, 28, 35] can be exacerbated – for instance – due to the output from GenAI technologies being perceived as knowledgeable expertise or because dominant members default to machine generated contributions without scrutiny.

3 DESIGNING FOR APPROPRIATE RELIANCE

Exploring the collaborative use of GenAI is still a nascent and largely unexplored research space [31, 32] and fully realizing the benefits of a group will require the discourse around GenAI to evolve and shift away from from individual use to consider facilitating collaborative multi-user engagements. Supporting the reality of knowledge work and creative problem solving will also need these systems to support both individual and group-based work and ensure smooth transitions between different modes of collaboration for seamless integration across different phases of the problem-solving process.

In this section, we discuss the key aspects that we believe need to be addressed in order to truly protect human critical thinking and design for appropriate reliance on technology when using GenAI in complex and open-ended problem-solving.

Detecting Over-reliance and Cognitive Pitfalls

The key question in human-GenAI interaction (at both group and individual levels) isn't whether or not we should but rather how much we can rely on AI. Some reliance on GenAI can be both appropriate and beneficial as the ability to offload certain aspects of our current processes to technology is precisely what will enable humans to engage in more creative and higher-order tasks. The real danger with GenAI arises when over-reliance on the machine generated output leads to disengagement from critical and creative thinking entirely.

This means that some deskilling is acceptable, as long as humans remain engaged in the higher-order thinking that GenAI cannot – and probably should not – replace. A foundational challenge ahead of us in going to be understanding which cognitive tasks or aspects of the problem-solving process we can afford to offload without compromising our ability to think independently, and ensuring that GenAI doesn't serve as an unwelcome source of cognitive atrophy.

Designing systems that foster appropriate reliance on GenAI will also require us to develop systems capable of detecting when users are overly-reliant on the technology and then intervene to promote cognitive engagement. While some recent work has explored interventions like cognitive forcing functions and reflective interactions [4, 40, 41] for individual users, we don't fully understand how to detect reliance patterns with multiple users. This will likely involve behavior sensing and real-time analysis of the evolving social dynamics and communication patterns within the group. For example, building on theories of group dynamics like the social influence theory [19] could help us detect when GenAI is becoming dominant or when people defer to it without much critical thought. Detecting when groups engage in social loafing (which is when individuals disengage and assume another party will carry the conversation or process forward) or social matching (which is the tendency for individuals to conform to others) [23], or detecting when GenAI agents reinforce preexisting group assumptions, all present promising pathways for research around how to understand group mechanisms in the presence of GenAI to avoid local maxima or mitigate group biases.

Designing Implicit Cues and Adaptive Interfaces

In order to ensure that humans remain active participants rather than passively accepting GenAI contributions in lieu of doing cognitive work, new mechanisms need to be designed that support the cognitive engagement of users. A key aspect here will be in balancing the influence of GenAI on the creative problem-solving process – this entails allowing generative technologies to provide new perspectives, novel ideas, and valuable insights, while also preventing it from becoming a dominant voice with undue influence on a group.

Complex problem-solving is already a cognitively demanding process, and adding GenAI to the mix – especially in collaborative processes where users are interacting with other humans as well – might only exacerbate this. Given the capabilities of the technology to rapidly synthesize deep knowledge and complex topics, these new technologies are more likely to be used in tasks where users might lack the skill or expertise for them to spot issues with misinformation or other inaccuracies, potentially making them more vulnerable to over-reliance. Therefore, we strongly believe over-reliance mitigation strategies must be seamless and non-intrusive, aligning with System 1 (intuitive) thinking [18] rather than demanding additional cognitive effort. The use of more implicit cues (rather than overt warnings) in adaptive interfaces could signal a change in the role, type of contribution, or interaction style for GenAI when over-reliance is detected. Subtle shifts in AI-behavior, like shifting to counter-arguments or prompting for more user interaction, could encourage deeper engagement that is more intuitive without disrupting the natural workflow.

Making use of more expressive and collaborative mediums like Mixed Reality (MR) [2] might also be worth exploring here. With MR, GenAI agents can take on embodiments and socio-spatial behaviors that can be intentionally designed to encourage appropriate reliance within groups. For example, the spatial positioning of GenAI agents within an MR space could more easily build on theories like proxemics [13], F-formation [20], and spatial syntax theory [10] to signal the appropriate influence or prominence a GenAI agent should have at any given point in time.

Fostering Work Practices that Protect Human Creativity and Agency

While technological and system-design approaches might help mitigate over-reliance on GenAI, it is critical to also explore the ways new and emerging work practices can be designed to optimize GenAI-assisted work while still keeping humans involved. Understanding how we organize ourselves collectively as human beings will be key to fostering environments that protect both the individual creative process and the group's critical thinking abilities.

Learning from existing collective intelligence frameworks and collaboration styles could help us understand the different ways of interacting with GenAI in a collaborative settings. For example, roundtable discussions encourage equal participation of all members, while design thinking exercises or processes like brainwriting help structure both individual and collective cognition to enhance productive collaboration. Labor division across roles can also foster critical discourse by allowing different group members to contribute multiple perspectives and considerations during decision making. These are well studied practices in human work, but not yet well understood when machines join human teams.

It is increasingly evident that GenAI will fundamentally reshape the future of work [7] and transform the dynamic of how people work with technology and one another, with many of its consequences still unfolding. It is therefore also critical to identify which GenAI-specific practices and organizational dynamics should be promoted to support the effective integration of this technology as it is increasingly adopted in the workforce.

Going beyond having GenAI as a static participant in problem-solving and work processes to enable adaptive configurations would give individuals and groups the agency to dynamically moderate GenAI's influence on group outcomes. This includes developing flexible interaction models that allow users to adjust the role GenAI plays in their work (i.e. an active brainstorming partner, a fact-checker or red teamer, a quiet observer that only intervenes under specific conditions, a facilitator, etc). We will also need to develop mechanisms for calibrating the level of GenAI interference, and changing the mode of engagement between human and machine – from synthesis to critical reflection and beyond.

Moreover, while roundtable discussions or group deliberations might help multiple users collectively fact-check or scrutinize GenAI, individual or sub-group breakout sessions could help limit how GenAI might influence the group as a whole. Future GenAI tools should therefore also be designed to support fluid transitions between individual and group engagement while facilitating different types of group structures and work styles.

Encouraging people to engage in different collaboration styles with the technology allows us to develop best practices for not just when to use GenAI but also *how* to use GenAI across various problem-solving tasks. In general, this approach would help identify where the technology serves as a catalyst for deeper engagement and augmented capabilities instead of just a superficial shortcut or replacement for human effort.

4 CLOSING REMARKS

Acknowledging that GenAI is a relatively new and not yet well-understood technology (with regard to cognitive consequences), it is important that HCI as a field gets ahead and proactively addresses the risks over-reliance and cognitive disengagement. We believe that a combination of approaches and multiple lenses is crucial to making measurable progress toward methods and frameworks to preserve human cognition. Beyond the interventional approaches, we also need ways to understand human behavior when using GenAI and establish methods for assessing whether or not these interventions are proving effective and useful. We hope this kick-starts a much longer conversation and body of work around successful human-AI teamwork, and we are enthusiastic about delving into these concepts and collectively sculpting the future of human-computer teams.

5 ABOUT THE AUTHORS

Janet G. Johnson is a Postdoctoral Research Fellow in the School of Information at the University of Michigan. Her research focuses on designing intelligent spaces for collaboration that leverage the spatial capabilities of Mixed Reality and the generative capacities of AI for complex problem-solving. Janet received her Ph.D. from UC San Diego where she worked on designing and evaluating spatial interfaces that can go beyond realism and serve as cognitive aids during collaborative physical tasks.

Steven R. Rick is a Postdoctoral Research Associate at MIT in the Center for Collective Intelligence, where he leads the design, development, and study of systems that aim to augment human creative problem-solving with computational teammates. His research spans the fields of collective intelligence, design, user experience, and human behavior sensing. Steven plays the role of both technology evangelist and technology skeptic to craft the most effective human-computer teams to solve real-world problems.

REFERENCES

[1] Volker Bilgram and Felix Laarmann. 2023. Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. *IEEE Engineering Management Review* 51, 2 (2023), 18–25.

[2] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (2002), 64–70.

[3] Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. 2023. Autonomous chemical research with large language models. *Nature* 624, 7992 (2023), 570–578.

[4] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction* 5, CSCW1 (2021), 1–21.

[5] Alice Cai, Steven R Rick, Jennifer L Heyman, Yanxia Zhang, Alexandre Filipowicz, Matthew Hong, Matt Klenk, and Thomas Malone. 2023. DesignAID: Using generative AI and semantic diversity for design inspiration. In *Proceedings of The ACM Collective Intelligence Conference*. 1–11.

[6] Shiye Cao, Anqi Liu, and Chien-Ming Huang. 2024. Designing for Appropriate Reliance: The Roles of AI Uncertainty Presentation, Initial User Decision, and User Demographics in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–32.

[7] Mauro Cazzaniga, Ms Florence Jaumotte, Longji Li, Mr Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Ms Marina Mendes Tavares. 2024. *Gen-AI: Artificial intelligence and the future of work*. International Monetary Fund.

[8] Edwin Creely. 2024. Exploring the Role of Generative AI in Enhancing Language Learning: Opportunities and Challenges. *International Journal of Changes in Education* (2024).

[9] Anil R Doshi and Oliver P Hauser. 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances* 10, 28 (2024), eadn5290.

[10] Pelin Dursun. 2007. Space syntax in architectural design. In *6th international space syntax symposium*. 01–56.

[11] Stephen M Fiore, Michael A Rosen, Kimberly A Smith-Jentsch, Eduardo Salas, Michael Letsky, and Norman Warner. 2010. Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human factors* 52, 2 (2010), 203–224.

[12] Vera Hagemann and Annette Kluge. 2017. Complex problem solving in teams: the impact of collective orientation on team process demands. *Frontiers in psychology* 8 (2017), 1730.

[13] Edward T Hall. 1966. The hidden dimension. *Garden City* (1966).

[14] Jennifer L Heyman, Steven R Rick, Gianni Giacomelli, Haoran Wen, Robert Laubacher, Nancy Taubenslag, Max Knicker, Younes Jeddi, Pranav Ragupathy, Jared Curhan, et al. 2024. Supermind Ideator: How scaffolding Human-AI collaboration can increase creativity. In *Proceedings of the ACM Collective Intelligence Conference*. 18–28.

[15] Wayne Holmes. 2023. The unintended consequences of artificial intelligence and education. (2023).

[16] Maia Jacobs, Melanie F Pradier, Thomas H McCoy Jr, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. 2021. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational psychiatry* 11, 1 (2021), 108.

[17] Janet G. Johnson, Macarena Peralta, Mansanjam Kaur, Ruijie Sophia Huang, Sheng Zhao, Ruijia Guan, Shwetha Rajaram, and Michael Nebeling. 2025. Exploring Collaborative GenAI Agents in Synchronous Group Settings: Eliciting Team Perceptions and Design Considerations for the Future of Work. *To Appear in Proceedings of the ACM on Human-Computer Interaction, CSCW (2025)*.

[18] Daniel Kahneman. 2011. Thinking, fast and slow. *Farrar, Straus and Giroux* (2011).

[19] Herbert C Kelman. 2006. Interests, relationships, identities: Three central issues for individuals and groups in negotiating their social environment. *Annu. Rev. Psychol.* 57, 1 (2006), 1–26.

[20] Adam Kendon. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. Vol. 7. CUP Archive.

[21] Geunwoo Kim, Pierre Baldi, and Stephen McAleer. 2023. Language models can solve computer tasks. *Advances in Neural Information Processing Systems* 36 (2023), 39648–39677.

[22] Mark Klein. 2011. How to harvest collective wisdom on complex problems: An introduction to the mit deliberatorium. *Center for Collective Intelligence working paper* 79 (2011).

[23] Nicholas W Kohn and Steven M Smith. 2011. Collaborative fixation: Effects of others' ideas on brainstorming. *Applied Cognitive Psychology* 25, 3 (2011), 359–371.

[24] Hao-Ping Hank Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. 2025. The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. (2025).

[25] Vadim Liventsev, Anastasiia Grishina, Aki Härmä, and Leon Moonen. 2023. Fully autonomous programming with large language models. In *Proceedings of the Genetic and Evolutionary Computation Conference*. 1146–1155.

[26] Jijian Lu, Ruxin Zheng, Zikun Gong, and Huifen Xu. 2024. Supporting teachers' professional development with generative AI: The effects on higher order thinking and self-efficacy. *IEEE Transactions on Learning Technologies* (2024).

[27] Thomas W Malone. 2018. *Superminds: The surprising power of people and computers thinking together*. Little, Brown Spark.

[28] Michael Pacanowsky. 1995. Team tools for wicked problems. *Organizational dynamics* 23, 3 (1995), 36–51.

[29] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R Varshney, Amit Dhurandhar, and Richard Tomsett. 2022. Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–22.

[30] Roberto Redaelli. 2024. Intentionality gap and preter-intentionality in generative artificial intelligence. *AI & SOCIETY* (2024), 1–8.

[31] Isabella Seeber, Eva Bittner, Robert O Briggs, Triparna De Vreede, Gert-Jan De Vreede, Aaron Elkins, Ronald Maier, Alexander B Merz, Sarah Oeste-Reiß, Nils Randrup, et al. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & management* 57, 2 (2020), 103174.

[32] William Seymour and Emilee Rader. 2024. Speculating About Multi-user Conversational Interfaces and LLMs: What If Chatting Wasn't So Lonely?. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*. 1–4.

[33] Hua Shen, Tiffany Knearem, Reshmi Ghosh, Kenan Alkiek, Kundan Krishna, Yachuan Liu, Ziqiao Ma, Savvas Petridis, Yi-Hao Peng, Li Qiwei, et al. 2024. Towards Bidirectional Human-AI Alignment: A Systematic Review for Clarifications, Framework, and Future Directions. *arXiv preprint arXiv:2406.09264* (2024).

[34] Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*. 769–778.

[35] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science* 185, 4157 (1974), 1124–1131.

[36] Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review* 11, 2 (2023), 7–30.

[37] Enwei Xu, Wei Wang, and Qingxia Wang. 2023. The effectiveness of collaborative problem solving in promoting students' critical thinking: A meta-analysis based on empirical literature. *Humanities and Social Sciences Communications* 10, 1 (2023), 1–11.

[38] Lixiang Yan, Samuel Greiff, Ziwen Teuber, and Dragan Gašević. 2024. Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour* 8, 10 (2024), 1839–1850.

[39] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems* 36 (2024).

[40] ShunYi Yeo, Gionnieve Lim, Jie Gao, Weiyu Zhang, and Simon Tangi Perrault. 2024. Help Me Reflect: Leveraging Self-Reflection Interface Nudges to Enhance Deliberativeness on Online Deliberation Platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–32.

[41] Liudmila Zavolokina, Kilian Sprenkamp, Zoya Katashinskaya, Daniel Gordon Jones, and Gerhard Schwabe. 2024. Think fast, think slow, think critical: designing an automated propaganda detection tool. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–24.