Tools for Thought: Understanding, Protecting, and Augmenting Human Cognition with Generative AI—From Vision to Implementation

Zelun Tony Zhang
zelun.tony.zhang@tum.de
TU Munich
Munich, Germany
Munich Center for Machine Learning (MCML)
Munich, Germany

Nick von Felten
nick.vonfelten@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Leon Reicherts
leon.reicherts@unisg.ch
University of St. Gallen
St. Gallen, Switzerland

Lev Tankelevitch
lev.tankelevitch@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Zhitong Guan
klarazt@utexas.edu
The University of Texas at Austin
Austin, USA

Sean Rintel
serintel@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Yue Fu
chrisfu@uw.edu
University of Washington
Seattle, USA

Jessica He
jessicahe@ibm.com
IBM Research
USA

Kenneth Holstein
kjholste@andrew.cmu.edu
Carnegie Mellon University
USA

Advait Sarkar
advait@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Gonzalo Ramos
gonzo.ramos@gmail.com
Independent Researcher
Kirkland, Washington, USA

Anuschka Schmitt
a.schmitt2@lse.ac.uk
The London School of Economics and Political Science
United Kingdom

Anjali Singh
anjali.singh@ischool.utexas.edu
The University of Texas at Austin
USA

Haotian Li
haotian.li@microsoft.com
Microsoft Research Asia
Beijing, China

Srishti Palani
srishti.palani@salesforce.com
Tableau Research
Palo Alto, California, USA

Peter Dalsgaard
dalsgaard@cavi.au.dk
Aarhus University
Aarhus, Denmark

Abstract

Building on the first Tools for Thought (TfT) workshop at CHI 2025, we invite researchers, designers, and practitioners to further operationalise approaches for the design, usage, and evaluation of Generative Artificial Intelligence (GenAI) as a TfT. The first goal of the workshop is to put into focus which *outcomes* a TfT should help people achieve to effectively augment their cognition while avoiding its erosion. Secondly, we will explore how to achieve these outcomes through design and usage *strategies*. Third, the workshop will also address what a TfT needs for its successful *adoption* and integration into people's flow, so that they can benefit from the tools' potential in their own terms. By focussing on these three research goals, the workshop aims to further develop and advance the multidisciplinary TfT community interested in exploring research frameworks, theories, methods, and approaches to conceptualising, designing, and researching GenAI as a TfT.

CHI EA '26, April 13–17, 2026, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/2026/04
https://doi.org/10.1145/3772363.3778733

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computing methodologies** → **Artificial intelligence**.

Keywords

generative AI, artificial intelligence, cognition, metacognition, critical thinking, reasoning, learning, diversity, creativity, sensemaking, autonomy, augmentation, intentionality, reflection, social science, research, design, workshop

ACM Reference Format:

Zelun Tony Zhang, Nick von Felten, Leon Reicherts, Lev Tankelevitch, Zhitong Guan, Sean Rintel, Yue Fu, Jessica He, Kenneth Holstein, Advait Sarkar, Gonzalo Ramos, Anuschka Schmitt, Anjali Singh, Haotian Li, Srishti Palani, and Peter Dalsgaard. 2026. Tools for Thought: Understanding, Protecting, and Augmenting Human Cognition with Generative AI—From Vision to Implementation. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3778733

1 Motivation and Goals

"Not another Generative AI workshop—again?!" At CHI 2025, the Tools for Thought workshop [19] differentiated itself from other GenAI-focused workshops by treating higher-order cognition as both its central design object and evaluation target. By *Tools for Thought (TfT)*, we refer to the characterisation of (sociotechnical) systems that put special emphasis on their impact on cognition — in both the short and long term — over mere productivity. Rather than focussing solely on task completion or an artefact being produced, this conceptualisation places the focus on how systems affect and can be designed to support higher cognitive functions, such as reasoning, memory, ideation, or critical thinking [20]. Last year's discussions started mapping (a) the ways in which GenAI, when used to support human cognition, can in turn affect it, and (b) how tools might best be designed to protect and augment it. Building on these workshop contributions and discussions [20], our aim in this year's iteration is to transition from the foundational to the operational. Specifically, we ask: what makes a TfT effective? What are the concrete mechanisms and appropriate outcomes of TfTs? And how can we establish and validate sensible metrics to measure their impact on human cognition? However, even though the focus of this workshop lies on the operationalisation of TfT, we recognise the importance of leaving space for reflection, definition, and further mapping of the field. As TfTs are an emerging research area, we expect to find gaps in our understanding of them. Thus, trying to pursue the operational will inevitably point us back to more abstract questions, which will be crucial for advancing TfT as a field.

This focus is timely. As GenAI tools proliferate in many aspects of our lives, so do our hopes and fears about their impact on human cognition. Prior work has found worsened learning outcomes and decreased metacognitive effort resulting from GenAI use, both in knowledge work and in education [3, 4, 7]. For instance, while GenAI can boost productivity in professional writing tasks [10], its impact on learning is more complex. Studies show that using LLMs for direct solutions can impair learning, whereas using them for explanation-seeking can be beneficial [8]. This can be further complicated by a perception-performance paradox, where students may perceive LLMs as more helpful than traditional study methods, such as note-taking, even when their learning outcomes are worse [6]. On the other hand, if designed and used in meaningful ways, there is much potential for cognitive augmentation. For example, AI assistance can enhance skill development by providing high-quality, personalised examples, acting as a powerful learning aid [9].

There is a need for concrete pathways toward GenAI as an effective and positive TfT. By moving from vision to implementation, we propose three interrelated, specific areas of contribution for this year's workshop (a) **TfT strategies** (Section 1.1): how GenAI can be designed and applied as a TfT. This involves mapping out proven design patterns and principles, and identifying successful usage strategies for the application of GenAI as a TfT. (b) **TfT outcomes** (Section 1.2): evaluating how a TfT affects human cognition and working processes and how it aligns with a person's goals. This includes both defining what desirable outcomes are, and how to measure them. (c) **TfT adoption** (Section 1.3): understanding what factors influence adoption or refusal, and helping individuals, teams, and/or organisations experience/see the value of a TfT (e.g. where in a task it is meaningful to be more cognitively involved and why).

1.1 Tools for Thought Strategies

Under this theme, we seek tangible strategies for realising GenAI-based TfT. We distinguish two broad types of strategies:

- **Design strategies:** These could be design principles or patterns that have supporting evidence of protecting or augmenting human cognition. For instance, during last year's workshop, one common thread that emerged was a strategy that can be broadly characterised as *process-oriented support* [24]. This strategy focusses on supporting users to reason *forward* to their own solution, rather than *backward* from an AI-generated one. What can such a type of support look like in various contexts? And what other design strategies exist?
- **Usage strategies:** Users themselves can also employ workflows or activities involving GenAI in a way that augments or protects their cognition, even when the system is not specifically designed as a TfT, such as general purpose chatbots like ChatGPT. For example, while much of the debate about GenAI in education is concerned with how easy it becomes for students to cheat, many students also find potentially learning-enhancing uses of GenAI, such as generating quizzes, adapting the learning content to their specific needs, and discussing ideas [21].

Many of last year's submissions [20] provide concrete examples for both types of strategies. However, they are mostly specific designs, with limited transferability to other domains or problem settings. How can we systematise these examples into tangible strategies that can reliably lead to effective and positive TfTs, given a new problem domain or context? Can we theoretically ground these strategies, and do we have convincing empirical evidence for their effectiveness in protecting and augmenting human cognition (see Section 1.2)? What are useful framings for the role of GenAI (e.g., AI as provocateur [14] or facilitator [25]) that can inspire effective design and usage strategies?

1.2 Tools for Thought Outcomes

One of the ways in which we can assess what makes a TfT effective is not only whether it helps users complete a task—such as writing a report, implementing a software feature, or completing an assignment—but whether it augments or supports the cognitive processes that underlie human thought. We therefore envision outcomes of TfTs as much more than task performance: They are about how tools shape, support, or extend thinking itself. TfT outcomes should capture results of thought-intensive activities such as decision-making, reflection, ideation, or exploration, contexts where conventional and more utilitarian notions of "efficiency" or "success" may not apply, or only represent a fraction of these activities' purpose. For such activities, the "outcome" of using a TfT may be more concerned with how it reshapes the process rather than its final outputs—for example, when supporting a sensemaking activity by taking the role of a socratic tutor and facilitating users' critical engagement with a topic.

As part of this workshop, we focus on defining what could count as a desirable outcome for TfTs that is aligned with people's goals, and how such an outcome could be meaningfully measured. Drawing from Rogers et al. [13] we distinguish three broad types of outcomes and welcome more:

- **Intermediary Outcomes**: Tangible artefacts, such as notes or diagrams, that emerge during TfT use. Though not ends in themselves, these artefacts scaffold thinking by externalising abstraction, inference, and reflection. The use of AI may "skip" certain intermediary outcomes (e.g. skipping low-fidelity prototypes since GenAI encourages jumping straight to high fidelity) but it may also introduce new ones.
- **Cognitive Outcomes**: Changes in understanding, learning, or cognitive skills, such as critical thinking and metacognitive engagement. For instance, a TfT might help a user form a clearer mental model or practice more accurate self-monitoring [17]. These outcomes are central to cognitive protection and augmentation.
- **Task Outcomes**: Though not necessarily the primary focus for TfTs, task outcomes, including performance, remain relevant. A good TfT should support cognition without undermining task completion (see Section 1.3).

Some outcomes may be hard to quantify or emerge only in users' own narratives of what they believe to have achieved. This calls for evaluation methods that are flexible and pluralistic, capturing not just what users produce, but also how they think and change in the process. Evaluations also attend to context—the same outcome may hold different meanings depending on the users' goals or circumstances of use. Useful evaluation approaches may include: process tracing (e.g., interaction logs, eye-tracking) to reveal interactions with a TfT and flow of thought; artefact analysis to examine the structure or integrative quality of outputs; cognitive and metacognitive assessments to measure understanding, learning, and reflection; longitudinal studies to identify patterns and track changes in cognitive style and problem-solving over time.

1.3 Tools for Thought Experience and Adoption

In times when GenAI promises productivity gains, design and usage strategies that introduce friction can be met with resistance, even when they are intended to improve cognitive outcomes. This resistance can arise from both users and stakeholders. For example, Ashktorab et al. [2] found that users had negative impressions of a cognitive forcing function that created more work for them. Kazemitabaar et al. [5] similarly discuss the need to balance performance with the amount of added task load and friction when designing cognitive engagement techniques to support programming education. In domains and environments that value productivity, TfTs may also be unwelcome to users and stakeholders who expect GenAI to accelerate workflows. A global study of 2500 executives and workers found that 96% of C-suite executives expect AI to boost productivity [12], and EY's AI Anxiety in Business Survey similarly found that 66% of employees were concerned about falling behind if they did not use AI [1].

In support of our goal of putting more into focus our understanding of implementing TfTs in practice, this workshop will also explore questions surrounding practical strategies to overcome barriers to adoption. For example, how might we foster people's intrinsic motivation to use GenAI in ways that support improved cognitive outcomes? What strategies can we employ to communicate the value of TfT to stakeholders? How might we manage potential trade-offs between productivity and cognitive outcomes to maximise the likelihood of their successful integration and sustained use? Are there strategies (see Section 1.1) that resolve the conflict between productivity and cognitive outcomes and allow users to see the value in and attain both?

2 Workshop Activities

We propose a 180-minute workshop, divided into two 90-minute sessions. To align with this year's format and to make the best use of participants' submissions, the workshop will be organised around small groups working on specific topics. Submissions should be within at least one of the areas of contribution laid out in Section 1. The group-based format should help participants quickly identify overlaps and target specific TfT aspects that they are all interested in and want to further operationalise during (and beyond) the workshop.

2.1 Before the Workshop

Workshop submissions consist of (a) a workshop paper (up to four pages without references, 2-column template) and (b) the completion of a Miro board template that prompts participants to focus on tangible outcomes related to the workshop themes. The template also allows participants to provide information on their current career stage, the main challenges in their current research, and key workshop interests. This will give us further context for group formation. The Miro boards of accepted submissions will be thematically clustered by the organisers to form workshop groups of four to five participants each.

Once groups are established, participants will be asked to read the submissions of their group members to establish a good understanding within groups prior to the workshop, and to provide a scaffold for conversations and collaborations during the workshop. We will emphasise the importance of this throughout the submission phase and before the workshop. During this stage, participants may also request switching their group.

2.2 During the Workshop

The bulk of the workshop time will be allocated to work within the groups. After a short ice-breaker session and aligning their interests, the groups will proceed to work in depth on a tangible output of their own choice (e.g., low-fidelity prototyping a new TfT in a hackathon style, mapping out a new framework, drafting design principles, developing theory, etc.). After the group work, each group will present their results in the form of short pitches to the plenum during the final part of the workshop. Table 1 shows our proposed workshop schedule.

Table 1: Proposed workshop schedule.

Time	Activity
10 min	Welcome and gathering in groups
10 min	Within-group ice-breaker session
20 min	Discussion of submissions and group alignment
5 min	Decision on output to pursue within each group
45 min	Group work on tangible outputs (such as a TfT concept or mockup, study proposal draft etc.)
Break	*Coffee break, with encouragement to glance at other groups' work-in-progress*
40 min	Continued group work on tangible outputs
45 min	Short pitches of group results to the plenary
5 min	Next steps and closing reflections

2.3 After the Workshop

Our pre-grouping and Miro collaboration are designed to encourage discussions and collaborations beyond the workshop. We will strongly encourage participants to continue and refine their group work afterwards. To further evolve the TfT research agenda, we will conduct a post-hoc gap-finding exercise to identify areas not fully covered by the workshop contributions. This allows us to move back from the operational to the foundational, with strong potential to uncover areas worth exploring, which again might spark future collaborations. Our post-workshop process will culminate in a synthesis report similar to last year's [20] that brings together the submissions, workshop outcomes, and the identified gaps from the gap-finding exercise. The synthesis report will be publicly shared.

3 Accessibility

As in the previous year, we encourage submissions in accessible formats and will use platforms that support speech captions. We will also contact participants in advance to identify any accessibility needs and accommodate them, with the support of the CHI accessibility chairs if necessary.

4 Offline Materials

We will use last year's workshop website (https://ai-tools-for-thought.github.io/workshop/) to share information, including advance publication of accepted submissions. Post-workshop, a written synthesis of activities and outputs, a showcase of curated artifacts, and results of the gap finding activity will be published on the website and advertised on social media. For asynchronous communication and media sharing among participants before, during, and after the workshop, we will make use of the existing TfT Discord channel.

5 Publishing Plans

Paper submissions will be stored and distributed via GitHub and linked through the workshop website. As in the previous year, a synthesis of the workshop contributions will be published on CEUR-WS or arXiv.

6 Organisers

Our team brings together a wide range of perspectives on TfTs, spanning both academic and industry researchers with diverse scientific backgrounds and levels of seniority. We include a mix of new and returning workshop organisers, with expertise spanning HCI, computer science, cognitive science, learning science, and information science, thus providing the breadth of perspectives needed to study and develop TfTs. All organisers will review submissions, take part in developing and assigning workshop groups, and moderate and document group discussions during the workshop. Other roles are distributed as listed in Table 2.

Tony Zhang is a postdoctoral researcher at TU Munich. He is currently researching how educational generative AI tools can be built for learner curiosity and active engagement rather than passive consumption. Previously, he researched how AI-based decision support tools can support users while actively engaging them in their decision-making task.

Nick von Felten is a psychologist and doctoral researcher at the University of St. Gallen, Switzerland. His current work investigates how human and AI biases interact and shape attitudes and decision-making, with the goal of designing AI tools calibrated to human cognitive tendencies [22, 23].

Leon Reicherts is a postdoctoral researcher at the University of St. Gallen. He was previously part of the *Tools for Thought* team at Microsoft Research. His research is concerned with how (Gen)AI tools affect augment human cognition—particularly decision-making and learning—and how they can be best designed to support it.

Lev Tankelevitch is a researcher at Microsoft Research within the *Tools for Thought* group. His research uses metacognition as a lens to augment intentionality and human-AI interaction in collaborative knowledge work. He has a background in behavioural science and in cognitive neuroscience. Find out more at: https://aka.ms/levt.

Zhitong Klara Guan is a PhD student at the University of Texas at Austin. Her current work investigates how Generative Interactive Information Retrieval can be conceptualised, designed and evaluated to augment essential human cognitive skills by supporting and extending the information seeking and use.

Sean Rintel is a Senior Principal Research Manager at Microsoft Research where he co-leads the *Tools for Thought* project [11]. His research is currently focused on AI collaboration and intentional meetings. See: https://aka.ms/seanrintel.

Yue Fu is a PhD candidate at the University of Washington Information School. His research focusses on understanding how

Table 2: Organisation tasks and designated leads for the workshop.

Task	Lead
Reviewing submissions	All organisers
Assigning workshop groups	All organisers
Group discussion moderation & documentation	All organisers
Lead organisers	Tony Zhang, Leon Reicherts, Nick von Felten
Submission and review coordination, author notifications	Zhitong Guan, Nick von Felten
Miro board set up and management	Leon Reicherts
Pre-workshop communication with participants	Yue Fu, Tony Zhang
Site handling before & during workshop	Jessica He, Srishti Palani
Workshop moderation	Peter Daalsgard
Website management	Haotian Li, Srishti Palani
Social media	Anjali Singh
Post-workshop artifacts showcase curation	Gonzalo Ramos, Srishti Palani
Post-workshop synthesis coordination	Anuschka Schmitt, Advait Sarkar
Post-workshop gap-finding activity coordination	Kenneth Holstein

people collaborate with AI to communicate, co-create, make sense, and learn.

Jessica He is a UX Designer at IBM Research. Her work focusses on designing transparency mechanisms and co-creative workflows that encourage people to engage more critically with generative AI tools and outputs, spanning topics such as AI attribution, collaborative knowledge work, and risk mitigation.

Ken Holstein is an Assistant Professor at Carnegie Mellon University's Human-Computer Interaction Institute. His research explores how AI systems can effectively augment and bring out the best of human expertise in real-world design and decision-making. This research has been recognised with paper awards at top conferences in HCI and AI, including CHI. Ken has previously co-organised numerous workshops on related topics at ACM CHI, CSCW, and FAccT.

Advait Sarkar is a researcher at Microsoft, and lecturer at the University of Cambridge and University College London. He studies the effects of GenAI on knowledge work, productivity, and creativity. He leads a research agenda aimed at enhancing critical thinking with GenAI [15, 16] in Microsoft's *Tools for Thought* group [11]. His article *"AI Should Challenge, Not Obey"* was the cover story in the October 2024 issue of Communications of the ACM [14].

Gonzalo Ramos is an independent researcher working at the intersection of HCI, Design, and AI to protect and augment people's agencies and capabilities. He is a graduate of the University of Toronto's DGP lab, as well as the Universidad de Buenos Aires. He was a Principal Researcher at Microsoft Research in Redmond and part of the Tools for Thought team at Microsoft Research Cambridge. He was also Senior Design Technologist, and as a UX Scientist at Amazon. You can find more information on his work here: https://www.linkedin.com/in/gonzaloramos/.

Anuschka Schmitt is an Assistant Professor at the London School of Economics. Her research examines how AI-based systems augment human work, emphasising how we can balance productivity gains with interrelated work motives and long-term outcomes such as knowledge preservation. Her work uses experimental and trace data methods.

Anjali Singh is a Postdoctoral Fellow at the University of Texas at Austin School of Information. Her research develops and examines learning technologies that support users' critical thinking, agency, and metacognitive engagement. Methodologically, her work integrates empirical lab and field studies with design-based research to create personalised, scalable, and interactive systems and interventions for learning and information seeking.

Haotian Li is a researcher in Microsoft Research Asia (MSRA). He has received his Ph.D. and B.Eng. from HKUST. His current research interest is understanding and enhancing human-AI collaboration for creativity and productivity with techniques from HCI and AI. For more details, please refer to https://haotian-li.com/.

Srishti Palani is a Senior Researcher at Tableau Research, Palo Alto, California. Her research interests lie at the intersection of human-centered AI, human-computer interaction and cognitive science. She studies how people interpret, use and evaluate AI for complex cognitive tasks like data analysis, decision-making, and creativity. Based on these insights, she develops interactive intelligent tools to support these cognitive abilities.

Peter Dalsgaard is a Full Professor and director of the Centre for Digital Creativity at Aarhus University. He explores digital technologies from a human-centered perspective with a focus on how humans use technology to think and create in new ways. His research combines studies of real-life use of digital systems in creative processes, experiments with prototypes of new digital technologies, and development of theories to understand the role and nature of digital tools in creative processes.

7 Note About Past Workshops

This workshop is a follow-up to the CHI 2025 *Tools for Thought* workshop, which brought together 56 officially registered participants, with 34 accepted submissions and 72 submissions received, culminating in a workshop synthesis [20] and an HCI journal special issue on tools for thought [18]. Last year's workshop focussed on mapping the field, this edition moves toward developing operational frameworks, principles, and tools.

8 Call for Participation

How can we design and use generative AI as *tools for thought* (TfT), i.e., tools that protect and augment people's higher cognitive functions such as reasoning, memory, imagination, or critical thinking, while also helping them achieve their goals? After mapping the landscape at last year's workshop, this year, we aim to go from vision to implementation, from foundational to operational.

To that end, we invite researchers, designers, practitioners, and provocateurs to submit a paper (up to four pages, excluding references) covering at least one of the following themes: *(1) TfT design or usage strategies, (2) TfT outcomes and their measurement,* and *(3) TfT experience and adoption.* Besides the paper, authors will be required to fill out a Miro board template, which will then be thematically clustered by the organisers to form participant groups for the workshop. The group work can take diverse forms, including, but not limited to, working on new frameworks, design principles, measurement techniques, theory, or even the creation of novel TfT concept/mockups in a mini hackathon. During the last part of the workshop, groups will bring together their key ideas in a group poster on Miro and pitch their results to the other participants.

Papers and the sharing links to the Miro boards can be submitted via the workshop website (https://ai-tools-for-thought.github.io/workshop/). Accepted papers will be published on the website. At least one author of each accepted submission must, and at most two authors can participate in the workshop. We aim for a total of 45 participants for the 180 minute workshop.

References

[1] EY Americas. 2023. New EY research reveals the majority of US employees feel AI anxiety amid explosive adoption. https://www.ey.com/en_us/newsroom/2023/12/ey-research-shows-most-us-employees-feel-ai-anxiety Accessed October 1, 2025.

[2] Zahra Ashktorab, Michael Desmond, Qian Pan, James Johnson, Michelle Brachman, Casey Dugan, Marina Danilevsky, and Werner Geyer. 2025. Emerging Reliance Behaviors in Human-AI Content Grounded Data Generation: The Role of Cognitive Forcing Functions and Hallucinations. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work.* 1–17.

[3] Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences* 122, 26 (2025), e2422633122.

[4] Yizhou Fan, Luzhen Tang, Huixiao Le, Kejie Shen, Shufang Tan, Yueying Zhao, Yuan Shen, Xinyu Li, and Dragan Gašević. 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology* 56, 2 (2025), 489–530.

[5] Majeed Kazemitabaar, Oliver Huang, Sangho Suh, Austin Z Henley, and Tovi Grossman. 2025. Exploring the design space of cognitive engagement techniques with ai-generated code for enhanced learning. In *Proceedings of the 30th International Conference on Intelligent User Interfaces.* 695–714.

[6] Pia Kreijkes, Viktor Kewenig, Martina Kuvalja, Mina Lee, Sylvia Vitello, Jake M. Hofman, Abigail Sellen, Sean Rintel, Daniel G. Goldstein, David M. Rothschild, Lev Tankelevitch, and Tim Oates. 2025. Effects of LLM Use and Note-Taking On Reading Comprehension and Memory: A Randomised Experiment in Secondary Schools. https://doi.org/10.2139/ssrn.5095149 social science research network:5095149

[7] Hao-Ping 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. In *Proceedings of the 2025 CHI conference on human factors in computing systems.* 1–22.

[8] Matthias Lehmann, Philipp B. Cornelius, and Fabian J. Sting. 2024. AI Meets the Classroom: When Does ChatGPT Harm Learning? https://doi.org/10.2139/ssrn.4941259 social science research network:4941259

[9] Benjamin Lira, Todd Rogers, Daniel G. Goldstein, Lyle Ungar, and Angela L. Duckworth. 2025. Learning Not Cheating: AI Assistance Can Enhance Rather than Hinder Skill Development. https://doi.org/10.48550/arXiv.2502.02880 arXiv:2502.02880 [cs]

[10] Shakked Noy and Whitney Zhang. 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *Science* 381, 6654 (July 2023), 187–192. https://doi.org/10.1126/science.adh2586

[11] Microsoft Research. 2024. Tools for Thought. https://www.microsoft.com/en-us/research/project/t4t/. Accessed: 2024-09-26.

[12] Bryan Robinson. 2024. *Employees Report AI Increased Workload.* https://www.forbes.com/sites/bryanrobinson/2024/07/23/employees-report-ai-increased-workload/ Forbes article.

[13] Yvonne Rogers, Leon Reicherts, Tony Zhang, and Mariam Hassib. 2025. Augmenting Human Cognition Through Generative AI. In *Handbook of Human-AI Collaboration*, Mohamed Chetouani, Andrzej Nowak, and Paul Lukowicz (Eds.). Springer. https://doi.org/10.13140/RG.2.2.14568.12801

[14] Advait Sarkar. 2024. AI Should Challenge, Not Obey. *Commun. ACM* (Sept. 2024), 5 pages. https://doi.org/10.1145/3649404 Online First.

[15] Advait Sarkar. 2024. Intention Is All You Need. *Proceedings of the 35th Annual Conference of the Psychology of Programming Interest Group (PPIG 2024)* (2024).

[16] Advait Sarkar, Xiaotong (Tone) Xu, Neil Toronto, Ian Drosos, and Christian Poelitz. 2024. When Copilot Becomes Autopilot: Generative AI's Critical Risk to Knowledge Work and a Critical Solution. In *Proceedings of the Annual Conference of the European Spreadsheet Risks Interest Group (EuSpRIG 2024).*

[17] Anjali Singh, Zhitong Guan, and Soo Young Rieh. 2025. Enhancing Critical Thinking in Generative AI Search with Metacognitive Prompts. *arXiv preprint arXiv:2505.24014* (2025).

[18] Sean Rintel Srishti Palani, Gonzalo Ramos and Hariharan Subramonyam (Eds.). 2025. Generative AI in Tools for Thought: Theory, Design & Assessment. Taylor & Francis. https://think.taylorandfrancis.com/special_issues/generative-ai-in-tools-for-thought-theory-design-assessment/ Upcoming Special Issue, Human-Computer Interaction.

[19] Lev Tankelevitch, Elena L. Glassman, Jessica He, Majeed Kazemitabaar, Aniket Kittur, Mina Lee, Srishti Palani, Advait Sarkar, Gonzalo Ramos, Yvonne Rogers, and Hari Subramonyam. 2025. Tools for Thought: Research and Design for Understanding, Protecting, and Augmenting Human Cognition with Generative AI. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* ACM, Yokohama, Japan, 804:1–804:8. https://doi.org/10.1145/3706599.3706745

[20] Lev Tankelevitch, Elena L. Glassman, Jessica He, Aniket Kittur, Mina Lee, Srishti Palani, Advait Sarkar, Gonzalo Ramos, Yvonne Rogers, and Hari Subramonyam. 2025. Understanding, Protecting, and Augmenting Human Cognition with Generative AI: A Synthesis of the CHI 2025 Tools for Thought Workshop. https://doi.org/10.48550/arXiv.2508.21036 arXiv:2508.21036 [cs]

[21] Nađa Terzimehić, Babette Bühler, and Enkelejda Kasneci. 2025. Conversational AI as a Catalyst for Informal Learning: An Empirical Large-Scale Study on LLM Use in Everyday Learning. https://doi.org/10.48550/arXiv.2506.11789 arXiv:2506.11789 [cs]

[22] Nick von Felten. 2025. Beyond Isolation: Towards an Interactionist Perspective on Human Cognitive Bias and AI Bias. arXiv:2504.18759 [cs.HC] https://arxiv.org/abs/2504.18759

[23] Nick von Felten, Johannes Schöning, Klaus Opwis, and Nicolas Scharowksi. 2025. Biased Minds Meet Biased AI: How Class Imbalance Shapes Appropriate Reliance and Interacts with Human Base Rate Neglect. arXiv:2511.14591 [cs.HC] https://arxiv.org/abs/2511.14591

[24] Zelun Tony Zhang and Leon Reicherts. 2025. Augmenting Human Cognition With Generative AI: Lessons From AI-Assisted Decision-Making. https://doi.org/10.48550/arXiv.2504.03207 arXiv:2504.03207 [cs]

[25] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* ACM, Hamburg, Germany, 351:1–351:19. https://doi.org/10.1145/3544548.3581131