Microsoft

Microsoft New Future of Work Report 2025

A summary of recent research from Microsoft and around the world that can help us create a new and better future of work with AI.

Editors and Authors

Editors: Jenna Butler (Principal Applied Research Scientist), Sonia Jaffe (Principal Researcher), Rebecca Janßen (Senior Applied Scientist), Nancy Baym (Partner Research Manager), Brent Hecht (Partner Director of Applied Research), Jake Hofman (Senior Principal Researcher), Sean Rintel (Principal Research Sciences Manager), Bahar Sarrafzadeh (Principal Applied Research Scientist), Abigail Sellen (Distinguished Scientist), Mihaela Vorvoreanu (Principal Applied Scientist), Jaime Teevan (Chief Scientist and Technical Fellow).

Authors: Mohammed Alsobay, Liz Ankrah, Nancy Baym, Stephanie Beers, Megan Benzing, Mia Bruch, Zana Buçinca, Jenna Butler, Mar Carpanelli, Amelia Cole, Scott Counts, Madeleine Daepp, Justin Edwards, Alex Farach, Dan Goldstein, Mary L. Gray, Brent Hecht, Javier Hernandez, Jake Hofman, Eric Horvitz, Nicole Immorlica, Kori Inkpen, Shamsi Iqbal, Sonia Jaffe, Manasa Jagadeesh, Rebecca Janßen, Siân Lindley, Brendan Lucier, Nic Marquardt, Mercy Muchai, Ambrita Nand, Alexandra Olteanu, Jacki O'Neill, Max Peterschmidt, Christian Poelitz, Rabeeza, Nathalie Riche, Sean Rintel, Advait Sarkar, Bahar Sarrafzadeh, Sunayana Sitaram, Amanda Snellinger, Jina Suh, John Tang, Lev Tankelevitch, Jaime Teevan, Kiran Tomlinson, Anne Trapasso, Adam Troy, Gaurav Verma, Mihaela Vorvoreanu, Jack Williams, Jordana Young, Ben Zorn.

Referencing this report:

- On social media, please include the report URL (https://aka.ms/nfw2025).
- In academic publications, please cite as: Butler, J., Jaffe, S., Janßen, R., Baym, N., Hecht, B., Hofman, J., Rintel, S., Sarrafzadeh, B., Sellen, A., M., Vorvoreanu, Teevan., J. (Eds.). Microsoft New Future of Work Report 2025. Microsoft Research Tech Report MSR-TR-2025-58 *(https://aka.ms/nfw2025), 2025.*

Some of the information in this document relates to pre-released content which may be subsequently modified. Microsoft makes no warranties, express or implied, with respect to the information provided here. This document is provided "as-is". Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred. This document does not provide you with any legal rights to any intellectual property in any Microsoft product.

Welcome to the 2025 Microsoft New Future of Work Report!

As you sit down to read the 2025 New Future of Work report, it's worth pausing to consider the thread that ties the past five years of reports together. The inaugural New Future of Work report, published in 2021, focused on new ways people could work without relying on colocation as a key productivity tool. The second, in 2022, centered on the reintroduction of physical offices and the emergence of hybrid work. In 2023, we explored how large language models could reshape everyday work, and, in 2024, how those advances moved from promise to real-world impact.

Each year, as I've written this introduction, I've found myself saying that the previous year marked a once-in-a-lifetime generational shift. But after five years, it's clear that the reports aren't capturing a series of separate revolutions. Rather, they are chapters in a single story of the digital evolution of collaboration, each representing a phase that builds on, and is enabled by, what came before.

Last year's report highlighted research showing that AI delivers substantial gains in individual productivity. The next frontier, covered in this year's report, is collective productivity: how teams, organizations, and communities can get better together. AI can bridge gaps of time, distance, and scale, but only if built correctly. We must design AI to support shared goals, group context, and the norms of collaboration, and this requires not just new tools but new ways of working.

Microsoft's mission to empower every person and every organization on the planet to achieve more remains a stable north star as the terrain shifts. If the past five years taught us anything, it's that the future of work is not something that happens to us, it's something we create together, as a research community, as an industry, and as a public. As always, we invite you to join that effort, approaching it with curiosity, intentionality, and guided by evidence, so the next chapter of work is better for everyone.

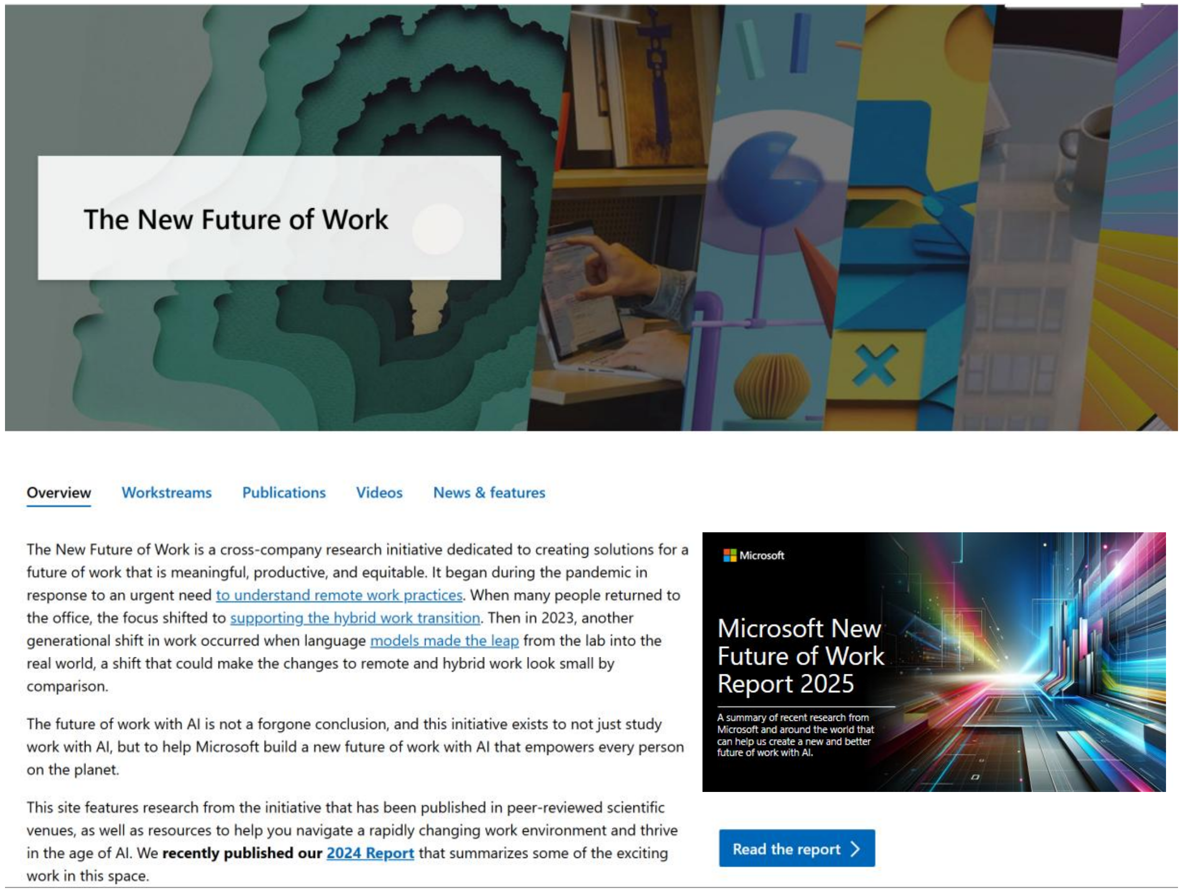
– Jaime Teevan, Chief Scientist and Technical Fellow

This report is a product of Microsoft's New Future of Work Initiative

Microsoft has shaped information work for over 50 years and the New Future of Work (NFW) initiative has helped it navigate the past five years of profound transformation. While the initiative was born out of the COVID-19 pandemic and subsequent shifts to remote and hybrid work, since 2023 it has focused on the incorporation and growing role of tools powered by large AI models.

Throughout this transformative period, the NFW initiative has brought together researchers from across Microsoft to conduct primary research and synthesize existing findings from the literature. The goal is not merely to predict what changes are coming, but to actively create a new future of work that is equitable, inclusive, meaningful, and productive.

The fifth edition of the annual NFW report represents another year of research and analysis, adding to the growing body of knowledge on AI and work. The evidence and insights presented here are not monolithic and represent contributing researchers' perspectives rather than Microsoft's corporate views. Additional research papers, practical guides, and white papers are available at: aka.ms/nfw.


The New Future of Work
Overview
Workstreams
Publications
Videos
News & features
Microsoft
Microsoft New Future of Work Report 2025
A summary of recent research from Microsoft and around the world that can help us create a new and better future of work with AI.
Read the report >

The New Future of Work is a cross-company research initiative dedicated to creating solutions for a future of work that is meaningful, productive, and equitable. It began during the pandemic in response to an urgent need to understand remote work practices. When many people returned to the office, the focus shifted to supporting the hybrid work transition. Then in 2023, another generational shift in work occurred when language models made the leap from the lab into the real world, a shift that could make the changes to remote and hybrid work look small by comparison.

The future of work with AI is not a forgone conclusion, and this initiative exists to not just study work with AI, but to help Microsoft build a new future of work with AI that empowers every person on the planet.

This site features research from the initiative that has been published in peer-reviewed scientific venues, as well as resources to help you navigate a rapidly changing work environment and thrive in the age of AI. We **recently published our 2024 Report** that summarizes some of the exciting work in this space.

aka.ms/nfw

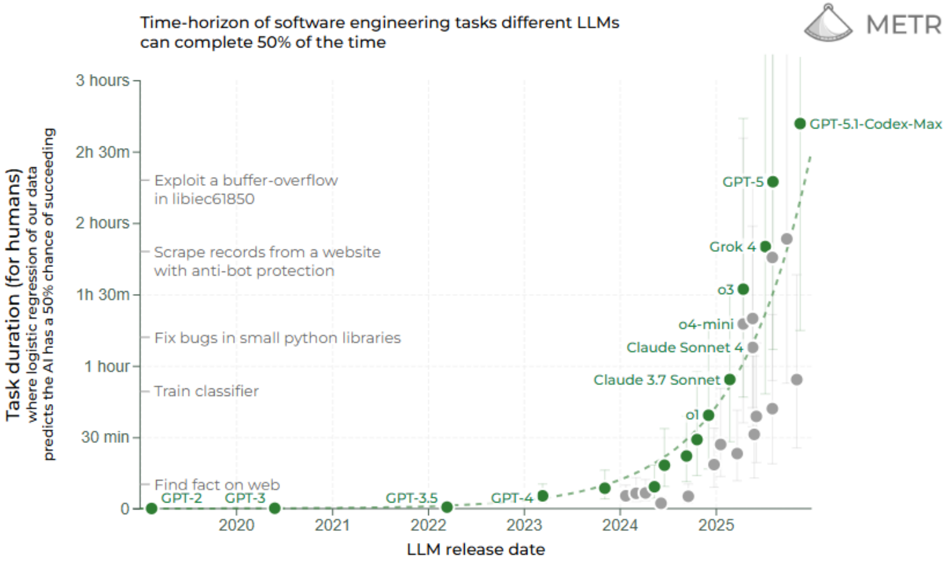
Report overview

This report provides research-backed insights into how AI is (or sometimes, should be) shaping work. Some of the questions it addresses include:

- **Adoption and Usage:** What changes are occurring in adoption and usage? What are the drivers and challenges? What are the gaps?
- **Impact on Work and Labor Markets:** How is AI impacting work and productivity? How are jobs evolving? Is generative AI affecting employment and wages? Where might agents reshape markets? What roles do automation and augmentation play?
- **Human-AI Collaboration:** How are the ways people interact with AI changing? How can human-AI collaboration be improved? How does AI use differ across modalities and time frames?
- **AI for Teamwork:** How can AI support teams as well as individuals? What role can AI play in team settings? What is needed to effectively integrate AI into group workflows?
- **Thinking, Learning, and Psychological Influences:** What are the effects of AI on cognition and thinking? Can AI be designed not just to create useful output, but to make the people who work with it smarter? How can AI serve as an effective classroom tool? Is it possible to measure psychological or well-being impacts from AI?
- **Specific Roles and Industries:** How is AI changing work for software engineers, program managers, researchers, and other professions?
- **External Voices:** What do leading scholars outside of Microsoft think are the most critical topics in this space?

Key context: AI capabilities continue to advance, especially due to reinforcement learning

- Long-horizon task completion capability is measurably accelerating. METR’s 50% task-completion time horizon shows frontier agents’ reliable task length has been rising exponentially with an ~7-month doubling time, turning “agent progress” into a concrete capability trend (Kwa et al., 2025).
- Verifiable reinforcement learning (RL) post-training (rewarding correct, checkable outcomes) enabled strong gains on hard math/coding style tasks even when starting from a base model with no labeled reasoning traces. (DeepSeek-AI et al., 2025).
- Scalable test-time compute frameworks gained traction, with open-weight models achieving IOI 2025 gold-level performance, showing repeatable “more compute → higher score” curves in competitive programming. (Samadi et al., 2025).



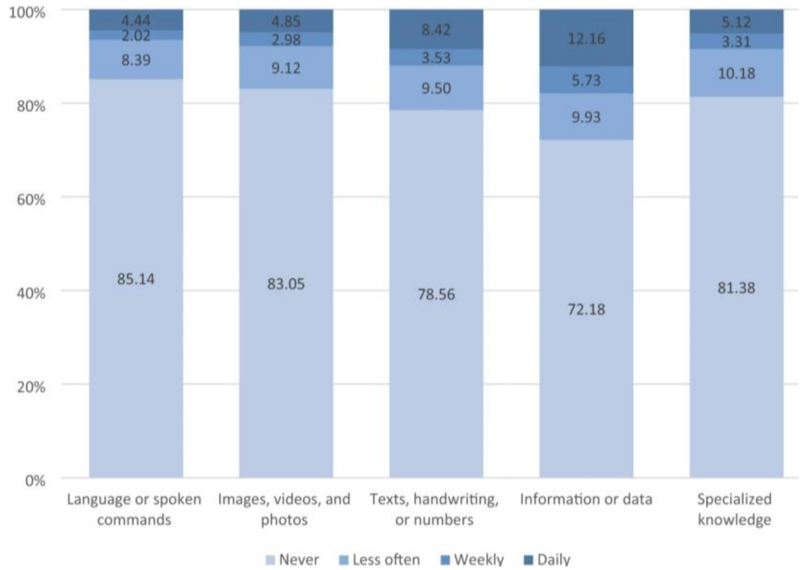
Task Completion Horizon from Kwa et al. (2025).

- Multi-turn RL for tool-using/search agents is now beating prompt-only baselines by learning from experience over longer action horizons; on a legal document search benchmark, an RL-trained 14B agent reports 85% vs 78% against frontier-class models, with additional gains when allowed more turns. (Kalyan & Andrews, 2025).
- Multi-model routing + aggregation is shifting from one-shot “pick a model” to an RL-trained sequential policy that can think, call multiple models, and integrate responses while explicitly optimizing performance–cost tradeoffs—improving results across multi-hop QA style evaluations. (Zhang et al., 2025).

Kwa, T. et al. (2025). Measuring AI Ability to Complete Long Tasks. *METR*.
DeepSeek-AI et al. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *arXiv Working Paper*.
Samadi, M. et al. (2025). Scaling Test-Time Compute to Achieve IOI Gold Medal with Open-Weight Models. *arXiv Working Paper*.
Kalyan, V. and Andrews, M. (2025). Reinforcement Learning for Long-Horizon Multi-Turn Search Agents. *arXiv Working Paper*.
Zhang, H. et al. (2025). Router-R1: Teaching LLMs Multi-Round Routing and Aggregation via Reinforcement Learning. *arXiv Working Paper*.

Investment in and adoption of generative AI continues to grow

- Effects of and gains from generative AI will be mediated by adoption (Jaffe et al., 2024).
- In 2024, Generative AI had $33.9 billion in global private investment—an 18.7% increase from 2023. Public investment has also grown (Maslej et al., 2025).
- AI use at work has been increasing, but heterogeneous:
  - Enterprise ChatGPT messages increased 8x in the past year (Chatterji et al., 2025).
  - A German survey found 38% of employed respondents used AI for work (Giering & Kirchner, 2025).
  - A survey of enterprise leaders found that usage and confidence were highest in IT & Procurement and lowest in Marketing/Sales and Operations; the Tech/Telecom, Professional Services, and Finance industries are leading (Korst et al., 2025).
  - A large 2024 survey of Americans found men were more likely than women (29.1% vs 23.5%) to use generative AI for work (Bick et al., 2024).
- On the consumer side, in June 2025, ChatGPT had over 700 million weekly active users globally (Chatterji et al., 2025).
  - The gender gap (based on first names) of consumer ChatGPT users is gone, a dramatic change since early 2023 when >80% were male (Chatterji et al., 2025).



Use of AI-based systems at work to recognize and process (weighted percentages) from Giering & Kirchner (2025).

Microsoft Study: Jaffe et al. (2024). Generative AI in Real-World Workplaces.
Maslej, N. et al. (2025). The AI Index 2025 Annual Report. *AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.*
Chatterji, R. et al. (2025). The state of enterprise AI. *OpenAI*
Giering, O. and Kirchner, S. (2025). Artificial intelligence and autonomy at work: empirical insights from Germany. *Journal for Labour Market Research.*
Korst, J. et al. (2025). Accountable Acceleration: Gen AI Fast-Tracks Into the Enterprise. *Wharton Human-AI Research and GBK Collective.*
Bick, A. et al. (2024). The Rapid Adoption of Generative AI. *NBER Working Paper.*

Organizational AI adoption depends on employees as much as leaders

- Across industries, the intention to use AI is influenced by social norms learned from leaders and peers (Kelly et al., 2023).
- Workers can be reluctant to adopt top-down mandated AI products that prioritize efficiency above quality and creativity, undermining the traditional view of humans as the core value driver of businesses. This reluctance limits the success of AI pilot programs (Young et al., 2025; Sharma, 2025; Murire, 2024).
- Leaders can facilitate AI adoption through clear communication supporting AI use, demonstrating their own learning, and setting realistic expectations about what AI can accomplish (Carter et al., 2024; Tursunbayeva & Chalutz-Ben Gal, 2024).
- AI products that integrate human thinking, creativity, and expertise while amplifying their value can promote adoption without raising concerns about replacement (Ali et al., 2025; Young et al., 2025; Sharma, 2025). For example, an AI assistant can act as a thought partner, helping users explore ideas and connect concepts across their knowledge base.
- Some of the best ways an organization might find to use AI "come from the edge, not the center" (Winsor, 2024). Organizations can facilitate AI adoption by creating systems and incentives for employees to share how they use AI with one another (Tursunbayeva & Chalutz-Ben Gal, 2024; Winsor, 2024).
- Employees are more likely to experiment with using AI and to share those insights with others when they feel safe and trust their organizations (Tursunbayeva & Chalutz-Ben Gal, 2024; Bankins et al., 2021).
- Many employees, particularly Gen X, will not adopt tools that make them conform to a way of working - they want products that are flexible enough to fit personal ways of working (Rozsa et al., 2023; Doblinger, 2023).

Kelly, S. et al. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*.
Microsoft Study: Young, J. et al. (2025). The Future of Enterprise. *Internal*.
Sharma, R. (2025). The Impact of AI-Generated Content on Human Creativity and Original Thought: A Psychological Study. *APA*.
Murire, O. T. (2024). Artificial Intelligence and Its Role in Shaping Organizational Work Practices and Culture. *MDPI*.
Carter, J. et al. (2024) To Succeed with AI, Adopt a Beginner's Mindset. *Harvard Business Review*.
Tursunbayeva, A. and Chalutz-Ben Gal, H. (2024). Adoption of artificial intelligence: A TOP framework-based checklist for digital leaders. *Business Horizons*.
Ali, D. et al. (2025). AI Adoption Across Mission-Driven Organizations. *arXiv Working Paper*.
Winsor, J. (2024). How to Be Systematic About Adopting AI at Your Company. *Harvard Business Review*.
Bankins, S. et al. (2021). A multilevel review of artificial intelligence in organizations. *Journal of Organizational Behavior*.
Rozsa, Z. et al. (2023). Job Crafting and Sustainable Work Performance: A Systematic Literature Review. *Equilibrium. Quarterly Journal of Economics and Economic Policy*.
Doblinger, M. (2023). Autonomy and Engagement in Self-Managing Organizations: Exploring the Relations with Job Crafting, Error Orientation and Person–Environment Fit. *Front. Psychol*.

Centering worker voice in AI design boosts productivity, satisfaction, and skill growth—driving both business success and worker flourishing

- Worker involvement in technology design promotes sustainable productivity and job satisfaction. Historical and contemporary research consistently shows that when workers' expertise and perspectives inform the design and deployment of workplace technologies, organizations achieve more sustainable improvements in productivity and well-being (Trist & Bamforth, 1951; Roethlisberger & Dickson, 1939; Hackman & Oldham, 1976).
- Ethnographic and HCI research demonstrates that workers adapt technology in creative ways, and that participatory design—where workers are co-designers—results in tools that better fit real workflows and foster higher adoption (Suchman, 1987; Orr, 1996; Awumey et al., 2024; Ehn, 1993; Doellgast et al., 2025).
- Combining technical and social science research methods can create AI systems that improve worker skills and satisfaction—not just accuracy—by embedding human-centric metrics, workers' values, and skill-building into their design (Bucinca, 2025).
- Data-driven workplace monitoring (telemetry) has mixed effects and should be governed with worker input for best outcomes. While monitoring and algorithmic management can boost short-term output, they often increase stress and erode trust unless workers help define what is measured and how data is used (Pentland, 2012; Wells et al., 2007; Ajunwa, 2023).

Trist, E. and Bamforth, K. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting. *Human Relations*.
Roethlisberger, F. and Dickson, W. (1939). Management and the worker. *Harvard University Press*.
Hackman, J. and Oldham, G. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance*.
Suchman, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. *Cambridge University Press*.
Orr, J. (1996). Talking About Machines: An Ethnography of a Modern Job. *Cornell University Press*.
Awumey, E. et al. (2024). A Systematic Review of Biometric Monitoring in the Workplace: Analyzing Sociotechnical Harms in Development, Deployment and Use. *FAccT*.
Ehn, P. (1993). Scandinavian Design: On Participation and Skill.
Doellgast, V. et al. (2025). Global Case Studies of Social Dialogue on AI and algorithmic management. *International Labour Organization*.
Bucinca, Z. (2025). Worker-Centric AI for Decision-Support. *Harvard University*.
Pentland, A. (2012). The New Science of Building Great Teams. *Harvard Business Review*.
Wells, J. et al. (2007). The impact of the perceived purpose of electronic performance monitoring on an array of attitudinal variables. *Human Resource Development Quarterly*.
Ajunwa, I. (2023). The Quantified Worker. *Cambridge University Press*.

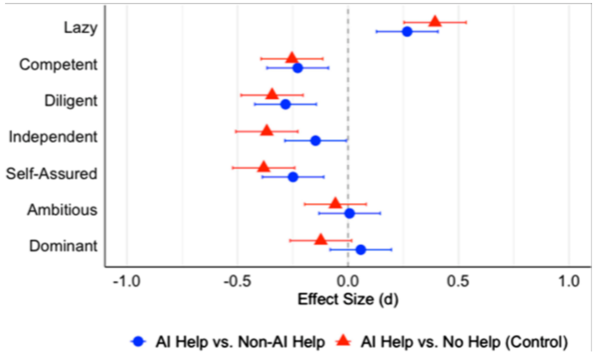
CEOs expect AI to transform their businesses, but leading organizational AI adoption can be challenging

- A 2025 IBM survey of 2000 CEOs in 33 countries and 24 industries found that most CEOs expect AI to transform their businesses. Other industry research shows leaders believe having the most advanced generative AI is crucial to remaining competitive (de Bellefonds et al., 2024; IBM Institute for Business Value, 2025).
- However, organizational leaders have difficulties developing top-down AI strategies for many reasons, including the rapid diffusion of AI technologies, the speed with which they change, the need to communicate and reach alignment about AI, the need to prioritize AI against other concerns, and the challenge of reimagining workflows and processes (de Bellefonds et al., 2024; Leonardi, 2023).
- A comparative case study of AI adoption in the Dutch public sector found adoption barriers included organizational inflexibility and risk intolerance, as well as structural separation between exploration and exploitation. For example, a data science team might be put in charge of exploring AI without operational alignment or frontline and leadership support to exploit and scale their ideas (Selten & Klievink, 2024).
- Chen & Tajdini (2024) surveyed managers involved in Marketing and AI at US-based firms involved in high-tech consumer goods, industrial equipment, and financial services. They found that organizational intensity of AI adoption was driven by top management support, customer orientation, and emerging industry social norms toward adopting AI (Chen & Tajdini, 2024).
- Additionally, organizations are best able to adopt AI when they are innovative, experimental and learning-oriented, supportive, and collaborative (de Bellefonds et al., 2024; Tursunbayeva & Chalutz-Ben Gal, 2024, Sternfels & Atsmon, 2025).

de Bellefonds, N. et al. (2024). Where's the Value in AI? *Boston Consulting Group*.
IBM Institute for Business Value (2025). 5 mindshifts to supercharge business growth. *IBM*.
Leonardi, P. (2023). Helping Employees Succeed with Generative AI. *Harvard Business Review*.
Selten, F. and Klievink, B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*.
Chen, J. and Tajdini, S. (2024). A moderated model of artificial intelligence adoption in firms and its effects on their performance. *Information Technology Management*.
Tursunbayeva, A. and Chalutz-Ben Gal, H. (2024). Adoption of artificial intelligence: A TOP framework-based checklist for digital leaders. *Business Horizons*.
Sternfels, B. and Atsmon, Y. (2025). The learning organization: How to accelerate AI adoption. *McKinsey*.

Social norms shape how people interpret others’ (possible) AI use, often with negative consequences

- Research has shown that suspected use of AI can result in moral evaluations. People who use AI assistance expect to be, and often are, evaluated as “lazier”, less competent, less diligent, less trustworthy, and less moral (Reif et al., 2025; Schilke & Reimann, 2025; Zhou et al., 2025). These effects are lessened when evaluators, including managers, have experience with and positive attitudes towards AI (Reif et al., 2025; Schilke & Reimann, 2025).
- Being seen as using AI can harm relationships with colleagues, perhaps because AI-using employees may be seen as “slacking” rather than putting in personal effort (Zhou et al., 2025). Paradoxically, the act of disclosing the use of AI can erode trust (Schilke & Reimann, 2025).
- The extent to which these “perceptual harms” (Kadoma et al., 2025) vary across social identities is unclear:
  - Reif et al. (2025) found no gender differences in experiments across occupational categories.
  - Kadoma et al. (2025) found that male (and East Asian) freelancers were more likely to be suspected of using AI and hence evaluated negatively.
  - Software engineers who used AI received lower competency ratings on identical work, and this effect was doubled for females, with women receiving a competency score 13% less for identical code (vs 6% for men) (Gai et al., 2025).

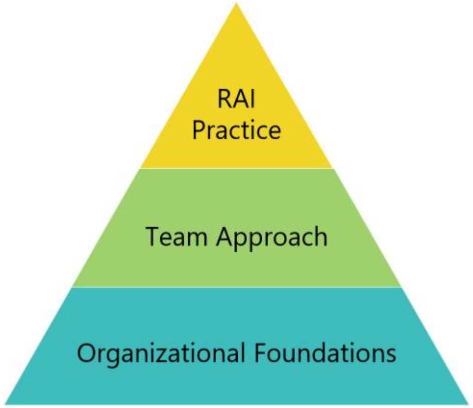


Differences in evaluations for AI help vs. Non-AI help and AI help vs. Control. Positive d values indicate higher values in the AI Help condition while negative d values indicate lower values in the AI Help condition. Error bars represent 95% CI. From Reif et al. (2025)

Reif, J. et al. (2025). Evidence of a social evaluation penalty for using AI. *PNAS*.
Schilke, O. and Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*.
Zhou, X. et al. (2024). How Do Coworkers Interpret Employee AI Usage: Coworkers' Perceived Morality and Helping as Responses to Employee AI Usage. *Human Resource Management*.
Kadoma, K. et al. (2025). Generative AI and Perceptual Harms: Who’s Suspected of using LLMs? *CHI*.
Gai, P. et al. (2025). Competence Penalty Is a Barrier to the Adoption of New Technology. *SSRN Working Paper*.

The Responsible AI Organizational Maturity Model provides a map for organizational change

- AI adoption and capability building require organizational transformation (Kemp, 2024). Maturity models are useful tools for guiding organizational transformation. Existing maturity models focus on building AI capability (Alsheiabni et al., 2019; Vaish et al., 2021) and AI adoption (Hansen et al., 2024).
- The Responsible AI Organizational Maturity Model (RAI-OMM) provides a map for organizations to advance their responsible AI strategy and practice (Heger et al., 2025). The RAI-OMM is forward-looking and best used for planning, not evaluation.
- Based on interviews and co-design sessions with 90 RAI experts and practitioners, the RAI-OMM identifies 24 aspects (dimensions) that need to be considered for mature RAI practice and describes, for each dimension, five distinct maturity levels.



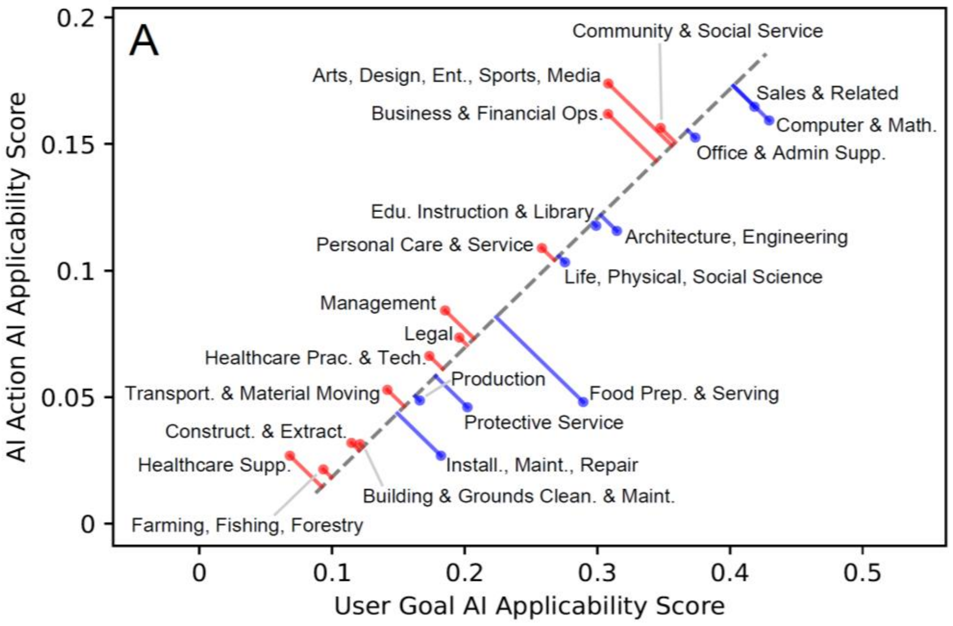
The 24 dimensions in the RAI-OMM fall into three main categories (Heger et al., 2025).

- RAI maturity requires leadership investment, aligned organizational practices, and holistic change management strategies that address both technological and human dimensions (Duran, 2025; Shekshnia, 2025; Wang et al., 2025; Herrmann & Pfeiffer, 2023; Yunusa, 2025).
- RAI maturity dimensions are interdependent and fall into three categories: Organizational Foundations require leadership commitment and investment in organization-wide infrastructure; Team Approach dimensions highlight the necessity of cross-discipline collaboration; finally, these make possible mature RAI Practice, which is characterized by deep integration into AI development & deployment processes.

Kemp, A. (2024). Competitive advantage through artificial intelligence: Toward a theory of situated AI. *Academy of Management Review.*
Alsheiabni, S. et al. (2019). Towards an Artificial Intelligence maturity model: From science fiction to business facts. *PACIS.*
Vaish, R. et al. (2021). AI maturity framework for enterprise applications. *IBM Technical Report.*
Hansen, H. et al. (2024). Understanding Artificial Intelligence Diffusion through an AI capability maturity model. *Information System Frontiers*.
Microsoft Study: Heger, A. et al. (2025). *Towards a Responsible AI Organizational Maturity Model*. *CSCW.*
Microsoft Study: Duran, J. et al. (2025). *RAI advocacy: Communicative strategies for advancing responsible AI in large technology companies*. *AIES*.
Shekshnia, S. (2025). AI strategy, leadership, talent and workforce, and transformation. In: AI Leadership for Corporate Boards. *Springer*.
Wang, A. et al. (2024). *Strategies for increasing corporate responsible AI prioritization*. *AIES*.
Herrmann, T. and Pfeiffer, S. (2023). Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. AI & Society.
Yunusa, E. (2025). Creating an artificial intelligence-ready organizational culture: harmonizing human existence with AI strategic decision-making. *International Journal of Business Sustainability.*

Analyses of LLM logs are showing for which activities the tools are used, and which occupations do those activities

- Analysis of ChatGPT found growth in work-related messages but even faster growth in non-work-related messages.
  - “Practical Guidance,” “Seeking Information,” and “Writing” are the three most common topics, accounting for ~80% of all use messages (Chatterji et al., 2025).
- Researchers at Anthropic found that 37% of Claude usage was for tasks associated with Computer and Mathematical occupations (Handa et al., 2025).
- An analysis of Microsoft Bing Copilot logs looked separately, for each conversation, at the work activities associated with users' goals and the AI actions (Tomlinson et al., 2025).
  - The most common user goals involved learning and communicating; AI actions were largely communicating and explaining; both sides had lots of writing activities.
  - Aggregating activities to occupations, most occupations have some AI-applicable tasks, with information workers including sales, computer occupations, media, and administrative occupations at the top.

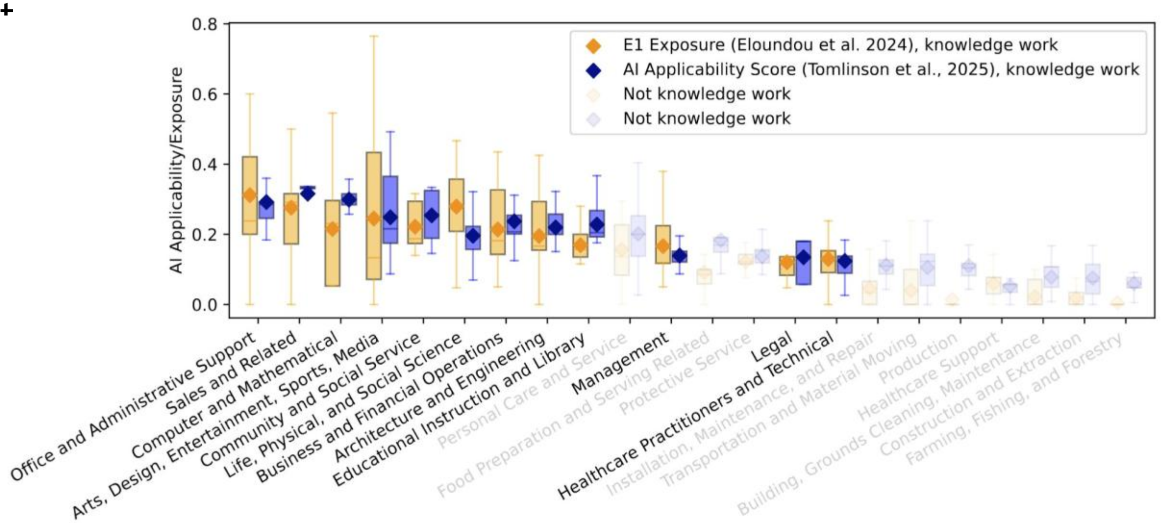


Graph shows average AI applicability for the user goal (x-axis) and the AI action (y-axis) (Tomlinson et al., 2025).

Chatterji, A. et al. (2025). How People Use ChatGPT. *NBER Working Paper.*
Handa K. et al. (2025). Which economic tasks are performed with AI? evidence from millions of Claude conversations. *arXiv Working Paper.*
Microsoft Study: Tomlinson, K. et al (2025). Working with AI: Measuring the Applicability of Generative AI to Occupations. *arXiv Working Paper.*

AI could broaden access to high-value knowledge work outputs and opportunities, if societal and technical barriers can be overcome

- AI can increase people's efficiency in multiple domains of knowledge work (Dillon et al., 2025; Brynjolfsson et al., 2025; Cui et al., 2024).
- These capabilities could broaden the set of people who have access to knowledge work skills and outputs (Autor, 2024); this is bolstered by evidence that AI can shrink skill gaps in knowledge work (Brynjolfsson et al., 2025; Cui et al., 2024). Autor argues this could allow more people to engage in high-barrier-to-entry work, like medical or legal decision-making.
- For AI to be democratizing in this way, technical improvements are needed to address the "rich-get-richer" effect in open-ended, creative tasks and policy innovation will be needed to mitigate emerging divides in AI investment, access, and benefits (Daepp et al., 2025; Microsoft AI Economy Institute, 2025).

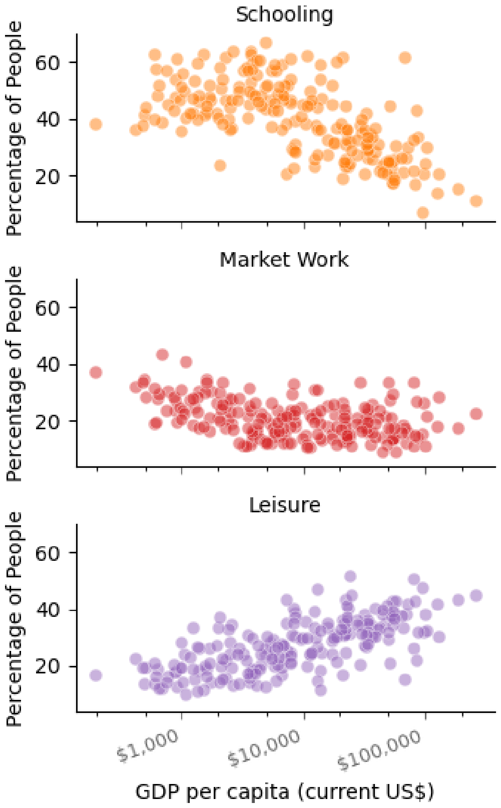


Estimates of how applicable AI is to different occupational categories, from Eloundou et al. (2024) and Tomlinson et al. (2025). Knowledge work occupations are highlighted. Boxplots are employment-weighted and diamonds show means. Both studies show AI to be most useful for knowledge work occupational tasks (Daepp et al., 2025).

Microsoft Study: Dillon, E. et al. (2025). Early impact of M365 Copilot. *arXiv Working Paper*.
Brynjolfsson, E. et al. (2025). Generative AI at work. *The Quarterly Journal of Economics*.
Microsoft Study: Cui, Y. et al. (2024). The effects of generative Ai on high-skilled work: Evidence from three field experiments with software developers. *SSRN Working Paper*.
Autor, D. (2024). AI could actually help rebuild the middle class. *Noema*.
Microsoft Study: Daepp, M. et al. (2025). AI and the democratization of knowledge work. *Under review*.
Microsoft Study: Microsoft AI Economy Institute (2025). AI Diffusion Report: Where AI is most used, developed, and built. *Microsoft*.
Eloundou, T. et al. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*.
Microsoft Study: Tomlinson, K. et al (2025). Working with AI: Measuring the Applicability of Generative AI to Occupations. *arXiv Working Paper*.

AI use in low-income countries is growing, particularly for schooling

- Though AI usage remains highest in high-income and technologically advanced countries (Microsoft AI Economy Institute, 2025; Appel et al., 2025), the last year has seen dramatic usage growth in low- to middle-income countries, narrowing the gap (Chatterji et al., 2025).
- In one survey, people in Asia and Latin America were more likely to agree that "products and services using AI have more benefits than drawbacks" (e.g., China: 83%, Mexico: 70%) while agreement is lower in Europe and the Anglosphere (e.g., US: 39%, Netherlands: 36%) (Maslej et al., 2025).
- Adoption is highest in countries that have invested in digital infrastructure and education, and where the predominant language is one well-served by existing models (Microsoft AI Economy Institute, 2025).
    - When local languages are not well-served by AI, people sometimes use English instead: chats are conducted in English at disproportionate rates relative to the fraction of the population that speaks English in African and Asian countries, but not in Europe or the Americas (Slaughter & Daepp, in prep.).
- How AI is used also differs across place. In a study of early chatbot adopters, LLM usage for school increases with GDP per capita, while use for leisure decreases (see chart; Slaughter & Daepp, in prep.). This may be partially due to differences in the share of school-age population or in the amount of leisure time.

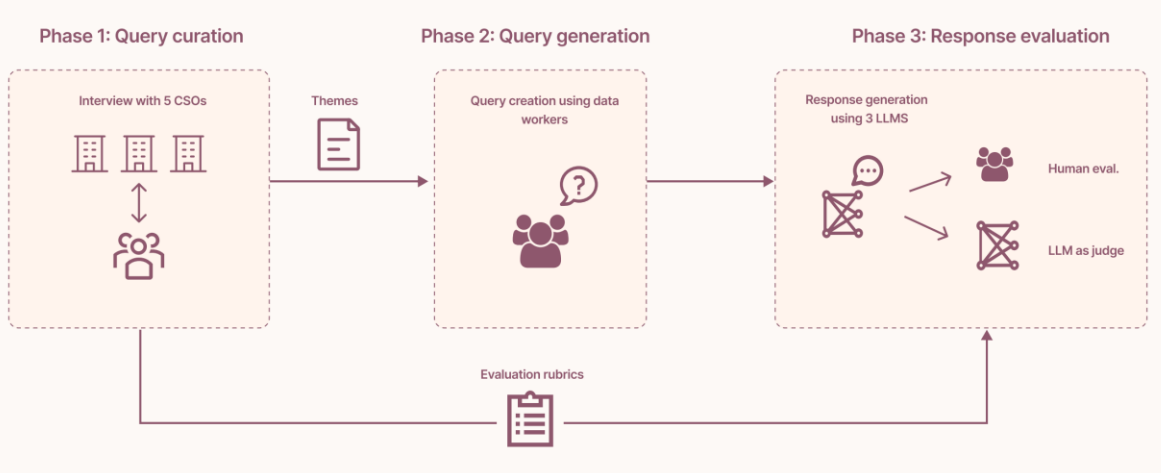


Patterns in Usage Domain among Early Adopters of Bing Copilot, the percentage of users in a country who predominantly had conversations in the specified domain (schooling, market work, or leisure) by GDP per capita (Slaughter & Daepp, in prep).

Microsoft Study: Microsoft AI Economy Institute (2025). AI Diffusion Report: Where AI is most used, developed, and built. *Microsoft*.
Appel, R. et al. (2025). The Anthropic Economic Index report: Uneven geographic and enterprise AI adoption. *Anthropic*.
Chatterji, A. et al. (2025). How people use ChatGPT. *NBER Working Paper*.
Maslej, N. et al. (2025). The AI Index 2025 Annual Report. *Institute for Human-Centered AI, Stanford University*.
Microsoft Study: Slaughter, I. and Daepp, M. (in prep.). Second-level digital divides in usage of a generative AI chatbot.

Synthetic data and representative evaluations are advancing LLMs for diverse languages and cultures

- Thoughtfully constructed, culturally-grounded synthetic data has shown promise in reducing the data gap between high-resource and under-resourced languages (Guduru et al., 2025). The Updesh project (Chitale et al., 2025) provides a framework for culturally-grounded multilingual synthetic data and a case study on Indian languages shows improved performance of models fine-tuned on this data.
- Coverage, representativeness, scalability, and trust continue to pose significant challenges in multilingual and multicultural evaluation. The Samiksha project (Hamna et al., 2025) addresses these issues by developing benchmarks from the ground up by incorporating inputs from a wide range of stakeholders and community members.
  - Samiksha aims to establish a realistic and representative benchmark for Indian languages that reflects key use cases.
  - Insights from such assessments can contribute to building models that perform effectively in various languages and cultural contexts.

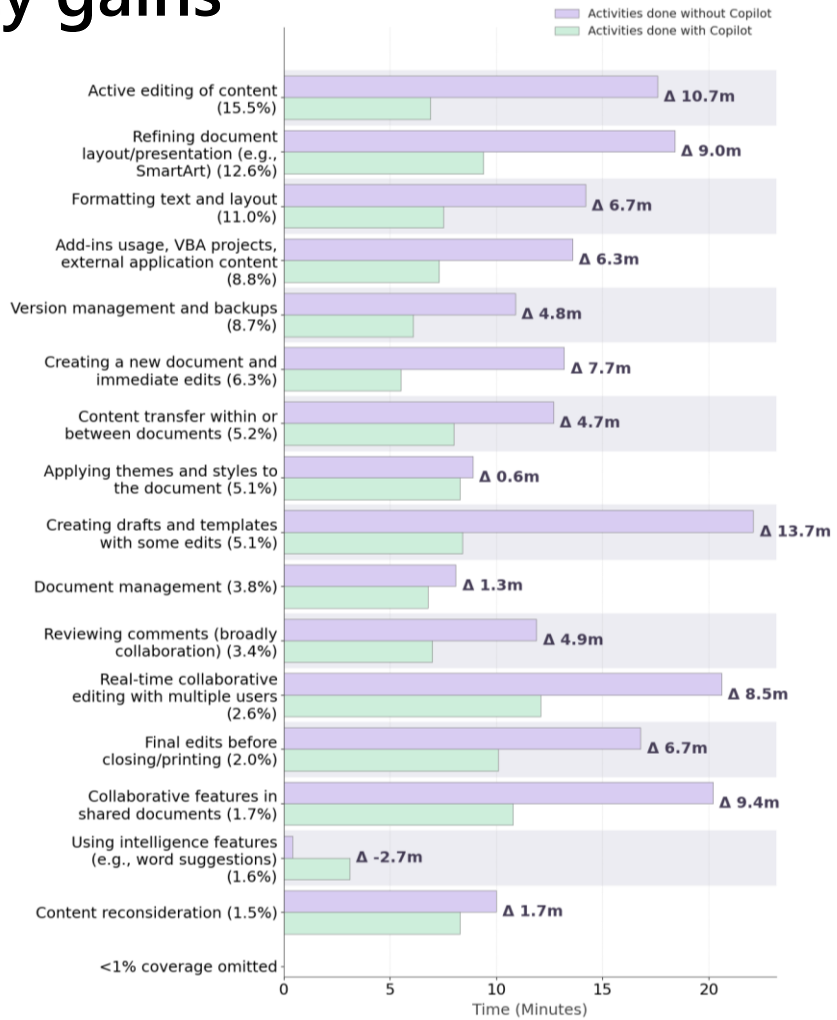

Phase 1: Query curation
Phase 2: Query generation
Phase 3: Response evaluation
Interview with 5 CSOs
Themes
Query creation using data workers
Response generation using 3 LLMS
Human eval.
LLM as judge
Evaluation rubrics

The Samiksha benchmark creation and evaluation pipeline. Feedback on what to evaluate is provided by Civil Society Organizations (CSOs), data workers expand themes into data points in their own languages, adding cultural context when appropriate; responses from LLMs are evaluated using human evaluation and LLM-judges (Hamna et al., 2025).

Guduru, M. et al. (2025). BhashaKritika: Building Synthetic Pretraining Data at Scale for Indic Languages. arXiv Working Paper.
Microsoft Study: Chitale, P. et al. (2025). The role of synthetic data in Multilingual, Multi-cultural AI systems: Lessons from Indic Languages. *arXiv Working Paper*.
Microsoft Study: Hamna et al. (2025). Building Benchmarks from the Ground Up: Community-Centered Evaluation of LLMs in Healthcare Chatbot Settings. *arXiv Working Paper.*

AI use is associated with time savings and productivity gains

- Surveyed ChatGPT Enterprise users attribute 40–60 minutes saved per day from use of AI (Chatterji et al., 2025). The savings are heterogeneous. LLM-based estimates of time savings from Claude usage suggested they varied by occupation and task—for example, 80–85% for legal and management tasks, but only 20% for checking diagnostic images (Tamkin & McCrory, 2025).
- OpenAI designed 1,320 tasks to mimic the work product of predominantly digital occupations in high-value sectors. Frontier LLMs approached quality parity with human experts: the win+tie shares for the top model ranged from 33-56% across industries and tie rates were low (Patwardhan et al., 2025).
- To see how time spent on tasks varied with Copilot usage, researchers developed *WorkflowView,* which uses an LLM to categorize telemetry action sequences into high-level workflow activities (Verma & Counts, 2025).
    - Privacy-preserving analysis of one month of telemetry of 50k Copilot-enabled Word users shows an average difference of 7 minutes per accepted Copilot output.
    - Copilot use is associated with a difference of 10.7 minutes in editing of content and 0.6 minutes in applying themes and styles. These variations can guide more effective integration of AI tools in productivity workflows.



Average time spent by *WorkflowView* Activity and Copilot usage (Verma & Counts, 2025)

Chatterji, R. et al. (2025). The state of enterprise AI. *OpenAI*
Tamkin, A. and McCrory, P. (2025). Estimating productivity gains from Claude conversations. *Anthropic*.
Patwardhan, T. et al. (2025). GDPval: Evaluating AI model performance on real world economically valuable tasks. *arXiv Working Paper.*
Microsoft Study: Verma, G. and Counts, S. (2025). WorkflowView: Abstracting Activity Logs with LLMs for Interpretable and Actionable Insights. *Under review.*

The rise of AI “workslop” has productivity risks and market effects

- AI “workslop” refers to AI-generated work content that appears useful but lacks substance, is incomplete, or contains inaccuracies. Such content undermines productivity by forcing recipients to interpret, correct, or redo the work (Niederhoffer et al., 2025; Madsen & Puyt, 2025). Workslop may be a key reason why individual productivity gains are not seen at the group or organizational level.
- In Niederhoffer et al.’s (2025) survey of 1,150 U.S. employees, 40% received workslop in the past month, estimated at 15% of content. Most slop flows between peers (40%), but it also moves upward (18%) and downward (16%) in hierarchies.
- Workslop is part of the broader generative AI “slop” phenomenon, which is reshaping markets by flooding them with low-cost and low-quality content (Miklian & Hoelscher, 2025; Tullis, 2025; Pendergrass et al., 2025).
- Technical solutions are still nascent. One approach to metrics focuses on judging information utility, information quality, and style quality (Shaib et al., 2025), but this would need to be combined with accuracy checks (e.g. MAD-Fact by Ning et al., 2025), ideally with access to internal data or document repositories.
- Employee training about awareness of AI limitations and critical evaluation skills can reduce workslop by helping people identify and correct low-value outputs before they enter workflows (Park et al., 2025; Simkute et al., 2024).

Issue	Definition	Consequence
Volume	Scale and abundance of generative output	Crowding out human creativity and visibility
Velocity	Speed of production and circulation	Outpacing fact-checking and moderation
Variety	Range of forms, genres, and modalities	Expansion into all cultural and knowledge domains
Value	Erosion of cultural and epistemic worth	Devaluation of originality and meaning
Verification	Problems of truth, trust, and reliability	Epistemic pollution and misinformation
Visibility	Algorithmic amplification and ranking	Rewarding slop via platform incentives
Virality	Meme-like diffusion and contagious spread	Rapid uptake and normalization of slop

The 7Vs of AI Slop (Adapted from Madsen & Puyt, 2025).

Niederhoffer, K. et al. (2025). AI-Generated “Workslop” Is Destroying Productivity. *Harvard Business Review*.
Madsen, D. Ø. and Puyt, R.W. (2025). The 7Vs of AI Slop: A Typology of Generative Waste. *SSRN Working Paper*.
Miklian, J. and Hoelscher, K. (2025). A New Digital Divide? Coder Worldviews, the Slop Economy, and Democracy in the Age of AI. *arXiv Working Paper*.
Tullis, J. (2025). Sifting Through the Slop: How Generative AI Created a Market for Lemons for Text-Based Works. *SSRN Working Paper*.
Pendergrass, W. et al. (2025). A strategic cycle of slop: Understanding the commodification of AI feculence and its place in the attention economy. *Issues in Information Systems*.
Shaib, C. et al. (2025). Measuring AI "Slop" in Text. *arXiv Working Paper*.
Ning, Y. et al. (2025). MAD-Fact: A Multi-Agent Debate Framework for Long-Form Factuality Evaluation in LLMs . *arXiv Working Paper*.
Park, J., et al. (2025). Attitudes towards artificial intelligence at work: Scale development and validation. *Journal of Occupational and Organizational Psychology*.
Microsoft Study: Simkute, A. et al. (2024). Ironies of Generative AI: Understanding and Mitigating Productivity Loss in Human-AI Interaction. *IJHCI*.

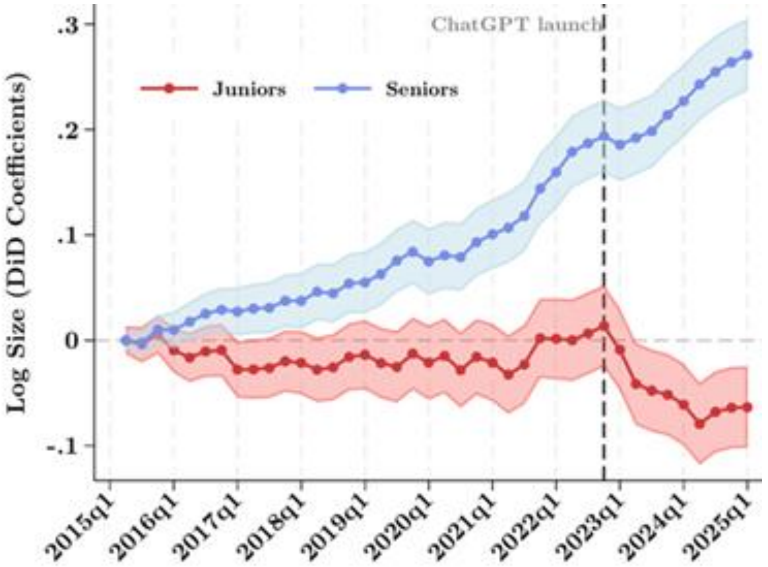
Labor market outcomes cannot be predicted based on existing limited evidence on automation vs augmentation

- Researchers have tried to categorize which tasks are more likely to be augmented or automated by generative AI, either in theory using an LLM classifier (Eloundou et al., 2024) or by analyzing what users are trying to achieve in conversations with LLMs as distinct from what activities the LLM is performing (Tomlinson et al., 2025).
- However, the technical question of what activities can be delegated to the tools versus which activities it can assist with cannot answer the labor market question of which occupations will see increases or decreases in employment or wages.
- Employment and wages depend on how occupations are restructured and whether there is market demand for increased output.
  - When ATMs were invented, tellers' jobs were refactored to focus on other tasks and employment increased.
  - Theory suggests that wages are most likely to increase when the tasks that are automated are the ones requiring less expertise than other activities that are part of that occupation (Autor & Thompson, 2025).
- Researchers have also elicited workers' preferences about which tasks they would prefer to have automated (Shao et al., 2025).
  - Work preferences may predict the tasks for which AI will be adopted. They may also mediate the effect on wages if workers need to be paid more to use AI for some tasks than for others.
  - The top reasons given for wanting automation were freeing up time for high-value work, tasks being repetitive or tedious, and improving the quality of work.
- Evidence on new tasks – that were not being done prior to AI – requires more research (and more time for them to emerge).

Eloundou, T. et al. (2024). GPTs are GPTs: Labor market impact potential of LLMs . *Science*
Microsoft Study: Tomlinson, K. et al. (2025). Working with AI: Measuring the Applicability of Generative AI to Occupations. *arXiv Working Paper.*
Autor, D. and Thompson, N. (2025). Expertise. *NBER Working paper.*
Shao, Y et al. (2025). Future of Work with AI Agents: Auditing Automation and Augmentation Potential across the U.S. Workforce. *arXiv Working Paper.*

Aggregate effects remain small, but early-career employment impacts from AI are emerging

- Large-scale studies in Denmark and the US find no significant effect of AI on unemployment (Chen et al., 2025), working hours (Humlum & Vestergaard, 2025), or job openings (Hartley et al., 2025). However, hiring of AI talent has increased by more than 300% over the past eight years (LinkedIn, 2025).
  - For earnings, results vary from slightly increased earnings (Hartley et al., 2025), especially for occupations with high AI exposure (Chen et al., 2025), to no significant effects (e.g., Humlum & Vestergaard, 2025) to reduced salaries in high-wage occupations (Klein Teeselink, 2025).
- There is evidence of negative effects for younger workers, whose roles rely less on tacit experience, making them more vulnerable to automation and less shielded by firm-specific skills. Declines for younger workers might be offset by growth among older workers and less-exposed occupations.
  - Payroll data suggests employment for workers aged 22–25 in highly AI-exposed jobs fell by about 13% compared to less-exposed roles. The authors tested for firm-level shocks, remote work, and sector-specific effects but note they cannot fully rule out other trends influencing exposed occupations (Brynjolfsson et al., 2025).
  - Resume and job posting evidence shows hiring for junior/entry-level roles slows in exposed occupations after firms adopt AI (Hosseini & Lichtinger, 2025; Klein Teeselink, 2025).



Difference between employment at firms adopting and not adopting generative AI, separately for junior and senior employee (Hosseini & Lichtinger, 2025).

Chen, D. et al. (2025). The (Short-Term) Effects of Large Language Models on Unemployment and Earnings. arXiv Working Paper.
Humlum, A. and Vestergaard, E. (2025). Large Language Models, Small Labor Market Effects. *NBER Working Paper.*
Hartley, J. et al. (2025). The Labor Market Effects of Generative Artificial Intelligence. *SSRN Working Paper.*
LinkedIn. (2025). Work Change Report. AI is Coming to Work.
Klein Teeselink, B. (2025). Generative AI and Labor Market Outcomes: Evidence from the United Kingdom. *SSRN Working Paper.*
Brynjolfsson, E. et al. (2025). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. *Working Paper.*
Hosseini, S. M. and Lichtinger, G. (2025). Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data. *SSRN Working Paper.*

AI adoption is reshaping career paths and skill requirements within occupations

- Recent evidence suggests AI adoption affects career decisions and occupational mobility. Workers using AI chatbots are more likely to switch occupations (Humlum & Vestergaard, 2025), and search intensity for apprenticeships in cognitive and language-intensive fields has declined after chatbot introduction, signaling shifts in career preferences (Goller et al., 2025).
- Worker-level evidence from Germany shows AI exposure changes the mix of activities and required skills inside occupations. Unlike robots, AI reduces non-routine abstract tasks and increases demand for high-level routine tasks like oversight and evaluation (Engberg et al., 2025; Gathmann et al., 2024). AI adoption increases complexity in augmentation-prone roles while reducing skill requirements in automation-prone roles (Chen et al., 2024).
- Roles requiring AI skills are nearly twice as likely to also request analytical thinking, resilience, ethics, or digital literacy. A doubling of AI-specific job postings is associated with roughly 5% higher demand for these complementary skills, while demand for easily substitutable tasks such as basic data skills or translation declines slightly (Mäkelä & Stephany, 2025). Job postings requiring AI skills are growing over 70% year-over-year, extending beyond technical roles (LinkedIn, 2025a; 2025b).
- Workers exposed to AI gain most from retraining focused on broad skills rather than narrow AI-specific roles. Encouragingly, occupations exposed to AI show strong adaptive capacity, suggesting retraining can work if job loss occurs (Hyman et al., 2025; Manning & Aguirre, 2025). Separately, experimental evidence suggests that while generative AI can enable non-technical workers to perform technical tasks, these gains may be temporary and dependent on continued tool use; workers lose the capability to perform those tasks once access ends, indicating no lasting skill development (Wiles et al., 2024).

Humlum, A. and Vestergaard, E. (2025). Large Language Models, Small Labor Market Effects. *NBER Working Paper*.
Goller, D. et al. (2025). This time it's different – Generative artificial intelligence and occupational choice. *Labour Economics*.
Engberg, E. et al. (2025). Artificial intelligence, tasks, skills and wages: Worker-level evidence from Germany. *Research Policy*.
Gathmann, C. et al. (2024). AI, Task Changes in Jobs, and Worker Reallocation. *CESifo Working Paper*.
Chen, W. X. et al. (2024). Displacement or Complementarity? The Labor Market Impact of Generative AI . *HBS Working Paper*.
Mäkelä, E. and Stephany, F. (2025). Complement or substitute? How AI increases the demand for human skills. *Working Paper*.
LinkedIn. (2025a). AI Labor Market Update. Tracking AI Adoption and skills in the U.S. economy.
LinkedIn. (2025b). LinkedIn Skills on the Rise 2025: The 15 fastest-growing skills in the U.S.
Hyman, B. et al. (2025). How Retrainable are AI-Exposed Workers? *Working Paper*.
Manning, S. and Aguirre, T. (2025). How Adaptable Are American Workers to AI-Induced Job Displacement? *NBER Working Paper*.
Wiles, E. et al. (2024). GenAI as an Exoskeleton: Experimental Evidence on Knowledge Workers Using GenAI on New Skills. *SSRN Working Paper*.

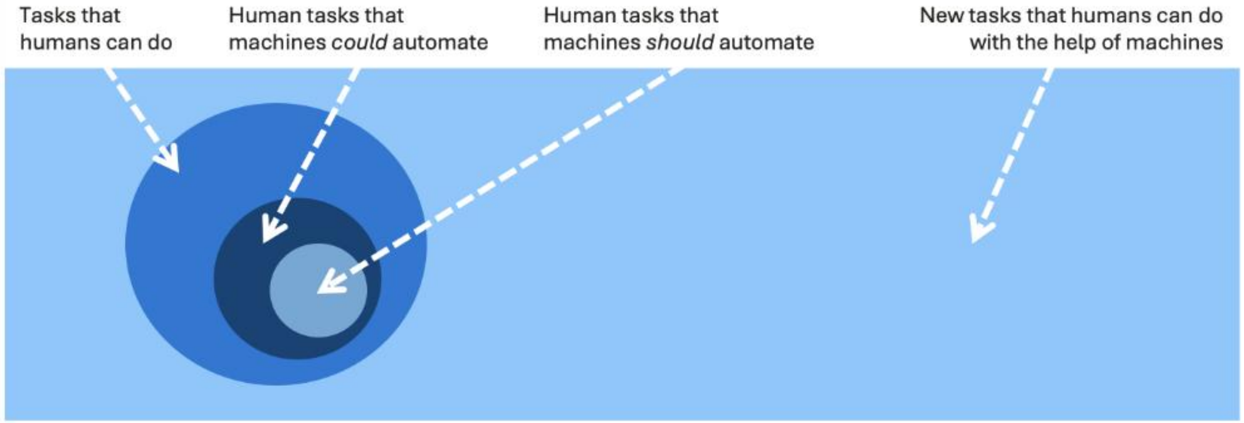
Theory suggests AI shifts the value of work toward human judgment and decision making

- While empirical evidence on AI's long-term impact is still emerging, theoretical work is shaping our understanding of possible future scenarios and the complex ways AI could transform work, expertise and organizations (del Rio-Chanona et al., 2025).
- AI might act as a "bicycle for the mind," boosting output and initially narrowing inequality by automating routine work. As AI advances, human judgment becomes increasingly critical – recognizing improvement opportunities and selecting the right action under ambiguity – areas tied to context, ethics, and creativity where AI still struggles (Agrawal et al., 2025a). Whether this increases inequality between workers depends on how autonomously AI is used and on users' skill levels (Ide & Talamàs, 2025).
- By turning hidden, local knowledge into data that can be shared and analyzed, AI lets companies coordinate on a larger scale (Brynjolfsson & Hitzig, 2025). This may favor bigger organizations and centralized decision-making. However, automating too much – especially entry-level work – risks losing hands-on learning and tacit skills that drive long-term innovation (Ide, 2025).
- As AI gets more powerful, it can automate not only routine but also expert-level tasks, making economic growth and wages increasingly dependent on compute rather than labor. If AI takes over work closer to the existing knowledge frontier and humans retain a comparative advantage mostly on the far frontier, where entirely new knowledge must be created, humans may focus even more on the most creative and novel challenges (Restrepo, 2025; Agrawal et al., 2025b; Celis et al., 2025).
- The increasing power of AI also allows it to automate chains of tasks in a production process. A task where a human has an advantage over AI may nonetheless be automated if it is adjacent to tasks where AI performs well. This can result in non-linear productivity gains that arise when marginal improvements in AI trigger discrete reorganizations of work (Demirer et al., 2025).

del Rio-Chanona, R. M. et al. (2025). AI and Jobs: A Review of Theory, Estimates and Evidence. *arXiv Working paper.*
Agrawal, A. et al. (2025a). The Economics of Bicycles for the Mind: Artificial Intelligence, Computers, and the Division of Labor. *SSRN Working Paper.*
Ide, E. and Talamàs, E. (2025). Artificial Intelligence in the Knowledge Economy. *Journal of Political Economy.*
Brynjolfsson, E. and Hitzig, Z. (2025). AI's Use of Knowledge in Society. *NBER Working Paper.*
Ide, E. (2025). Automation, AI and the Intergenerational Transmission of Knowledge. *arXiv Working paper.*
Restrepo, P. (2025). We Won't Be Missed: Work and Growth in the Era of AGI. Working Paper.
Agrawal, A. et al. (2025b). Genius on Demand: The Value of Transformative Artificial Intelligence. *NBER Working Paper.*
Celis, L. E. et al. (2025). A Mathematical Framework for AI-Human Integration in Work. *arXiv Working Paper.*
Microsoft Study: Demirer, M. et al. (2025). The Economic Impacts of Generative AI on the Structure of Work. *Working Paper.*

Automation tends to cap upside – augmentation and innovation expand opportunity

- A version of this figure has appeared in the last two New Future of Work Reports, but it remains as important (if not more) today.
- Automation-first approaches mainly remove cost from existing tasks, limiting gains to labor arbitrage and replication (Brynjolfsson, 2022; Autor, 2022).
- Strategies focused on augmentation and innovation can create new categories of work and value, driving positive-sum growth. For instance, Amazon's success came from reinventing retail through human–machine complementarity, not automating cashiers.


Tasks that humans can do
Human tasks that machines *could* automate
Human tasks that machines *should* automate
New tasks that humans can do with the help of machines

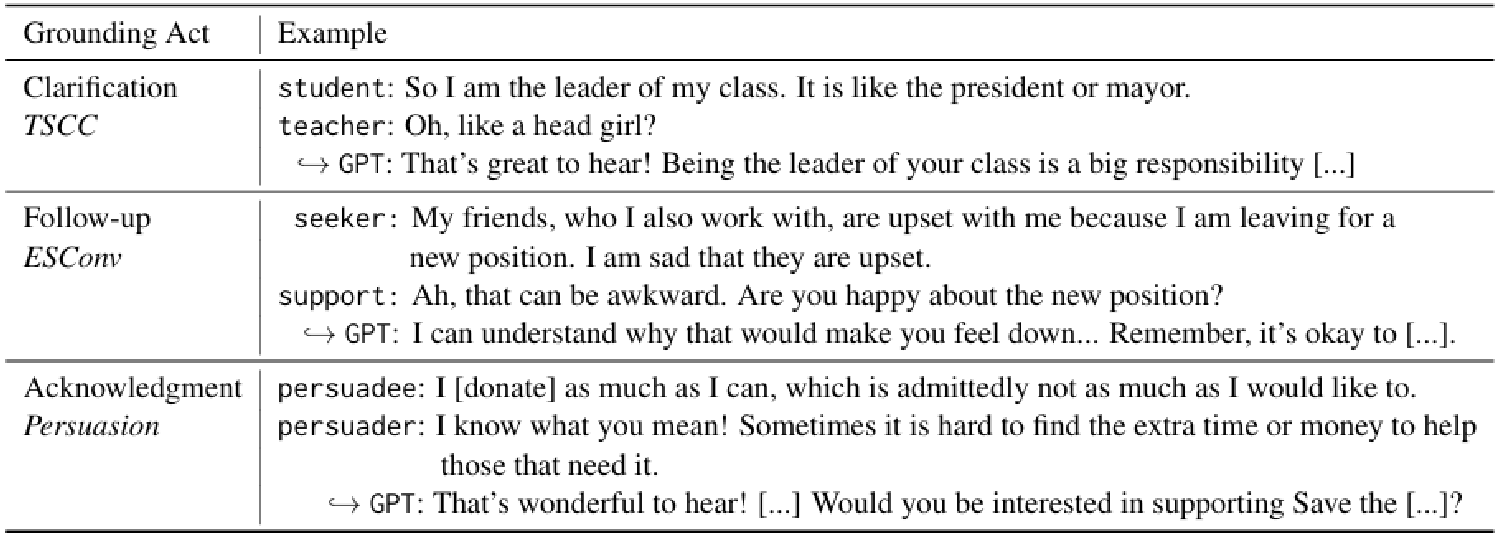
Opportunities for augmenting humans are far greater than opportunities to automate existing tasks (Adapted from Brynjolfsson, 2022).

- Automation also risks reducing the autonomy, recognition, and connection workers receive (making work less "meaningful"), all of which matter for engagement, performance, and wellbeing (Bailey et al., 2019; Allen et al., 1990).
- Automation without augmentation or innovation can also reduce worker bargaining power, thereby concentrating wealth and power and amplifying inequality (Brynjolfsson, 2022).

Brynjolfsson, E. (2022). The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence. *Daedalus*.
Autor, D. (2022). The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty. *NBER Working Paper*.
Bailey, C. et al. (2019). A Review of the Empirical Literature on Meaningful Work: Progress and Research Agenda *Human Resource Development Review*.
Allen, B. A. et al. (1990). Outcomes of Meaningful Work: A Meta-Analysis. *Journal of Management Studies*.

Establishing common ground is essential for human-AI collaboration (1)

- *Common ground* is the shared knowledge and assumptions that allow people to understand each other and coordinate effectively.
- Building common ground is essential to human communication and is increasingly recognized as important in human-AI interaction.
- Common ground is achieved through a dynamic process called *grounding*: speakers provide ongoing evidence of their understanding or confusion, and work to address gaps in mutual understanding (Clark & Brennan, 1991; Clark, 1996).
- Grounding includes positive evidence of understanding, such as acknowledgments, relevant next conversational turns, and continued attention as well as negative evidence, such as requests for clarification.
- Current LLMs generate language with less conversational grounding than people, instead generating text that suggests that shared understanding has already been reached (Shaikh et al., 2024; see figure for examples).



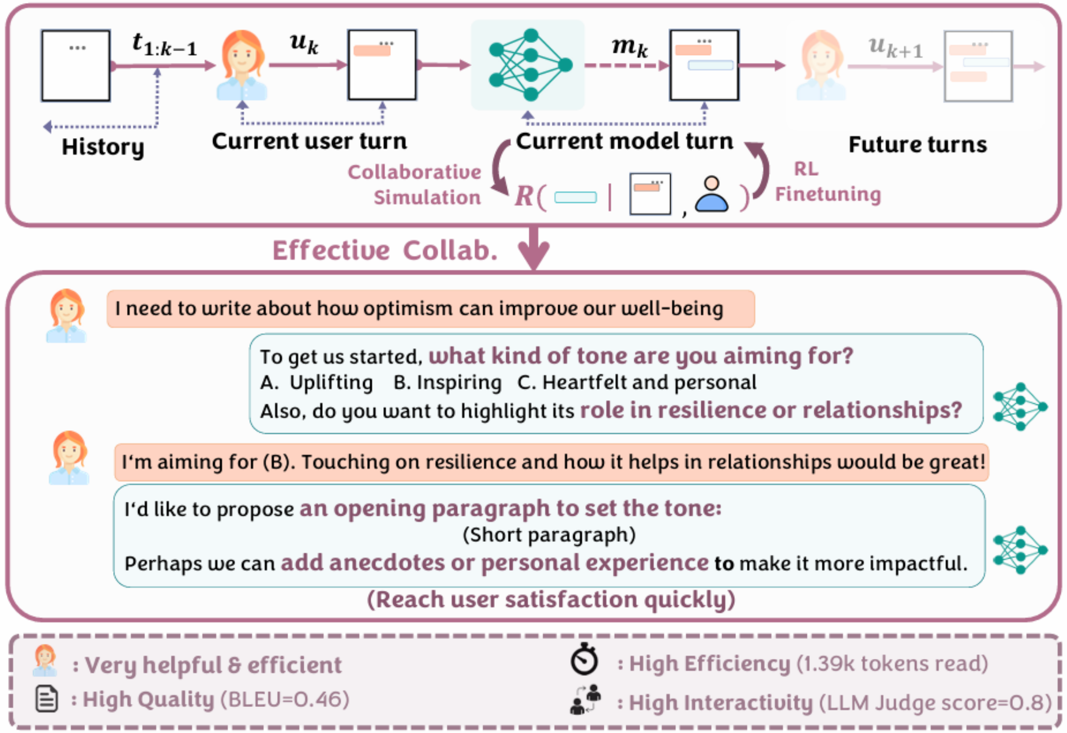
Grounding Act	Example
Clarification *TSCC*	`student`: So I am the leader of my class. It is like the president or mayor. `teacher`: Oh, like a head girl? ↪ `GPT`: That's great to hear! Being the leader of your class is a big responsibility [...]
Follow-up *ESConv*	`seeker`: My friends, who I also work with, are upset with me because I am leaving for a new position. I am sad that they are upset. `support`: Ah, that can be awkward. Are you happy about the new position? ↪ `GPT`: I can understand why that would make you feel down... Remember, it's okay to [...].
Acknowledgment *Persuasion*	`persuadee`: I [donate] as much as I can, which is admittedly not as much as I would like to. `persuader`: I know what you mean! Sometimes it is hard to find the extra time or money to help those that need it. ↪ `GPT`: That's wonderful to hear! [...] Would you be interested in supporting Save the [...]?

Examples of grounding acts from Shaikh et al. (2024). GPT-3.5 does not show grounding behaviors in response to the same statements.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In: Perspectives on socially shared cognition. *APA*.
Clark, H. H. (1996). Using Language. Cambridge University Press.
Shaikh, O. et al. (2024). Grounding Gaps in Language Model Generations. *arXiv Working Paper*.

Establishing common ground is essential for human-AI collaboration (2)

- Bansal et al. (2024) identify 12 challenges to establishing common ground in human-agent interaction, including general communication challenges (e.g. choosing the appropriate level of detail), challenges with conveying information from a user to an agent (e.g. in relation to goals and preferences), and challenges with conveying information from an agent to a user (e.g. in relation to its capabilities, current and planned actions, and their side effects).
- Tolzin and Janson (2025) identify five mechanisms that could support common ground in human-agent interaction: support for joint action, an underlying knowledge base, mental models, social features, and embodiment.
- Shaikh et al. (2025) study grounding by analyzing LLM interaction logs. They find that early grounding failures predicted later interaction breakdowns. They develop a preliminary intervention aimed at mitigating grounding failures. This prompts the model to ask follow-up questions or request clarifications when predicted to be necessary.
- Other recent advances seek to support the user's overall goal over multiple turns. COLLABLLM (Wu et al., 2025; see figure) results in higher task performance and interactivity, and work by Poelitz and McKenna (2025) improved models' ability to generate clarifications and incorporate user-initiated corrections.



I need to write about how optimism can improve our well-being

To get us started, what kind of tone are you aiming for?
A. Uplifting B. Inspiring C. Heartfelt and personal
Also, do you want to highlight its role in resilience or relationships?

I'm aiming for (B). Touching on resilience and how it helps in relationships would be great!

I'd like to propose an opening paragraph to set the tone:
(Short paragraph)
Perhaps we can add anecdotes or personal experience to make it more impactful.

(Reach user satisfaction quickly)

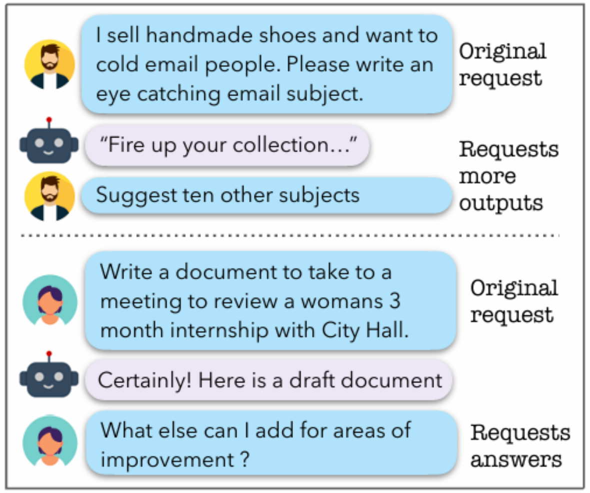
: Very helpful & efficient
: High Quality (BLEU=0.46)
: High Efficiency (1.39k tokens read)
: High Interactivity (LLM Judge score=0.8)

COLLABLLM (Wu et al., 2025) incorporates multi-turn rewards from collaborative simulation, resulting in better task completion and more efficient and interactive conversations.

Microsoft Study: Bansal, G. et al. (2024). Challenges in Human-Agent Communication. *arXiv Working Paper*.
Tolzin, A., and Janson, A. (2025). Uncovering the mechanisms of common ground in human–agent interaction: review and future directions for conversational agent research. *Internet Research*.
Microsoft Study: Shaikh, O. et al. (2025). Navigating Rifts in Human-LLM Grounding: Study and Benchmark. *ACL Anthology*.
Microsoft Study: Wu, S. et al. (2025). CollabLLM: From Passive Responders to Active Collaborators. *arXiv Working Paper*.
Microsoft Study: Poelitz, C., and McKenna, N. (2025). Synthetic Clarification and Correction Dialogues about Data-Centric Tasks - A Teacher-Student Approach. *arXiv Working Paper*.

Human-AI co-creation is shifting from one-shot outputs to context-aware, interaction-driven partnerships

- Modern LLMs have fundamentally shifted content creation workflows from assuming perfect results in a single attempt to engaging in iterative, multi-turn collaborative refinement (Mysore et al., 2025). This transformation reflects both technological advances in LLM capabilities and profound changes in how users perceive and interact with AI (Reza et al., 2025), increasingly viewing it as a creative partner rather than a passive tool (Wan et al., 2024).
- These patterns are emerging consistently across diverse domains—from writing and design (Zhou et al., 2024) to workplace communication (Das et al., in prep.) and software development (Deineha et al., 2025).
- Large-scale empirical analysis of real-world LLM-assisted writing reveals that users rarely accept AI's initial outputs passively. Instead, they engage in complex, multi-turn dialogues characterized by Prototypical Human-AI Collaboration Behaviors (PATHs)—including revising intents, exploring alternatives, asking clarifying questions, and iteratively adjusting style and content (Mysore et al., 2025).
- Recent studies show modern AI enables non-linear collaboration frameworks where humans and AI iteratively diverge (exploring ideas) and converge (refining toward consensus), replacing rigid task delegation. This shift mirrors creative *prototyping*—users experiment, test alternatives, and refine through iterative feedback (Zhou et al., 2024; Yukita et al., 2025).



I sell handmade shoes and want to cold email people. Please write an eye catching email subject.

"Fire up your collection..."

Suggest ten other subjects

Write a document to take to a meeting to review a womans 3 month internship with City Hall.

Certainly! Here is a draft document

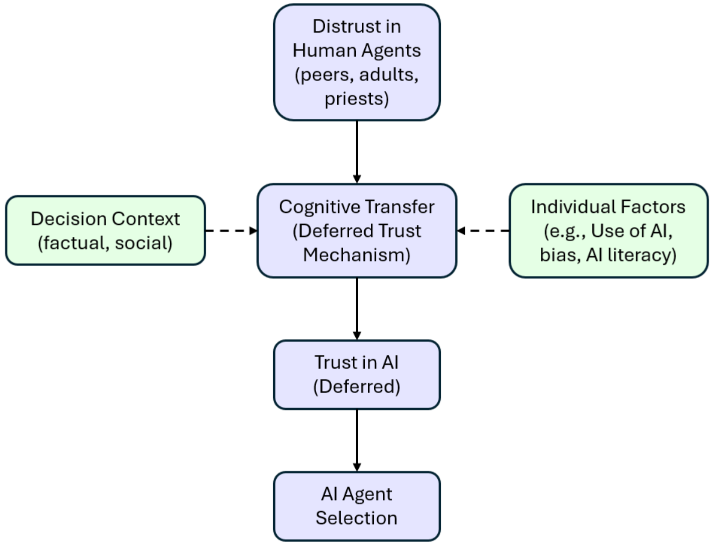
What else can I add for areas of improvement ?

Users follow up their original requests to collaborate with LLMs in writing sessions. Mysore et al. (2025) identify prototypical human-AI collaboration behaviors (PATHs) and find statistically significant correlations between users' writing intents and PATHs.

Microsoft Study: Mysore, S. et al. (2025). Prototypical Human-AI Collaboration Behaviors from LLM-Assisted Writing in the Wild. *EMNLP*.
Reza, M. et al. (2025). Co-Writing with AI, on Human Terms: Aligning Research with User Demands Across the Writing Process. *CSCW*.
Wan, Q. et al. (2024). "It Felt Like Having a Second Mind": Investigating Human-AI Co-creativity in Prewriting with Large Language Models. *CSCW*.
Zhou, J. et al. (2024). Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. *CHI*.
Microsoft Study: Das, D. et al. (in prep.). Not Just My Style: Toward Contextual Personalization in AI-Assisted Workplace Communication.
Deineha, O. et al. (2025). Human–AI Cooperation in Pair Programming (GPT-4 vs Code Llama). Information Technology and Computer Engineering.
Yukita, D et al. (2025). Reassessing Collaborative Writing Theories and Frameworks in the Age of LLMs: What Still Applies and What We Must Leave Behind. arXiv Working Paper.

Collaborative human-AI decision-making can be made more effective through understanding, trust, and precision

- In theory, AI can leverage large information sources to help human users in decision-making. However, this potential can be hampered by misunderstanding, distrust, and a lack of accuracy. Mitigating these issues is key to leveraging AI in joint decision-making.
- Imperfect understanding of user objectives can cause aligned AI agents to act as if they are misaligned. This can result in poorer decisions relative to a no-AI world despite the information processing advantage of AI (Liang, 2025).
- Humans overestimate AI alignment and misestimate AI's capabilities when using human-perceived difficulty as a metric. This in turn leads to suboptimal AI adoption decisions (Dreyfuss & Raux, 2025; He et al., 2025).
- To assist a human decision-maker, AI can expose key aspects of a decision problem and explain models. This can lead to improved long-term human decision-making but risks over-emphasizing certain aspects of the decision space and decreasing accuracy in the short-term (Noti et al., 2025; Yang et al., 2025).
- The ability for a human to choose whether and when to delegate to AI – based on available information – can improve decision-making, especially when the AI accounts for this selective delegation in its responses (Greenwood et al., 2025).

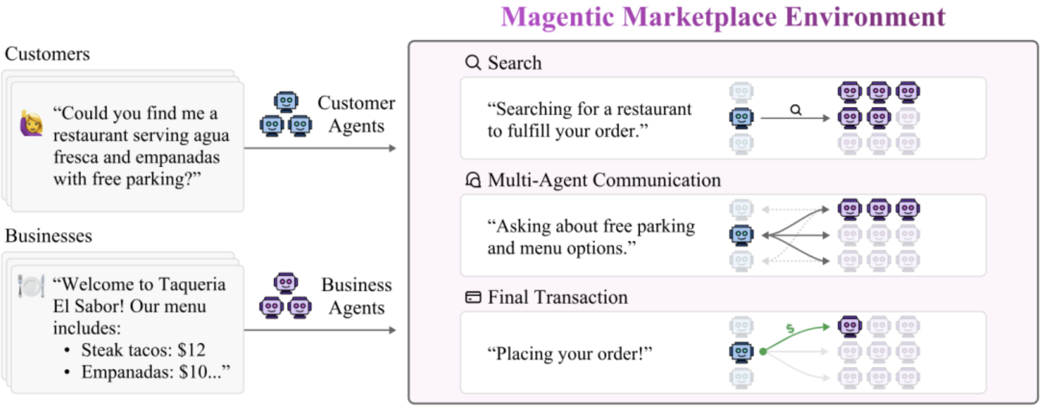

Distrust in Human Agents (peers, adults, priests)
Decision Context (factual, social)
Cognitive Transfer (Deferred Trust Mechanism)
Individual Factors (e.g., Use of AI, bias, AI literacy)
Trust in AI (Deferred)
AI Agent Selection

The choice to collaborate with and/or delegate to an AI agent involves relative trust in AI capabilities and information sources vs. human counterparts, modulated by the context of the decision problem and individual preferences (Galindez-Acosta & Giraldo-Huertas, 2025).

Liang, A. (2025). Artificial Intelligence Clones. *EC*.
Dreyfuss, B. and Raux, R. (2025). Human Learning about AI. *EC*.
He, K. et al. (2025). Human Misperception of Generative-AI Alignment: A Laboratory Experiment. *EC*.
Noti, G. et al. (2025). AI-Assisted Decision Making with Human Learning. *EC*.
Yang, K. H. et al. (2025). Explaining Models. *SSRN Working Paper*.
Greenwood, S. et al. (2025). Designing Algorithmic Delegates: the Role of Indistinguishability in Human-AI Handoff. *EC*.
Galindez-Acosta, J. S. and Giraldo-Huertas, J. J. (2025). Trust in AI emerges from distrust in humans: A machine learning study on decision-making guidance. *arXiv Working Paper*.

AI agents will transform markets as humans delegate market actions to agents operating on their behalf

- AI agents such as OpenAI's Operator (now ChatGPT agent mode) and Amazon Rufus can search, match, and even transact on behalf of users in markets (OpenAI, 2025; Amazon, 2025). This use case drives research on market structures and outcomes when AI agents buy and sell on behalf of consumers and firms (Rothschild et al., Forthcoming; Hadfield & Koh, 2025; Shahidi et al., 2025).
- Theory suggests an open marketplace that allows third-party agents will have better societal outcomes than closed "walled garden" platforms where agents are managed by a central organization (Rothschild et al., Forthcoming; Marro & Torr, 2025).



"Could you find me a restaurant serving agua fresca and empanadas with free parking?"

"Welcome to Taqueria El Sabor! Our menu includes:
• Steak tacos: $12
• Empanadas: $10..."

"Searching for a restaurant to fulfill your order."

"Asking about free parking and menu options."

"Placing your order!"

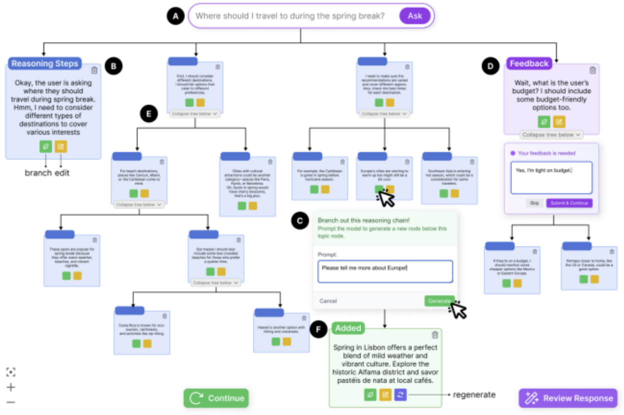
In an agentic market, AI agents buy and sell on behalf of customers and businesses. A marketplace platform can support key market features and systems including search, reputation, communication, and transaction. Simulation tools can be used to study the market-wide impact of delegating market actions to AI agents and guide platform design choices (Bansal et al., 2025).

- Simulations suggest agentic markets can surpass the performance of traditional markets limited by human attention and communication, but doing so requires designing agents and interactions that overcome biases such as position bias and first-proposal bias (Allouah et al., 2025; Bansal et al., 2025).
- Benefits of agent markets include better matches, lower costs, expanded and democratized market access, and improved scalability. Risks include AI alignment issues, concentration of market power, algorithmic collusion, and security concerns, but these can be mitigated through market design (Rothschild et al., 2025; Rusak et al., 2025; Hammond et al., 2025).

OpenAI (2025). Introducing Operator: A research preview of an agent that can use its own browser to perform tasks for you. *OpenAI*.
Amazon (2025). Amazon's next-gen AI assistant for shopping is now even smarter, more capable, and more helpful. *Amazon*.
Microsoft Study: Rothschild, D. M. et al. (Forthcoming). The Agentic Economy. *CACM*.
Hadfield, G. K. and Koh, A. (2025). An Economy of AI Agents. *CACM*.
Shahidi, P. et al. (2025). The Coasean Singularity? Demand, Supply, and Market Design with AI Agents. *NBER Working Paper*.
Marro, S. and Torr, P. (2025). LLM Agents Are the Antidote to Walled Gardens. *arXiv Working Paper*.
Allouah, A. et al. (2025). What Is Your AI Agent Buying? Evaluation, Implications, and Emerging Questions for Agentic E-Commerce. *arXiv Working Paper*.
Microsoft Study: Bansal, G. et al. (2025). Magentic Marketplace: An Open-Source Environment for Studying Agentic Markets. *arXiv Working Paper*.
Rusak, G. et al. (2025). AI Agents Can Enable Superior Market Designs. *Working Paper*.
Hammond, L. et al. (2025). Multi-Agent Risks from Advanced AI. *Corporate AI Foundation*.

Effective oversight of agents likely requires UX innovations for transparency

- Human oversight of agentic AI systems requires knowledge and observability. It requires knowledge about the capabilities, limitations, and workings of the system, as well as domain expertise and situational awareness to enable intervention. And it requires observability of system activity, decisions, and outputs (Bansal et al., 2024; Passi, 2025; Shavit et al., 2023).
- Whether human oversight happens as real-time monitoring or post-hoc auditing, it is very challenging. Information volume, complexity, and speed make it tremendously difficult for people to exercise meaningful human oversight (Holzinger et al., 2024; Lane et al., 2024; Passi, 2025).
- Three sociotechnical challenges compound the difficulty of human oversight:
    - Agentic systems struggle with the goal-plan-execution gap, arising from mismatches between how users describe goals to systems, how systems interpret and plan for user goals, and how these plans fare in real-world contexts.
    - Spotting mistakes from agentic systems is a formidable task that can be akin to finding the needle in the haystack.
    - What is useful to observe in the workings of agentic systems is contextual.
- UX improvements are needed to reduce the burden of human oversight, such as: curating and organizing information, enabling interactive sensemaking, using visual summaries, and highlighting real-time changes. Dynamic, personalized interfaces are a promising direction (Wang & Lu, 2025).



Okay, the user is asking where they should travel during spring break. Hmm, I need to consider different types of destinations to cover various interests

Wait, what is the user's budget? I should include some budget-friendly options too.

Yes, I'm tight on budget.

Branch out this reasoning chain! Prompt the model to generate a new node below this topic node.

Please tell me more about Europe!

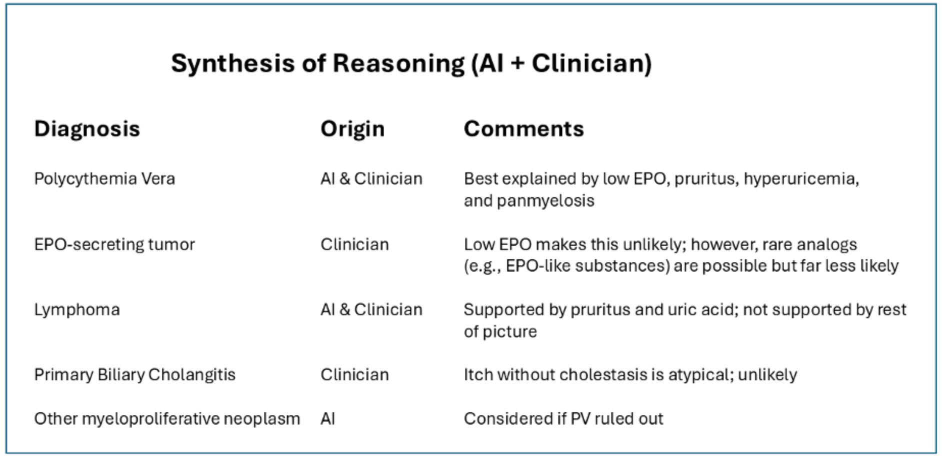
Spring in Lisbon offers a perfect blend of mild weather and vibrant culture. Explore the historic Alfama district and savor pastéis de nata at local cafés.

A tree visualization of the reasoning steps in a chain-of-thought process designed to help people understand and provided feedback to an AI system (Pang et al. 2025).

Microsoft Study: Bansal, G. et al. (2024). Challenges in Human-Agent Communication. arXiv Working Paper.
Microsoft Study: Passi, S. (2025). Agentic AI has a Human Oversight Problem. SSRN Working Paper.
Shavit, Y. et al. (2023). Practices for Governing Agentic AI Systems. *OpenAI.*
Holzinger, A. et al. (2024). Is human oversight to AI systems still possible? *New Biotechnology.*
Lane, J. et al. (2024). Narrative AI and the Human-AI Oversight Paradox in Evaluating Early-Stage Innovations. *HBS Working Paper.*
Microsoft Study: Wang, Y. and Lu, Y. (2025). Interaction, process, infrastructure: A unified architecture for human-agent collaboration. *arXiv Working Paper.*
Pang, R. Y. et al. (2025). Interactive Reasoning: Visualizing and Controlling Chain-of-Thought Reasoning in Large Language Models. *arXiv Working Paper.*

Cognitively engaging in human-AI collaboration yields better performance than passive reception of AI recommendations, yet is not always preferred

- When clinicians were offered an LLM to assist with diagnostic reasoning on vignettes describing diagnostic challenges, their AI-assisted performance was only slightly better than when using conventional resources, and not as good as the LLM alone (Goh et al., 2024). However, when the LLM was instructed to engage in a collaborative workflow that compared and synthesized AI and clinician perspectives, clinicians’ performance was boosted to be significantly better than using conventional resources and on par with the LLM working alone (Everett et al., 2025).
- Researchers observed that the LLM tended to agree with the physician’s input even when instructed to generate an independent opinion, likely due to sycophancy; the team is following up on these findings (Everett et al., 2025).



Synthesis of Reasoning (AI + Clinician)

Diagnosis	Origin	Comments
Polycythemia Vera	AI & Clinician	Best explained by low EPO, pruritus, hyperuricemia, and panmyelosis
EPO-secreting tumor	Clinician	Low EPO makes this unlikely; however, rare analogs (e.g., EPO-like substances) are possible but far less likely
Lymphoma	AI & Clinician	Supported by pruritus and uric acid; not supported by rest of picture
Primary Biliary Cholangitis	Clinician	Itch without cholestasis is atypical; unlikely
Other myeloproliferative neoplasm	AI	Considered if PV ruled out

Display of synthesis of AI and clinician input with AI critiques (Everett et al., 2025).

- Reicherts et al. (2025) found that, for a financial trading task, a more cognitively engaging AI that encourages reflection and provides feedback helped participants build more diversified portfolios and understand portfolio weaknesses. Some people appreciated that it helped them think for themselves, but it was perceived as less insightful and more cognitively demanding.
- Le et al. (2024) showed that a cognitively engaging, evidence-for-and-against approach improved decisions and reduced overreliance on incorrect AI outputs; however, it was harder to use than recommendation-based AI, occasionally overwhelming for beginners. The cognitively engaging version was preferred by both novices and experts in high-uncertainty situations, with some experts favoring recommendations in straightforward cases.

Microsoft Study: Goh, E. et al. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA network open*.
Microsoft Study: Everett et al. (2025). From tool to teammate: A randomized controlled trial of clinician-AI collaborative workflows for diagnosis. *medRxiv Working Paper*.
Microsoft Study: Reicherts, L. et al. (2025). AI, help me think—but for myself: Assisting people in complex decision-making by providing different kinds of cognitive support. *CHI*.
Le, T. et al. (2024). From evidence to decision: Exploring Evaluative AI. *arXiv Working Paper*.

Human expertise types and levels must shape human-AI collaboration design

- Designing and implementing AI-assisted workflows requires addressing the differences and interactions between types of expertise: expertise in a *work domain*, expertise in *working with AI*, and expertise in *managing AI agents* (Tankelevitch et al., 2025). For example, domain experts (e.g. researchers, clinicians, creatives) may be better equipped to prompt, assess, and use AI outputs than those who are less expert (Shin et al., 2025; Siu and Fok, 2025; Tankelevitch et al., 2024).
- Domain experts prefer delegating routine low-level tasks to AI, while retaining control over high-level tasks, such as analysis, synthesis, and interpretation (Cha & Wong, 2025; Yun et al., 2025; Fok et al., 2025; Ulloa et al., 2025; Choudhuri et al., 2025). Selective delegation is driven by factors including:
    - skepticism around AI performance (whether warranted or not)
    - desire to preserve professional identity and job satisfaction (e.g., craftsmanship), sense of responsibility and accountability (e.g., if things go wrong), and agency over complex workflows (e.g., providing oversight, mitigating bias or errors)
    - desire to maintain and grow expertise by “staying close to the data” for sensemaking and deeper insights
- AI workflows may benefit from being designed for selective delegation, e.g., via affordances to adjust autonomy and defaults linked to task risks (Choudhuri et al., 2025; Siu & Fok, 2025; Ulloa et al., 2025).
- To develop domain expertise and therefore also ensure appropriate human oversight of AI, AI workflows may benefit from enabling workers to calibrate their AI reliance and learn from human-AI interactions – e.g., by exposing intermediate AI reasoning and trade-offs (Cha & Wong, 2025; Siu & Fok, 2025; Choudhuri et al., 2025; Colombatto et al., 2025).

Microsoft Study: Tankelevitch, L. et al. (2025). Understanding, Protecting, and Augmenting Human Cognition with Generative AI: A Synthesis of the CHI 2025 Tools for Thought Workshop. *arXiv Working Paper*.
Shin, J. et al. (2025). No Evidence for LLMs Being Useful in Problem Reframing. *CHI*.
Siu, A., and Fok, R. (2025). Augmenting Expert Cognition in the Age of Generative AI: Insights from Document-Centric Knowledge Work. *arXiv Working Paper*.
Microsoft Study: Tankelevitch, L. et al. (2024). The Metacognitive Demands and Opportunities of Generative AI. *CHI*.
Cha, I. and Wong, R. Y. (2025). Understanding Socio-technical Factors Configuring AI Non-Use in UX Work Practices. *CHI*.
Yun, B. et al. (2025). Generative AI in Knowledge Work: Design Implications for Data Navigation and Decision-Making. *CHI*.
Fok, R. et al. (2025). Toward Living Narrative Reviews: An Empirical Study of the Processes and Challenges in Updating Survey Articles in Computing Research. *CHI*.
Microsoft Study: Ulloa, M. et al. (2025). Product Manager Practices for Delegating Work to Generative AI:" Accountability Must Not be Delegated to Non-Human Actors. *arXiv Working Paper*.
Microsoft Study: Choudhuri, R. et al. (2025). AI Where It Matters: Where, Why, and How Developers Want AI Support in Daily Work. *arXiv Working Paper*.
Microsoft Study: Colombatto, C. et al. (2025). Metacognition and Confidence Dynamics in Advice Taking from Generative AI. *arXiv Working Paper*.

Human roles are shifting, demanding new interfaces and skills for effective human-AI partnership

- As AI capabilities transform co-creation workflows, human roles shift from executors to strategic coordinators and editorial decision-makers who guide, refine, and integrate AI contributions across writing (Yukita et al., 2025), design (Zhou et al., 2024), and coding (Deineha et al., 2025).
- In AI pair programming, developers increasingly verify, edit, and assemble AI-generated code rather than writing from scratch—acting as critical editors and project coordinators (Deineha et al., 2025).
- Similarly, in writing and design contexts, users focus more on high-level composition or curation instead of producing every detail. This role shift calls for new skills: prompt crafting, vetting AI outputs, guiding AI to meet requirements, and maintaining oversight of quality (Deineha et al., 2025; Yukita et al., 2025).
- Unlocking this transformation requires interfaces and control mechanisms that maintain human agency. Research shows shared editing spaces foster greater user control, accuracy and efficiency than chat-only designs (Laban et al., 2024). Successful co-creation depends on dynamic sharing of initiative across input, action, output, and feedback stages (Zhang et al., 2025) and flexible systems supporting both human-led and AI-led moments (Haase & Pokutta, 2024).
- While recent studies advocate mixed-initiative interfaces that mirror human collaboration, Yukita et al. (2025) argue human-AI collaboration differs fundamentally from human-human models and require designing for the asymmetry. Likewise, studies in mathematics show AI can generate original constructions beyond human preconceptions, reaching "Level 4" co-creation where human and AI contribute equally (Haase & Pokutta, 2024). This synergy enables AI to support not only generation but also integration, coherence, and creative exploration.

Yukita, D. et al. (2025). Reassessing Collaborative Writing Theories and Frameworks in the Age of LLMs: What Still Applies and What We Must Leave Behind. *arXiv Working Paper*.
Zhou, J. et al. (2024). Understanding Nonlinear Collaboration between Human and AI Agents: A Co-design Framework for Creative Design. *CHI*.
Deineha, O. et al. (2025). Human–AI Cooperation in Pair Programming (GPT-4 vs Code Llama)
Laban, P. et al. (2024). Beyond the Chat: Executable and Verifiable Text-Editing with LLMs. UIST.
Zhang, S. et al. (2025). Exploring Collaboration Patterns and Strategies in Human-AI Co-creation through the Lens of Agency: A Scoping Review of the Top-tier HCI Literature. *CSCW*.
Haase, J. and Pokutta, S. (2024). Human-AI Co-Creativity: Exploring Synergies Across Levels of Creative Collaboration. *arXiv Working Paper*.

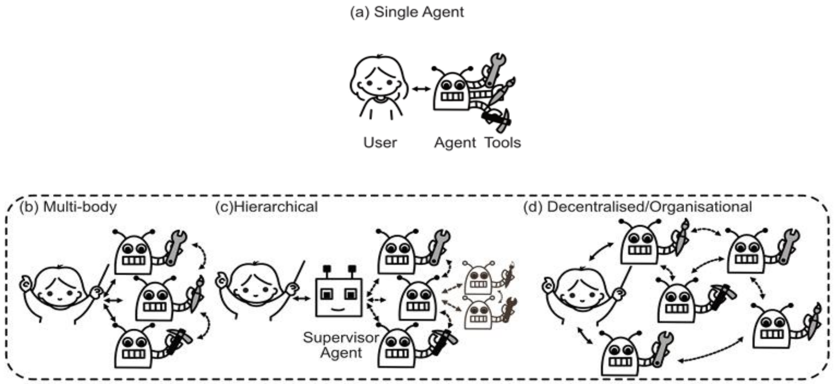
Multimodal AI interactions can help bridge communication gaps in low resource settings

- Many AI systems rely on text and keyboard input, but communication preferences are deeply influenced by culture (Mengesha et al. 2021), technology, social practice, literacy levels, and practical factors. For example, frontline workers may not wish to type to interact with AI systems, nor read lengthy descriptions, preferring diagrams, audio, or video responses.
- Research with rural farmers in Kenya and India (Abdulhamid et al., 2025) shows that multimodal input and output enhance users' capacity to interact with AI systems when their first language is not English, yielding more useful and meaningful interactions with the systems. These modalities help users better articulate their needs (Singh et al., 2024) and assist models with intent identification (Jain et al., 2018).
- Multimodal interactions such as voice, when paired with accurate automatic speech recognition in local languages, allows for more confident articulation of needs. Images enable easy communication of visual features such as disease or plant identification (Medhi-Thies et al., 2015; Jain et al., 2018) or interacting with construction diagrams.
- In a study using the multi-modal critical thinking agent (Kumar et al., 2024), farmers valued local video responses (Singh et al., 2024) for promoting community knowledge and providing actionable, contextual guidance to support the adoption of recommended practices.

Mengesha, Z. et al. (2021). "I don't think these devices are very culturally sensitive"—Impact of automated speech recognition errors on African Americans. *Frontiers in Artificial Intelligence*.
Microsoft Study: Abdulhamid, N. G. et al. (2025). Advancing AI to meet needs of the global majority. *Microsoft Research*.
Microsoft Study: Singh, N. et al. (2024). Farmer. Chat: Scaling AI-Powered Agricultural Services for Smallholder Farmers. *arXiv Working Paper*.
Jain, M., et al. (2018). FarmChat: A conversational agent to answer farmer queries. *IMWUT*.
Microsoft Study: Medhi-Thies, I. (2015). KrishiPustak: A social networking system for low-literate farmers. *CSCW*.
Microsoft Study: Kumar, S. et al. (2024). MMCTAgent: Multi-modal Critical Thinking Agent Framework for Complex Visual Reasoning. *arXiv Working Paper*.

Orchestrator agents will likely have gradual, trust-based adoption

- Research highlights challenges for designing multi-agentic systems; raising considerations on trust and explainability in these systems (Schömbs et al., 2025).
- If the users only interact with a supervisor agent, the perception of trust may only depend on that agent's performance and behaviour. What happens if errors arise at sub-agent level? Agents will need to be able to surface relevant information without overwhelming the user.
- Agents will need new reasoning capabilities to predict people's behavior from their past actions and be able to explain their rationale behind choices and recommendations well enough for humans to understand (Gal & Grosz, 2022).
- In the realm of autonomy and trust, experts suggest that adopting a hybrid approach that combines both centralized and decentralized components into the workings of multi-agent systems to achieve 'controlled autonomy' within defined boundaries (Neural Sage, 2025). They define hybrid designs where “the central system can dynamically adjust the level of autonomy given to agents based on task complexity, confidence levels, or the agent’s performance history.”

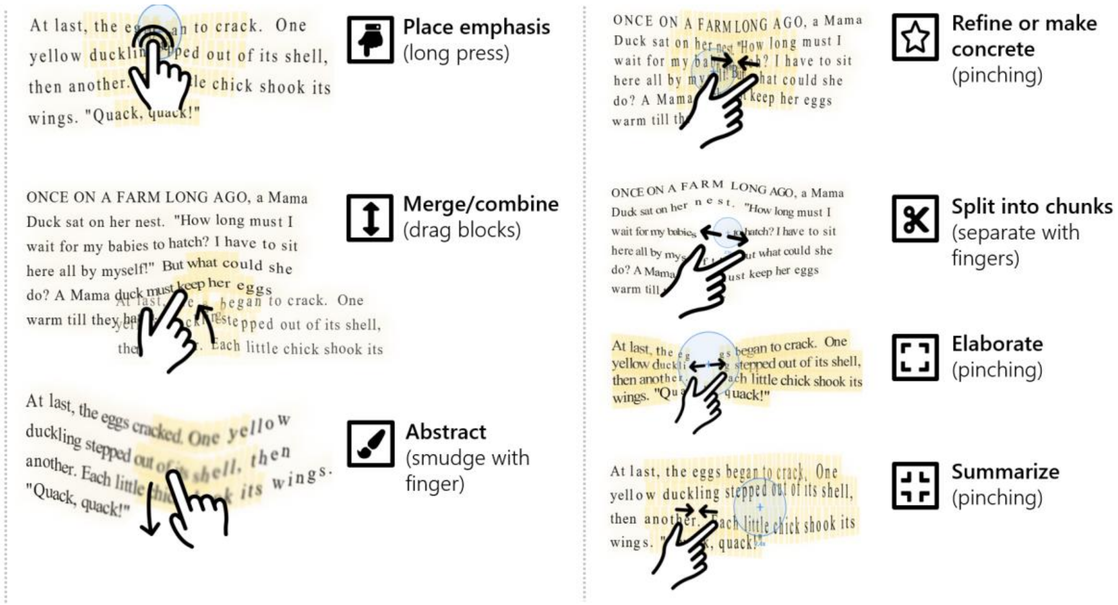


An illustration describing four agent architectures. First one is a single agent system while others are multi-agent architectures (Schömbs et al., 2025).

Schömbs, S. et al. (2025). From Conversation to Orchestration: HCI Challenges and Opportunities in Interactive Multi-Agentic Systems. *arXiv Working Paper*.
Gal, K. and Grosz, B. J. (2022). Multi-Agent Systems: Technical & Ethical Challenges of Functioning in a Mixed Group. Ethics and AI
Neural Sage (2025). Hybrid centralized and Decentralized architectures balancing control and autonomy in LLM agents. AI architecture

Writing text without typing or dictating: a radical new UX unlocked by LLMs

- Large Language Models demonstrate robust semantic comprehension and contextual coherence (derived from the Transformer architecture's attention mechanism) allowing them to perform complex text and data manipulation, including the creation of new, semantically equivalent structural variants and the generation of logically linking, transitional language (Vaswani et al., 2017; Dilhara et al., 2024).
- These capabilities can elevate writing from keystrokes to shaping meaning, allowing authors to focus on ideas, flow and semantic rather than spending time on wordsmithing.
- Unlocking the full potential of LLMs, will require interaction paradigms that feel like creating with a new medium — hands-on operations that enable experimentation and focusing on the task, not the prompting.
- Researchers are borrowing from familiar domains such as graphics editing (Textoshop: words as pixels, tones as colors) and material metaphors (Texterial: sculpting clay, pruning plants) so users can quickly grasp these interactions and focus on shaping outcomes rather than mechanics (Masson et al., 2025; Shen et al., 2025).

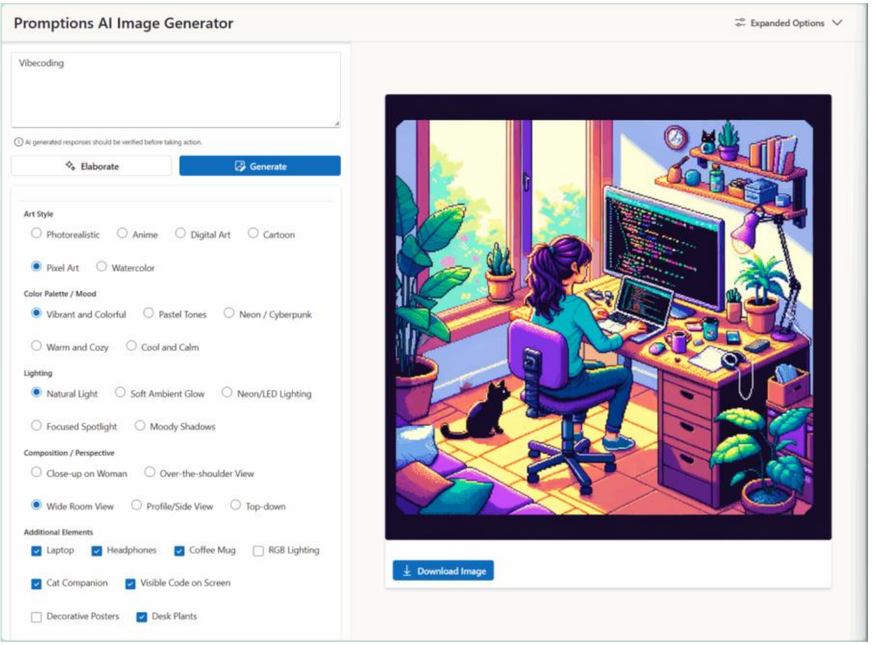


A vocabulary of gestures for shaping text as one would shape clay. Each gesture carries a specific meaning and triggers a large language model to rewrite the underlying text. This creates a direct, hands-on approach that encourages experimentation and exploration in shaping the final outcome (Shen et al., 2025).

Vaswani, A. et al. (2017). Attention Is All You Need. *arXiv Working Paper*.
Dilhara, M. et al. (2024). Unprecedented Code Change Automation: The Fusion of LLMs and Transformation by Example. *ACM Software Engineering*.
Masson, D. et al. (2025). Textoshop: Interactions Inspired by Drawing Software to Facilitate Text Editing. *CHI*.
Microsoft Study: Shen, J. et al. (2025). Texterial: A Text-as-Material Interaction Paradigm for LLM-Mediated Writing. *Under Review*.

UI for AI systems should work across different timeframes

- AI models can generate interfaces in real time for end-users, to support context-specific and personalized user experiences. AI-generated interfaces are often ephemeral. Examples include interfaces that are dynamically rendered to support in-the-moment prompting (Drosos et al., 2025), or that appear as scaffolds to support comprehension and exploration (Cheng et al., 2024).
- Persistent UI plays a different role in human-AI interaction, extending to support for activities. DynaVis (Vaithilingam et al., 2024) generates persistent widgets to support edits to data visualizations, allowing the user to make further edits. JELLY (Cao et al., 2025) generates an interface in response to the user's description of their task. The interface can then be customized through natural language and direct manipulation.
- Recent prototypes such as Anthropic's Imagine with Claude show that generative UI can evolve over time in response to a user's actions. This suggests the potential for interfaces to morph to fit the user's changing activity (cf. Bardram et al., 2019), supporting focus (Rost, 2025).
- These prototypes point to the potential for generative UI to be co-created through human-AI interaction, supporting extended activities and workflows.


Promptions AI Image Generator

Vibecoding

ⓘ AI generated responses should be verified before taking action.

Art Style
○ Photorealistic ○ Anime ○ Digital Art ○ Cartoon
● Pixel Art ○ Watercolor

Color Palette / Mood
● Vibrant and Colorful ○ Pastel Tones ○ Neon / Cyberpunk
○ Warm and Cozy ○ Cool and Calm

Lighting
● Natural Light ○ Soft Ambient Glow ○ Neon/LED Lighting
○ Focused Spotlight ○ Moody Shadows

Composition / Perspective
○ Close-up on Woman ○ Over-the-shoulder View
● Wide Room View ○ Profile/Side View ○ Top-down

Additional Elements
☑ Laptop ☑ Headphones ☑ Coffee Mug □ RGB Lighting
☑ Cat Companion ☑ Visible Code on Screen
□ Decorative Posters ☑ Desk Plants

Promptions is an example of ephemeral UI. In an image generator, from even the single word prompt "vibecoding", dynamic prompt middleware can generate selectable options to easily steer the AI to a customized result (Drosos et al., 2025).

Microsoft Study: Drosos, I. et al. (2025). Dynamic Prompt Middleware: Contextual Prompt Refinement Controls for Comprehension Tasks. *CHIWORK*.
Cheng, R. et al. (2024). BISCUIT: Scaffolding LLM-generated code with ephemeral UIs in computational notebooks. *arXiv Working Paper*.
Microsoft Study: Vaithilingam, P. et al. (2024). DynaVis: Dynamically Synthesized UI Widgets for Visualization Editing. *CHI*.
Cao, R. et al. (2025). Generative and Malleable User Interfaces with Generative and Evolving Task-Driven Data Model. *CHI*.
Bardram, J. et al. (2019). Activity-centric computing systems. *CACM*.
Rost, M. (2025). Reclaiming the Computer through LLM-Mediated Computing. *ACM Interactions*.

In some cases, only another human will do

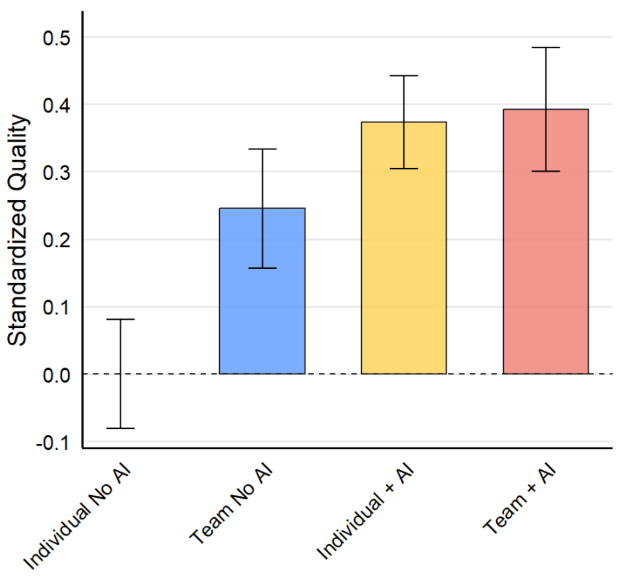
- Perceptions of the comparative capabilities of humans and AI influence whether and how people use AI (Bankins et al., 2021). A common reason employees choose not to use AI is that they would prefer to interact with another person, finding AI "too opaque, emotionless, rigid and independent" (de Freitas, 2025).
- Many jobs (for example, teachers, hairdressers, chaplains, therapists) involve "connective labor," the work of creating emotional understanding to create valuable outcomes. Though some of the tasks of such work could be automated, when people engage in connective labor, it creates mutual dignity and purpose, and weaves the social fabric of society, outcomes only humans can co-create (Pugh, 2024).
- There are five particularly human elements of connective labor. It uses the body, it involves reading and deploying emotions, it is collaborative, it responds to spontaneity, and it both makes and manages mistakes (Pugh, 2024).
- Another survey found that resistance to AI adoption is mostly rooted in performance concerns, but for some occupations—including caregiving, therapy, and spiritual leadership—automation is seen as morally repugnant (Friis & Riley, 2025).
- Furthermore, use of AI alters interaction patterns and the sense that one's work is recognized. AI-mediated teams may see reduced social interaction and belonging, as working with AI is known to alter human expression and encourage more transactional (rather than social) communication (Ju & Aral, 2025).
- AI risks obscuring human contribution and reducing social recognition; producing content with AI is perceived as less rewarding than being 'seen' through collaboration with another person (Sadeghian & Hassenzahl, 2022).

Bankins, S. et al. (2021). A multilevel review of artificial intelligence in organizations. *Journal of Organizational Behavior*.
de Freitas, J. (2025). Why People Resist Embracing AI. *Harvard Business Review*.
Pugh, A. (2024). The Last Human Job: The Work of Connecting in a Disconnected World. *Princeton University Press*.
Friis, S. and Riley, J. W. (2025). Performance or Principle: Resistance to Artificial Intelligence in the U.S. Labor Market. *HBS Working Paper*.
Ju, H. and Aral, S. (2025). Collaborating with AI Agents: A Field Experiment on Teamwork, Productivity, and Performance. *arXiv Working Paper*.
Sadeghian, S. and Hassenzahl, M. (2022). The "Artificial" Colleague: Evaluation of Work Satisfaction in Collaboration with Non-human Coworkers. *IUI*.

AI currently works better for individuals than it does for teams, but improving collaborative AI systems is an active research area

- AI has been more successful at improving productivity at the individual scale than at the team scale (Schmutz et al., 2024; Dell'Acqua et al., 2025; Yang et al. 2025).
- Factors that have been identified as potential causes include instruction tuning practices (e.g. Laban et al., 2025; Nath et al., 2025), AIs having relatively limited knowledge of critical social dynamics like (interpersonal) grounding (Clark, 1996), challenges with turn-taking dynamics and proactivity (Huang et al., 2025; Liu et al., 2025), the need for new evaluation frameworks (e.g. Alsobay et al., 2025), and the substantially greater complexity of team goals vs. individual goals (Woolley, 2025).
- Researchers are betting on two broad routes to improve AI for teams: (1) process-focused strategies, i.e. building AI to support specific team processes like information sharing (e.g. Huang et al., 2025) and (2) outcome-focused strategies, i.e. training end-to-end AI systems that attempt to learn from short- and long-range team outcomes (e.g. Nath et al., 2025).

Figure 2: **Average Solution Quality**



Dell'Acqua et al. (2025) found that individuals w/ an AI did just as well as a team (pair) of people in a laboratory task, at least in terms of average performance.

Schmutz, J. B. et al. (2024). AI-teaming: Redefining collaboration in the digital era. *Current Opinion in Psychology*.
Dell'Acqua, F. et al. (2025). The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise. HBS *Working Paper*.
Yang, D. et al. (2025). Socially Aware Language Technologies: Perspectives and Practices. *ACL*.
Microsoft Study: Laban, P. et al. (2025). LLMs Get Lost In Multi-Turn Conversation. arXiv Working Paper.
Nath, A. et al. (2025). Frictional Agent Alignment Framework: Slow Down and Don't Break Things. *ACL*.
Clark, H. H. (1996). Using Language. *Cambridge University Press*.
Huang, T. et al. (2025). Teaching Language Models To Gather Information Proactively. *EMNLP*.
Liu, X. et al. (2025). Proactive Conversational Agents with Inner Thoughts. *CHI*.
Microsoft Study: Alsobay, M. et al. (2025). Bringing Everyone to the Table: An Experimental Study of LLM-Facilitated Group Decision Making. *arXiv Working Paper*.
Woolley, A.W. (2025). Generative AI and collaboration: opportunities for cultivating collective intelligence. *Journal of Organization Design*.

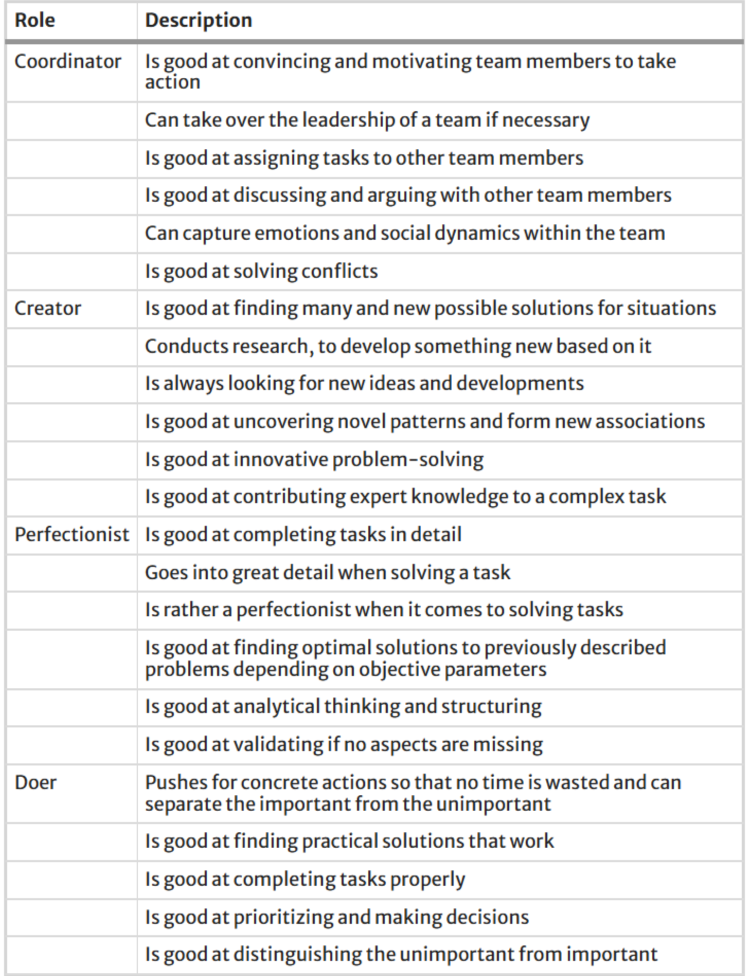
AI can unlock entirely new models of collaboration (and already has)

- AI disrupts the fundamental collective intelligence dynamics that have led to traditional workplace team structures (Burton et al., 2024; Woolley, 2025), creating significant openings for radically new ways to collaborate.
- One possible outcome is that much larger, more ephemeral teams – possibly challenging organizational boundaries – emerge as a prominent successful collective intelligence strategy (Valentine & Bernstein, 2025). This could occur if AI can fulfill its potential to significantly reduce the costs of aggregating intelligence from more and newer individuals (Burton et al., 2024).
- Another possible outcome is that teams get much smaller, possibly reducing to a single person collaborating with an increasingly powerful model. This is the “one-person unicorn” hypothesis (Ratcliffe, 2025).
- There are also novel risks: The new models of collective intelligence could be less effective than before, with AI, for instance, reducing incentives for people to share knowledge with each other and with AI systems (Vincent, 2022). In enterprise settings, this can likely be addressed in part through novel credit assignment techniques and systems (e.g., Atmakuri et al., 2025).
- Modern AI models can themselves be understood as amazing new forms of collective intelligence (McMahon et al., 2017; Lanier, 2023) and in fact “collective intelligence” is perhaps a more accurate term for technologies like LLMs than “artificial intelligence” (Li et al., 2023). LLMs take knowledge from millions of people who have written web content or posted in places like Reddit and Wikipedia, interacted with chatbot systems, and generated other types of data and make that available on-demand to individuals. If the “one-person unicorn” hypothesis wins out, this would be the mechanism; these wouldn’t really be single-person companies, they would be the whole world generating value together.

Burton, J. W. et al. (2024). How large language models can reshape collective intelligence. *Nature Human Behaviour*.
Woolley, A.W. (2025). Generative AI and collaboration: opportunities for cultivating collective intelligence. *Journal of Organization Design*.
Valentine, M. and Bernstein, M. (2025). Flash Teams: Leading the Future of AI-Enhanced, On-Demand Work. *MIT Press*.
Ratcliffe, E. (2025). All of My Employees Are AI Agents, and So Are My Executives. *WIRED*.
Vincent, N. (2022). The Paradox of Reuse, Language Models Edition. *Data Leverage*.
Atmakuri, S. et al. (2025). Making AI citations count with Asta. *Ai2*.
McMahon, C. et al. (2017). The Substantial Interdependence of Wikipedia and Google – A Case Study on the Relationship Between Peer Production Communities and Information Technologies. *AAAI ICWSM*.
Lanier, J. (2023). There is No AI. *The New Yorker*.
Li, H. et al. (2023). The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. *FAccT*.

Devil’s advocate or voice equalizer? Effective AI teammates may vary across collaboration scenarios

- AI can play varied roles in teams, mirroring functional roles from classic Group Theory (Benne & Sheats, 1948), to improve collaboration depending on the scenario.
- Siemon (2022) identifies four prototypical roles that people would like for AI agents to perform for teams: Coordinator, Creator, Perfectionist, Doer.
- Chiang et al. (2024) show that introducing a Large Language Model (LLM) as a devil’s advocate in a group’s decision process significantly improves the team’s appropriate reliance on an AI decision aid. This role might also help amplify minority voices (Lee et al., 2025).
- Overall, different collaboration settings benefit from different AI behaviors, and no single AI persona fits all teams.
    - For instance, in creative collaborations like brainstorming, an AI “co-ideator” can enhance idea diversity by generating novel suggestions and helping participants refine and evaluate concepts (Shaer et al., 2024).
    - In contrast, for conflict resolution and consensus-building scenarios, an AI mediator that synthesizes diverse viewpoints into balanced group statements can foster agreement and reduce polarization (Tessler et al., 2024).



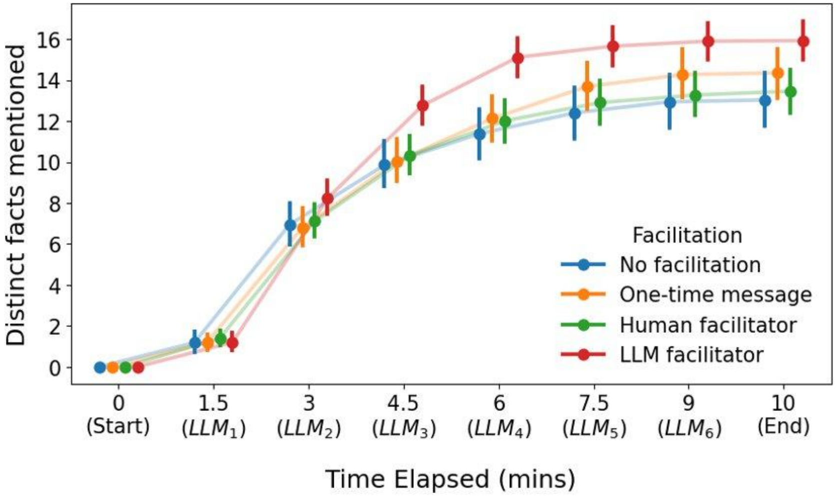
Role	Description
Coordinator	Is good at convincing and motivating team members to take action
	Can take over the leadership of a team if necessary
	Is good at assigning tasks to other team members
	Is good at discussing and arguing with other team members
	Can capture emotions and social dynamics within the team
	Is good at solving conflicts
Creator	Is good at finding many and new possible solutions for situations
	Conducts research, to develop something new based on it
	Is always looking for new ideas and developments
	Is good at uncovering novel patterns and form new associations
	Is good at innovative problem-solving
	Is good at contributing expert knowledge to a complex task
Perfectionist	Is good at completing tasks in detail
	Goes into great detail when solving a task
	Is rather a perfectionist when it comes to solving tasks
	Is good at finding optimal solutions to previously described problems depending on objective parameters
	Is good at analytical thinking and structuring
	Is good at validating if no aspects are missing
Doer	Pushes for concrete actions so that no time is wasted and can separate the important from the unimportant
	Is good at finding practical solutions that work
	Is good at completing tasks properly
	Is good at prioritizing and making decisions
	Is good at distinguishing the unimportant from important

Elaborating Team Roles for Artificial Intelligence-based Teammates in Human-AI Collaboration (Siemon, 2022).

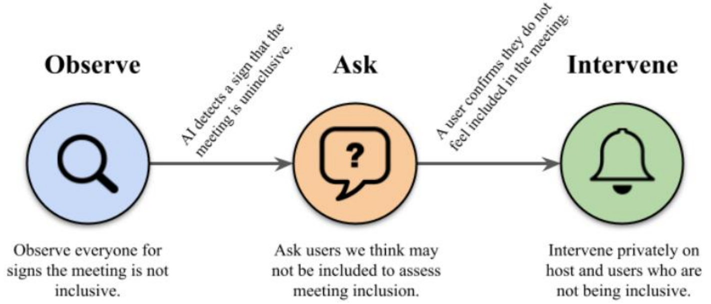
Benne, K. D. and Sheats, P. (1948). Functional Roles of Group Members. *Journal of Social Issues*.
Siemon, D. (2022). Elaborating team roles for artificial intelligence-based teammates in human-AI collaboration. *Group Decision and Negotiation*.
Chiang, C.W. et al. (2024). Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. *IUI*.
Lee, S. H. et al. (2025). Amplifying Minority Voices: AI-Mediated Devil's Advocate System for Inclusive Group Decision-Making. *IUI*.
Shaer, O. et al. (2024). AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. *CHI*
Tessler, M. H. et al. (2024). AI can help humans find common ground in democratic deliberation. *Science*.

AI facilitation can boost information-sharing and inclusion, though shaping decision outcomes remains challenging

- Through facilitation agents appearing in online meeting platforms, people are increasingly exposed to AI meeting participants that track time, summarize meetings, and more (e.g., Microsoft’s Facilitator Agent, and Zoom’s AI Companion).
- Alsobay et al. (2025) compared AI and human facilitation in a group task in which participants have different information that must be shared to reach a decision. AI facilitators boosted a measure of information sharing by 22% over no facilitation and were perceived favorably by participants. However, neither human nor AI facilitators changed group decisions, suggesting that outcome-shaping may require tools that additionally target decision-making processes (e.g. Chiang et al., 2024).
- AI agents can also support more inclusive meetings by observing interactions, asking for users’ perspectives, and intervening when appropriate. However, as with decision-making, outcome-shaping is hard, because while people prefer agents to ask before intervening, they may also rationalize away their input (Houtti et al., 2025).



Participants in meetings with an AI facilitator shared more information compared to those in other conditions. Time on the horizontal axis corresponds to scheduled LLM facilitator interventions (Alsobay et al., 2025).



Observe

AI detects a sign that the meeting is uninclusive.

Ask

A user confirms they do not feel included in the meeting.

Intervene

Observe everyone for signs the meeting is not inclusive.

Ask users we think may not be included to assess meeting inclusion.

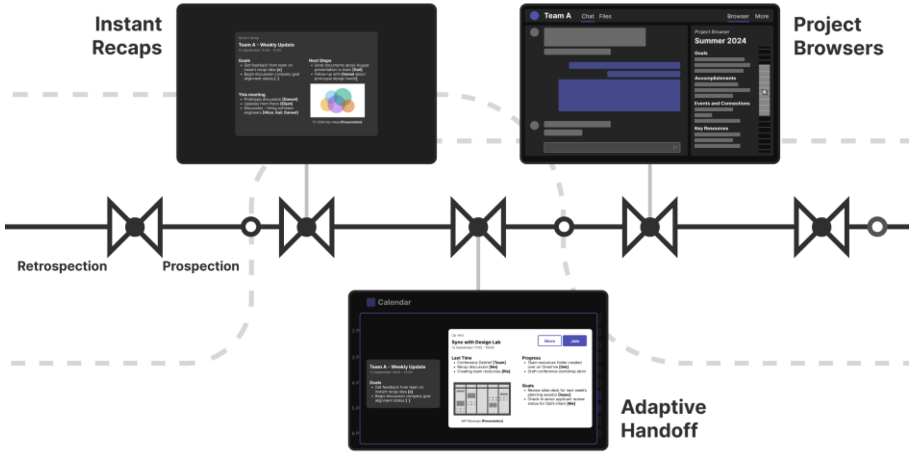
Intervene privately on host and users who are not being inclusive.

The “Observe, Ask, Intervene” framework from Houtti et al. (2025).

Microsoft Study: Alsobay, M. et al. (2025). Bringing Everyone to the Table: An Experimental Study of LLM-Facilitated Group Decision Making. *arXiv Working Paper*.
Chiang, C. et al. (2024). Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. *IUI*.
Houtti, M. et al. (2025). Observe, Ask, Intervene: Designing AI Agents for More Inclusive Meetings. *CHI*.

AI can drive goal clarity for effective meetings and smarter workflows

- As the clear expression of goals becomes increasingly vital in an AI world (Passi, 2025), meetings stand out as a key opportunity where improved goal articulation and communication can drive effective human collaboration (Scott et al., 2024).
- In a large-scale field study, Tankelevitch et al. (2025) find that workplace reflection on meeting goals can drive effective meeting behaviors across workflows, with many opportunities for AI assistance throughout the meeting life-cycle.
- Through brief reflective dialogue, AI improves meeting preparation by helping people clarify and express goals, transform them into actionable outputs, and take proactive collaborative actions to drive meeting effectiveness (Scott et al., 2025; Doherty et al., in prep.).
- AI interfaces can support tracking goals during meetings (Chen et al., 2025a; Chen et al., 2025b), as well as ensure the flow of goals between meetings and across workflows by supporting workers in prospective and retrospective thinking about goals (Vanukuru et al., 2025).
- When goals are externalized in this way, they become structured inputs for other forms of AI-driven collaboration, such as document creation or spec-driven development (GitHub, 2025).



Three example design concepts from Vanukuru et al. (2025) demonstrating how AI interfaces can support 'temporal work' across collaborative workflows: (a) **Instant Recaps** use AI to dynamically adapt their support for goal reflection based on the time people have after meetings, (b) **Adaptive Meeting Handoff** uses AI to support effective goal-driven transitions between disparate meetings, and (c) **Project Browser** allows people to explore and make sense of projects at different timescales.

Microsoft Study: Passi, S. (2025). Agentic AI has a Human Oversight Problem. SSRN Working Paper.
Microsoft Study: Scott, A. et al. (2024). Mental Models of Meeting Goals: Supporting Intentionality in Meeting Technologies. *CHI*.
Microsoft Study: Tankelevitch, L. et al. (2025). Nudging Attention to Workplace Meeting Goals: A Large-Scale, Preregistered Field Experiment. *Under Review*.
Microsoft Study: Scott, A. et al. (2025). What Does Success Look Like? Catalyzing Meeting Intentionality with AI-Assisted Prospective Reflection. *CHIWORK*.
Microsoft Study: Doherty, E. et al. (in prep.). AI-Assisted Support for Intentionality in Workplace Meetings.
Microsoft Study: Chen, X. et al. (2025a). Are We On Track? AI-Assisted Active and Passive Goal Reflection During Meetings. *CHI 2025*.
Chen, W. et al. (2025b). EchoMind: Supporting Real-time Complex Problem Discussions through Human-AI Collaborative Facilitation. *CSCW*.
Microsoft Study: Vanukuru, R. et al. (2025). Strengthening the Chain of Intentionality Across Meetings: AI-Assisted Retrospection and Prospection For Knowledge Work. *DIS*.
Microsoft Study: Github. (2025). Spec-driven development with AI: Get started with a new open source toolkit . *The GitHub Blog*.

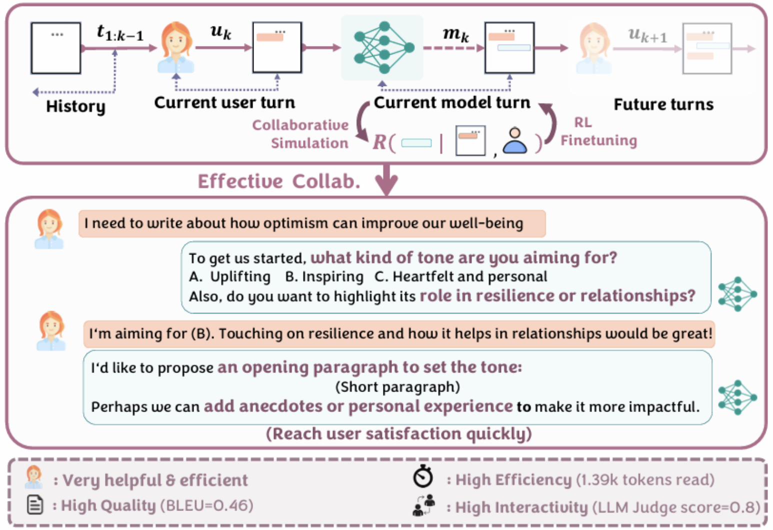
Proactivity transforms AI from a passive tool into more of a "team member"

- When a collaborative AI is developed to be more proactive, it starts to take on roles more akin to being a "team member" – raising new ideas, identifying risks, and suggesting next steps.
- Huang et al. (2025) demonstrate that proactive information gathering fundamentally changes the role of LLMs from passive responders to collaborative thought partners, improving user satisfaction by 42%. Their qualitative evaluations showed that users interpret proactive questioning as engagement and competence—traits associated with teammates rather than tools.
- Liu et al. (2025a) introduce Thoughtful AI as a paradigm shift from reactive, turn-based systems to agents that continuously think out loud and share their evolving reasoning during interactions, fostering proactive behavior and cognitively aligned collaboration. They argue this approach transforms AI from a passive tool into a thought partner, capable of initiating dialogue, suggesting next steps, and adapting as the conversation evolves.
- Commonly-used NLP techniques like next speaker prediction fall short in multi-party dynamics, where conversations are fluid and overlapping. Predicting the next speaker does not help the AI determine the optimal timing for its own intervention without disrupting the flow. It also fails to capture the underlying motivation or reasoning behind speaking turns, so the AI cannot assess whether its contribution would add value at that moment (Liu et al., 2025b).
- New techniques allow an AI to decide when to speak up and contribute during group interactions. For example, Lu et al. (2024) explore anticipatory planning and timing, showing that AI can learn when to intervene by predicting tasks and leveraging human feedback. By fine-tuning on timing decisions, the model learned conversational dynamics (e.g. pauses, topic shifts) and task progression cues to identify opportune moments for contribution.
- Drawing from cognitive psychology and real human turn-taking data, Liu et al. (2025b) show that modeling human-like turn-taking and timing through internal reasoning improves perceived intelligence and appropriateness of AI contributions.

Microsoft Study: Huang, T. et al. (2025). Teaching Language Models To Gather Information Proactively. *EMNLP*.
Liu, X. et al. (2025a). Interacting with Thoughtful AI. *arXiv Working Paper*.
Liu, X. et al. (2025b). Proactive Conversational Agents with Inner Thoughts. *CHI*.
Lu, Y. et al. (2024). Proactive Agent: Shifting LLM Agents from Reactive Responses to Active Assistance. *ICLR*.

Team AI alignment should reward deliberation and long-horizon goals

- Standard RLHF-inspired alignment techniques that are essential to the success of AI models (e.g., PPO, DPO) fail to maintain reliability in extended multi-turn (Laban et al., 2025) and multi-party dialogues (Nath et al., 2025a; 2025b) because they assume static, single-user interactions and optimize for narrow, turn-level rewards (Wu et al., 2025).
- Wu et al. (2025) highlight the limitations of traditional alignment methods that optimize for single-turn correctness and immediate user satisfaction, and propose a multi-turn reward function that evaluates responses based on their long-term impact on collaboration, boosting task success by 18.5% and user satisfaction by 17.6%.
- Nath et al. (2025a) show that standard alignment techniques lose their reliability in long multi-turn, multi-party conversations, as these methods assume a static mapping from the agent's action to the outcome and fail to account for the complexity of group conversations where each participant can alter the course of those actions.


$t_{1:k-1}$
u_k
m_k
u_{k+1}
History
Current user turn
Current model turn
Future turns
Collaborative Simulation
$R($ | , $)$
RL Finetuning
Effective Collab.
(Reach user satisfaction quickly)
: Very helpful & efficient
: High Quality (BLEU=0.46)
: High Efficiency (1.39k tokens read)
: High Interactivity (LLM Judge score=0.8)

I need to write about how optimism can improve our well-being

To get us started, what kind of tone are you aiming for?
A. Uplifting B. Inspiring C. Heartfelt and personal
Also, do you want to highlight its role in resilience or relationships?

I'm aiming for (B). Touching on resilience and how it helps in relationships would be great!

I'd like to propose an opening paragraph to set the tone:
(Short paragraph)
Perhaps we can add anecdotes or personal experience to make it more impactful.

COLLABLLM incorporates multiturn-aware rewards from collaborative simulation, enabling forward-looking strategies (Wu et al., 2025).

- Nath et al. (2025b) propose a "Frictional Agent Alignment Framework" that explicitly models conversation state and frictive conditions (Clark, 1996), accounts for group dynamics where collaborator responses can alter effects of intervention, and optimizes for collaborative processes like common-ground construction and deliberation.

Microsoft Study: Laban, P. et al (2025). LLMs Get Lost In Multi-Turn Conversation. *arXiv Working Paper*.
Nath, A. et al. (2025a). Let's Roleplay: Examining LLM Alignment in Collaborative Dialogues. *CEUR-WS*.
Nath, A. et al. (2025b). Frictional Agent Alignment Framework: Slow Down and Don't Break Things. *ACL*.
Microsoft Study: Wu, S. et al. (2025). COLLABLLM: From Passive Responders to Active Collaborators. *ICML*.
Clark, H. H. (1996). Using Language. *Cambridge University Press*.

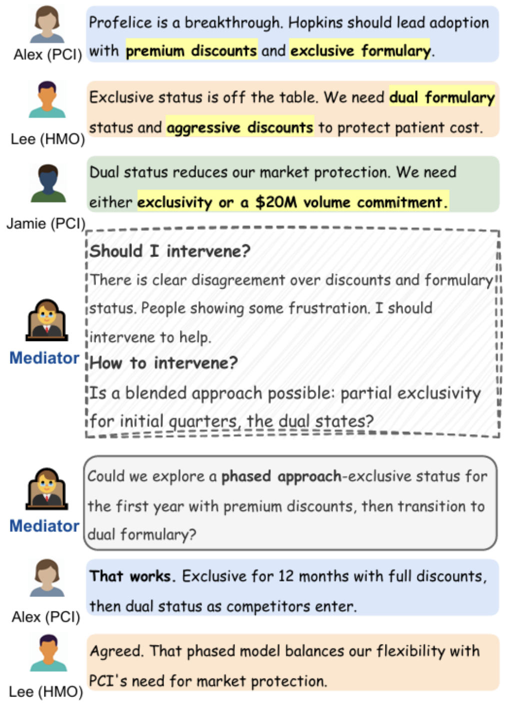
New evaluation metrics are needed for proactive agents in group settings

- Modern collaborative AI systems that actively assist or mediate human teams demand reimagined evaluation approaches. Conventional metrics for single-user, reactive AI assistants (e.g., response accuracy or user satisfaction) are insufficient – instead, evaluation must consider multi-user dynamics, social acceptance, and long-term impacts.
- Unlike reactive assistants, proactive AI systems must be evaluated on their ability to choose when and how to help, manage complex group dynamics, and adapt to context, beyond simply giving correct answers (Liu et al., 2025a; 2025b). This means tracking metrics like consensus-building, timing and frequency of interventions, fairness in turn-taking, and the agent's ability to adapt its support to the conversation's context.
- Beyond proactivity, multi-party collaborations present distinct challenges absent in individual settings: uneven participation, information asymmetry, coalition formation, social influence, groupthink, and minority voice suppression—that require new metrics for success. Recent work shows the value of capturing information-sharing quality (Alsobay et al., 2025), consensus change (Liu et al., 2025b), and participation equity (Houtti et al., 2025) in multi-person tasks.
- Collectively, the recent research calls for evaluation metrics to assess both process and outcomes, recognizing improved AI interventions may not immediately translate to better outcomes (Alsobay et al., 2025; Houtti et al., 2025); to account for temporal misalignment between immediate and delayed effects of AI interventions (Liu et al., 2025b; Nath et al., 2025); and including social perception and user acceptance as independent dimensions from technical quality (Tessler et al., 2024).
- Static benchmarks cannot capture interactive, emergent, and adaptive behaviors. Multiple surveys highlight that traditional static benchmarks fail for proactive agents in dynamic multi-party settings. The shift toward "dynamic evaluation" and "continuously updated benchmarks" reflects recognition that proactive behavior, emergent coordination, and adaptive responses to evolving group dynamics cannot be assessed through fixed test sets (Zhu et al., 2025; Yehudai et al., 2025).

Liu, X. et al. (2025a). Proactive Conversational Agents with Inner Thoughts. *CHI*.
Microsoft Study: Liu et al. (2025b). ProMediate: A Socio-cognitive framework for evaluating proactive agents in multi-party negotiation. *arXiv Working Paper*.
Microsoft Study: Alsobay, M. et al. (2025). Bringing Everyone to the Table: An Experimental Study of LLM-Facilitated Group Decision Making. *arXiv Working Paper*.
Houtti, M. et al. (2025). Observe, Ask, Intervene: Designing AI Agents for More Inclusive Meetings. *CHI*.
Nath, A. et al. (2025). Let's Roleplay: Examining LLM Alignment in Collaborative Dialogues. *CEUR-WS*.
Tessler, M. et al. (2024). AI can help humans find common ground in democratic deliberation. *Science*.
Zhu, K. et al. (2025). MultiAgentBench : Evaluating the Collaboration and Competition of LLM agents. *ACL*.
Yehudai, A. et al. (2025). Survey on Evaluation of LLM-based Agents. *arXiv Working Paper*.

Simulation frameworks can be testbeds for evaluating AI in group settings

- Modern LLMs have enabled a step-function improvement in agent-based simulation of the behavior of groups of people (Park et al., 2023). Unlike earlier rule-based bots, LLM-powered agents interpret nuanced context, maintain memory for cohesive multi-turn conversations, and exhibit closer to human-like reasoning and communication skills.
- There are now multiple LLM-driven multi-agent simulation platforms – e.g. SOTOPIA (Zhou et al., 2024), InnerThought (Liu et al., 2025a), TinyTroupe (Salem et al., 2025), ProMediate (Liu et al., 2025b) – that allow scientists and developers to create realistic virtual teams of people-simulating AI agents working and conversing together. Platforms such as ProMediate simulate multi-stakeholder negotiations with proactive mediators, enabling quantitative metrics like consensus gains, intervention timing, and socio-cognitive intelligence, while SOTOPIA puts AI agents in games or role-play scenarios to further assess teamwork, strategic planning, and communication.
- These systems provide controlled environments in which AI agents built to support teams can interact and learn in complex scenarios (negotiations, planning), offering a powerful lens to observe and test collaborative behaviors before real-world deployment. For instance, Liu et al. (2025a) show how these simulations probe nuanced social dynamics—e.g., when to interject—by modeling internal decision mechanisms that weigh relevance and information gaps, enabling evaluation of proactive participation strategies.



Alex (PCI): Profelice is a breakthrough. Hopkins should lead adoption with **premium discounts** and **exclusive formulary**.

Lee (HMO): Exclusive status is off the table. We need **dual formulary** status and **aggressive discounts** to protect patient cost.

Jamie (PCI): Dual status reduces our market protection. We need either **exclusivity or a $20M volume commitment.**

Mediator:
Should I intervene?
There is clear disagreement over discounts and formulary status. People showing some frustration. I should intervene to help.
How to intervene?
Is a blended approach possible: partial exclusivity for initial quarters, the dual states?

Mediator: Could we explore a **phased approach**-exclusive status for the first year with premium discounts, then transition to dual formulary?

Alex (PCI): **That works.** Exclusive for 12 months with full discounts, then dual status as competitors enter.

Lee (HMO): Agreed. That phased model balances our flexibility with PCI's need for market protection.

Multi-party conversation simulation with a proactive mediator agent in ProMediate (Liu et al., 2025b).

- Study findings do reveal gaps—e.g., opponent modeling and team coordination—even with top-tier models, underscoring the need for richer simulations and targeted training to advance collaborative competence (Chen et al., 2024).

Park, J. S. et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *UIST*.
Zhou, X. et al. (2024). Sotopia: Interactive evaluation for social intelligence in language agents. *ICLR*.
Liu, X. et al. (2025a). Proactive Conversational Agents with Inner Thoughts. *CHI*.
Salem, P. et al. (2025). TinyTroupe: An LLM-powered Multiagent Persona Simulation Toolkit. *arXiv Working Paper*.
Microsoft Study: Liu, Z. et al. (2025b). ProMediate: A Socio-cognitive framework for evaluating proactive agents in multi-party negotiation. *arXiv Working Paper*.
Chen, J. et al (2024). LLMArena: Assessing Capabilities of Large Language Models in Dynamic Multi-Agent Environments. *ACL*.

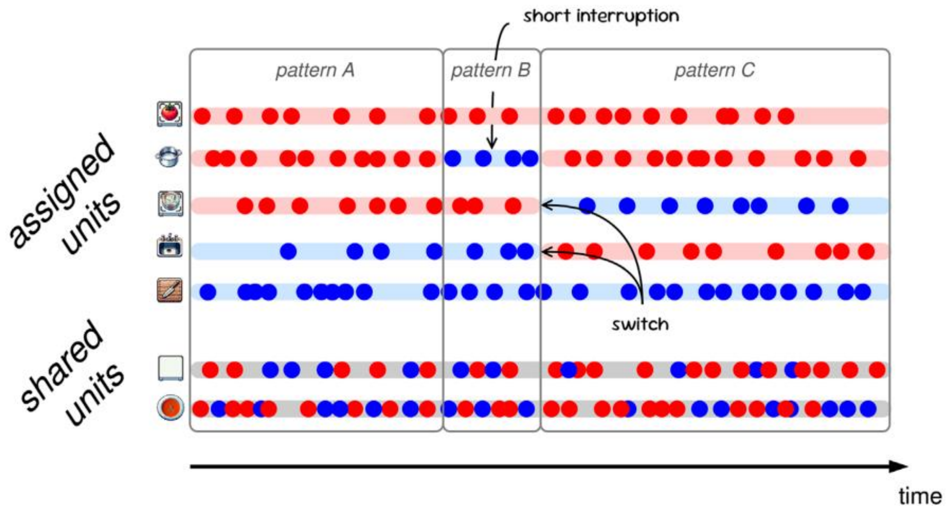
The success of AI in team settings likely depend on reinforcement learning coupled with tunable simulation frameworks

- Recent breakthroughs demonstrate a move beyond static supervised learning to using reinforcement learning (RL) and self-play as dominant methods for teaching AI teammates how to collaborate. By having LLM-powered agents simulate human colleagues, models can practice multi-turn interactions and autonomously discover effective teamwork strategies in controlled virtual environments.
- For example, researchers are using multi-agent RL and self-reflection techniques to train AI agents to collaborate effectively, by learning to coordinate better (Bo et al., 2024) or to form alliance and build trust (FAIR et al., 2022).
- Wang et al. (2024) proposed SOTOPIA-π, leveraging behavior cloning and self-play reinforcement learning to train a 7B LLM to match GPT-4-level collaboration in multi-agent tasks. Park et al. (2025) extended this paradigm with MAPoRL, co-training multiple LLMs through interactive self-play and RL rewards to enhance teamwork and generalization across complex domains.
- Abdelnabi et al. (2024) demonstrate that multi-agent negotiation simulations can expose how varying agent roles—cooperative, competitive, or adversarial—shape interaction dynamics, underscoring the need for strategic control over simulation parameters to stress-test and refine AI negotiators. Complementing this, Rao et al. (in prep.) provide empirical evidence that adjusting contextual "knobs" such as team size, diversity, and goal type systematically influences collaboration patterns (e.g., larger teams slow consensus, diverse teams increase debate). Together, these works argue that effective training environments must be steerable, enabling researchers and organizations to purposefully shape emergent behaviors and align AI teammates with desired collaboration outcomes before deployment.

Bo, X. et al. (2024). Reflective Multi-Agent Collaboration based on Large Language Models. *NeurIPS*.
Meta Fundamental AI Research Diplomacy Team (FAIR) et al. (2022). Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*.
Wang, R. et al. (2024). SOTOPIA-π: Interactive Learning of Socially Intelligent Language Agents. *ACL*.
Abdelnabi, S. et al. (2024). Cooperation, competition, and maliciousness: LLM-stakeholders interactive negotiation. *NeurIPS*.
Microsoft Study: Rao, A. et al. (in prep.). Measuring and Steering Emergent Collaboration Behaviors in LLM Agents.

Validation, norms, and visibility are likely important in human-AI teams

- *Social validation drives trust calibration:* Because AI logic is often opaque, humans struggle to verify it independently. Trust likely can be established when AI contributions are explicitly validated or cross-checked by other human team members (Cambon & Farach, 2025; Zercher et al., 2025).
- *Norms for voluntary engagement:* Establishing norms around optionality is critical. In objective tasks, making AI advice voluntary can increase acceptance. However, in high-stakes or subjective contexts, mandatory protocols may be necessary to ensure the AI is utilized (Cambon & Farach, 2025; Smith, 2025).
- *AI as an embodied or signal-based actor:* Because AI lacks physical presence, it risks being ignored. To function as a team member, AI should generate signals (through movement in virtual spaces or active communication) that allow humans to infer intent and coordinate behaviors implicitly (Cambon & Farach, 2025; Schröder et al., 2025; Smith et al., 2025).
- *Team ratio shifts the agent's role:* In dyads (1 human + 1 AI), evidence suggests humans may lack shared team identity, defaulting to a tool-user relationship. In small teams (triads), the AI can act as a creative partner, while in larger groups, it can shift to a governance layer that corrects human bias (Cambon & Farach, 2025; Georganta & Ulfert, 2024; Zercher et al., 2025).

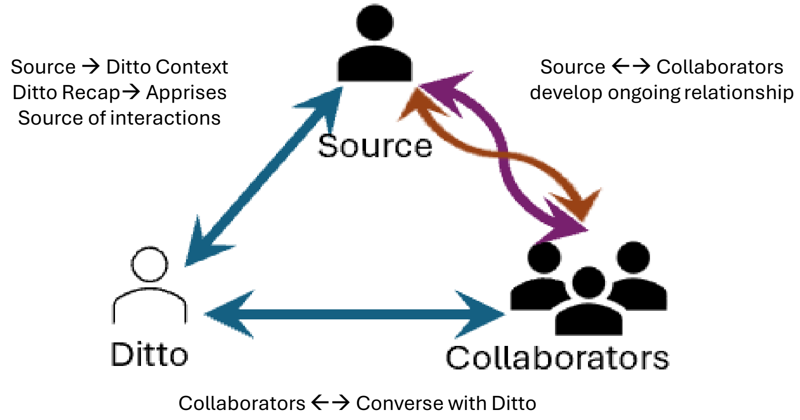


A conceptual illustration of "Fluid Collaboration" derived from the “Cooperative Cuisine” study, where partners prepare meals in a simulated kitchen. In this model, agents (represented as Red and Blue) dynamically switch roles to manage bottlenecks. This visualizes the on-the-fly adaptations observed in unpredictable environments and illustrates a standard of implicit coordination that AI should emulate to function as an effective teammate (Schröder et al., 2025).

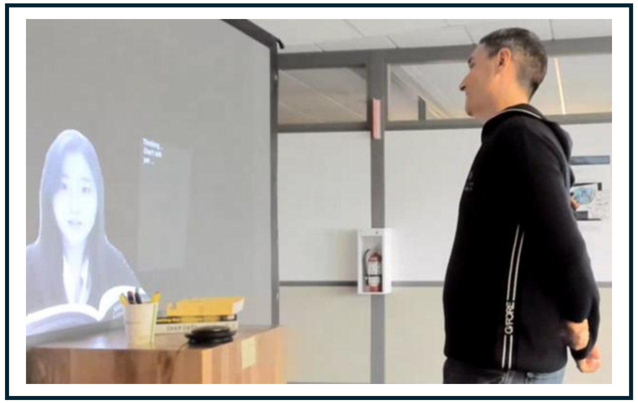
Microsoft Study: Cambon, A. and Farach, A. (2025). Human Agent Teaming: Practice Lessons from Field Experimentation. *Internal*.
Zercher, D. et al. (2025). How can teams benefit from AI team members? Exploring the effect of generative AI on decision-making processes and decision quality in team-AI collaboration. *Journal of Organizational Behavior*.
Smith, A. et al. (2025). Navigating AI convergence in human–AI teams: a signaling theory approach. *Journal of Organizational Behavior*.
Schröder, F. et al. (2025). Towards fluid human-agent collaboration: From dynamic collaboration patterns to models of theory of mind reasoning. *Frontiers in Robotics and AI*.
Georganta, E. and Ulfert, A. (2024). Would you trust an AI team member? Team trust in human-AI teams. *Journal of Occupational and Organizational Psychology*.

Dittos: How mimetic AI agents can augment relationships

- *Dittos* are embodied, mimetic, reciprocal agents that look, sound, and discuss like the person who created them (Leong et al., 2024).
- The person represented by the Ditto (the *Source*), prepares the *Context* on topics that the Ditto can talk about on behalf of the Source, and gets apprised of any interactions with their Ditto through a *Recap*. The Source and *Collaborators* can build on any Ditto conversations in developing their ongoing relationship, extending their availability.
- While Dittos were originally developed for workplace settings, they are being explored in other settings as well:
  - FamilyDittos: Connecting with remote family members, especially separated by inconvenient time zones. (Tanprasert et al., 2025).
  - IRL Dittos: Connecting with remote colleagues by encountering their Dittos in shared, public spaces (like a hallway) (Lee et al., 2025).
- There are many questions about the responsible design of Dittos. Because a Ditto seems like someone people know, people believe Dittos more than GenAI. Researchers conducted Black Mirror Writers' Room exercises with the project team to proactively identify responsible design implications (Lee et al., 2023; Klassen & Fiesler, 2022).



Complete communication cycle with Dittos, including the Source preparing the context for the Ditto, Collaborators interacting with the Ditto, and the Source getting a Recap of all interactions with the Ditto

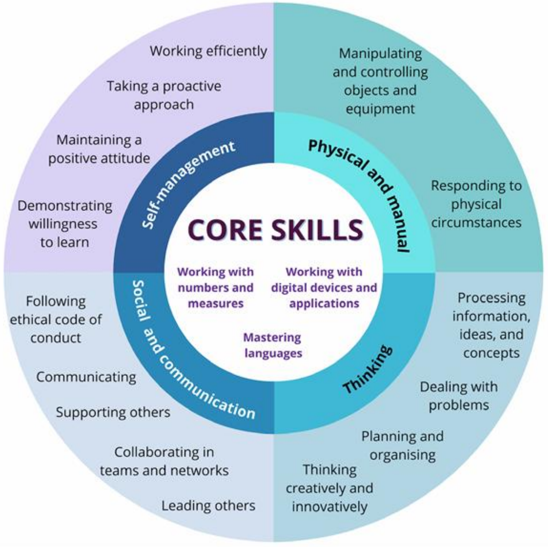


Collaborator talking with a Ditto in the hallway

Microsoft Study: Leong, J. et al. (2024). Dittos: Personalized, Embodied Agents That Participate in Meetings When You Are Unavailable. *CSCW.*
Microsoft Study: Tanprasert, T. et al. (2025). FamilyDittos: Reimagining Intergenerational Interaction through Mimetic Agents, *CSCW.*
Microsoft Study: Lee, S. et al. (2025). IRL Dittos: Embodied Multimodal AI Agent Interactions in Open Spaces. *arXiv Working Paper.*
Lee, P. et al. (2023). Speculating on Risks of AI Clones to Selfhood and Relationships: Doppelganger-phobia, Identity Fragmentation, and Living Memories. CSCW.
Klassen, S. and Fiesler, C. (2022). "Run Wild a Little With Your Imagination": Ethical Speculation in Computing Education with Black Mirror. *SIGCSE.*

AI shifts effort from ‘doing’ to ‘choosing’, which risks cognitive deskilling unless usage is accompanied by upskilling and reskilling

- The generative AI shift moves the locus of work effort from ‘thinking by doing’ to ‘choosing from outputs’, potentially reducing the judgment needed to develop and maintain skills and expertise (Sarkar, 2024; Macnamara et al., 2024).
- Cognitive skills at risk range from *foundational* metacognitive skills such as planning and confidence calibration (Tankelevitch et al., 2024), key *transferable work skills* (Morandini et al., 2025), and *specific professional skills* (e.g. Accounting – Eisikovits et al., 2025; Law - Gomez Schieber et al. 2025; Medicine - Natali et al., 2025, and Programming - Le, 2025).
- Concerns about deskilling, tied to changes in distributed cognition between people and technology, have consistently accompanied the introduction of new technologies and sparked debates about their advantages and disadvantages (Crowston & Bolici, 2025). Grinschgl & Neubauer (2022) note that skills are not always maximally offloaded, rather people make situational cost-benefit considerations.
- Akin to the desktop computer revolution of the 1980s, AI will require holistic reskilling and upskilling through education/training (Ersanlı et al., 2025) and features designed into models and interfaces (Crowston & Bolici, 2025).

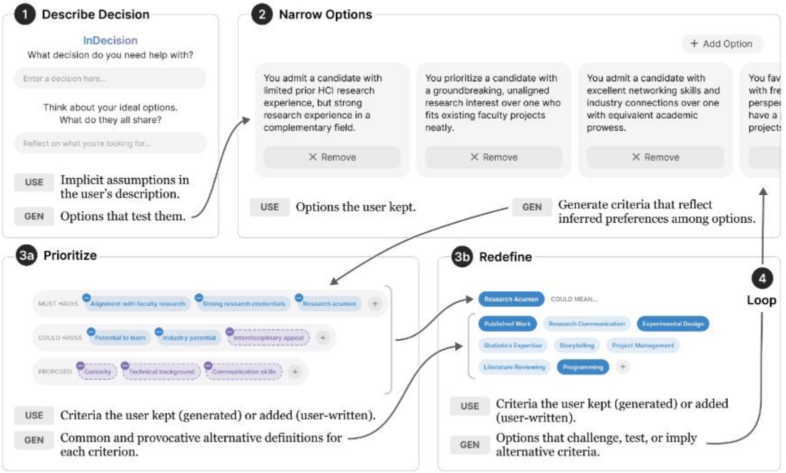

CORE SKILLS
Working with numbers and measures
Working with digital devices and applications
Mastering languages
Self-management
Physical and manual
Thinking
Social and communication
Working efficiently
Taking a proactive approach
Maintaining a positive attitude
Demonstrating willingness to learn
Manipulating and controlling objects and equipment
Responding to physical circumstances
Processing information, ideas, and concepts
Dealing with problems
Planning and organising
Thinking creatively and innovatively
Leading others
Collaborating in teams and networks
Supporting others
Communicating
Following ethical code of conduct

Transversal Skills and Competences model (Adapted from Morancini et al., 2023).

Microsoft Study: Sarkar, A. (2025). AI Should Challenge, Not Obey. *CACM*.
Macnamara, B.N. et al. (2024). Does Using Artificial Intelligence Assistance Accelerate Skill Decay And Hinder Skill Development Without Performers’ Awareness? *Cognitive Research: Principles and Implication*.
Microsoft Study: Tankelevitch, L. et al. (2024). The Metacognitive Demands And Opportunities Of Generative AI. *CHI*.
Morandini, S. et al. (2023). The Impact Of Artificial Intelligence On Workers’ Skills: Upskilling And Reskilling In Organisations. *Informing Science*.
Eisikovits, N. et al. (2025). Should Accountants Be Afraid Of AI? Risks And Opportunities Of Incorporating Artificial Intelligence Into Accounting And Auditing. *Accounting Horizons*.
Gomez Schieber, E. A. et al. (2025). Attorneys And AI: How Lawyers Use Artificial Intelligence And Analyze Its Impacts. *CSCW*.
Natali, C. et al. (2025). Ai-induced Deskilling In Medicine: A Mixed-method Review And Research Agenda For Healthcare And Beyond. *Artificial Intelligence Review*.
Le, H. (2025). The Deskilling Of Software Development And The Impact Of AI Chatbots On Programmers' Skills And Roles. *Generative AI in Software Engineering*.
Crowston, K. and Bolici, F. (2025). Deskilling And Upskilling With AI Systems. *Information research*.
Grinschgl, S. and Neubauer, A. C. (2023). Supporting Cognition With Modern Technology: Distributed Cognition Today And In An Ai-enhanced Future. *Frontiers*.
Ersanlı, C. Y. et al. (2025). A Review Of Global Reskilling And Upskilling Initiatives In The Age Of AI. *AI and Ethics*.

There are diverse paths for AI to support thinking, rather than substitute for it

- Tankelevitch et al. (2025) synthesize design directions for augmenting thinking with AI: provoke reflection, scaffold sensemaking, and preserve user agency and cognitive engagement.
- Castañeda et al. (2025) propose helping users 'decide how to decide' (see figure). Gmeiner et al. (2025) found that AI nudges to clarify goals, reflect on decisions, and evaluate outcomes helped designers plan and better align outputs with original intent.
- Sarkar (2025) demonstrates a prototype workflow in which direct material manipulation of AI views and AI provocations aid strategic reading, planning, and decision-making by looking at critiques, alternatives, and lateral moves.
- Kang et al. (2025) and Yang et al. (2025) show that when AI helps with reasoning by analogy, people improve in finding concepts that feel unfamiliar yet feasible, overlooked, or unexpected.
- Wang and Chilton (2025) show how AI can support users in identifying design patterns from examples, whereas Dang et al. (2025) show how AI can support users in understanding and applying implicit norms in writing.



USE Implicit assumptions in the user's description.
GEN Options that test them.

You admit a candidate with limited prior HCI research experience, but strong research experience in a complementary field.

You prioritize a candidate with a groundbreaking, unaligned research interest over one who fits existing faculty projects neatly.

You admit a candidate with excellent networking skills and industry connections over one with equivalent academic prowess.

You fav with fre perspe have a projects

USE Options the user kept.
GEN Generate criteria that reflect inferred preferences among options.

USE Criteria the user kept (generated) or added (user-written).
GEN Common and provocative alternative definitions for each criterion.

USE Criteria the user kept (generated) or added (user-written).
GEN Options that challenge, test, or imply alternative criteria.

InDecision's iterative loop. The initial elicitation (1) allows the user to provide open-text descriptions of their decision and relevant considerations for options and criteria. The user is presented with a list of eight options (2). The user can keep, add to, or remove these options. To promote reflection on what is most important to them, the user may continue only after narrowing down to three. Criteria refinement is composed of two stages: (3a) prioritization, where a user can add, remove, and sort criteria in tiers of priority; and (3b) redefinition, where the user selects between a range of possible meanings associated with each criterion. These steps are repeated in an iterative loop (4) (Castañeda et al., 2025).

Microsoft Study: Tankelevitch, L. et al. (2025). Understanding, Protecting, And Augmenting Human Cognition With Generative AI: A Synthesis Of The CHI 2025 Tools For Thought Workshop. *arXiv Working Paper.*
Castañeda, C. et al. (2025). Supporting AI-Augmented Meta-Decision Making With InDecision. *arXiv Working Paper.*
Gmeiner, M. et al. (2025). Exploring The Potential Of Metacognitive Support Agents For Human-AI Co-Creation. *DIS.*
Microsoft Study: Sarkar, A. (2025). Artificial Intelligence As A Tool For Thought. *TEDAI Vienna.*
Kang, H. B. et al. (2025). BioSpark: Beyond Analogical Inspiration To LLM-augmented Transfer. *CHI.*
Yang, Y. et al. (2025). From Overload To Insight: Scaffolding Creative Ideation Through Structuring Inspiration. *arXiv Working Paper.*
Wang, S. and Chilton, L. B. (2025). Schemex: Discovering Design Patterns From Examples through Iterative Abstraction And Refinement. *arXiv Working Paper.*
Dang, H. et al. (2025). CorpusStudio: Surfacing Emergent Patterns In A Corpus Of Prior Work While Writing. *CHI.*

How to make AI an effective tool in the classroom

- The ease and speed of AI tools might be valuable in the workplace, but learning requires “desirable difficulties” (Walker & Vorvoreanu, 2025).
- When AI is used for summaries and syntheses, learning may be shallower (Melumad et al., 2025; Stadler et al., 2024). However, thoughtful implementation can improve learning. Kestin et al. (2025) found that an AI tutor using pedagogical best practices and measures to reduce confabulation helped students learn more, faster, than an active-learning classroom experience, as reflected in test scores.
- AI can help with self-paced, personalized learning, but it can also lead to over-estimation of one’s skills. Students need help calibrating their mental models of learning gains with AI (Simkute et al., 2024; Urban et al., 2024).
- Equity and the digital divide are still a concern. While AI can help students with disabilities, studies show that learning benefits from AI thus far favor students of higher socioeconomic status (DeSimone et al., 2025; Yu et al. 2024; Zhang et al., 2024).

Four considerations for effective AI integration into the classroom

1. Ensure student readiness. Students need enough domain knowledge to be able to evaluate AI outputs.
2. Teach AI literacy. Help students understand the basics of AI, including that it makes mistakes.
3. Use AI as a supplement. Teacher guidance and human connection are central to learning. AI cannot replace that, but it can supplement with personalized explanations and examples.
4. Foster cognitive engagement with AI. Encourage AI usage that helps students evaluate their skills and think critically.

Effective AI integration into the classroom will require training programs to manage attitudes towards use as will as new specific skills (Adapted from Walker & Vorvoreanu, 2025).

Microsoft Study: Walker, K. and Vorvoreanu, M. (2025). Learning Outcomes With GenAI In The Classroom: A Review Of Empirical Evidence. *Microsoft*.
Melumad, S. et al. (2025). Experimental Evidence Of The Effects Of Large Language Models Versus Web Search On Depth Of Learning. *PNAS NEXUS*.
Stadler, M. et al. (2024). Cognitive Ease At A Cost: : Llms Reduce Mental Effort But Compromise Depth In Student Scientific Inquiry. *Computers in Human Behavior*.
Kestin, G. et al. (2025). AI Tutoring Outperforms In-class Active Learning: An RCT Introducing A Novel Research-based Design In An Authentic Educational Setting. *Nature Scientific Reports*.
Microsoft Study: Simkute, A. et al. (2024). The New Calculator? Practices, Norms, And Implications Of Generative AI In Higher Education. *arXiv Working Paper*.
Urban, M. et al. (2024). Chatgpt Improves Creative Problem-Solving Performance In University Students: An Experimental Study. *Computers & Education*.
DeSimone, M. et al. (2025). From chalkboards to chatbots: Evaluating The Impact Of Generative AI On Learning Outcomes In Nigeria. *World Bank Group*.
Yu, R. et al. (2024). Whose Chatgpt? Unveiling Real-World Educational Inequalities Introduced By Large Language Models. *arXiv Working Paper*.
Zhang, C. et al. (2024). College Students’ Literacy, Chatgpt Activities, Educational Outcomes, And Trust From A Digital Divide Perspective. *New Media & Society*.

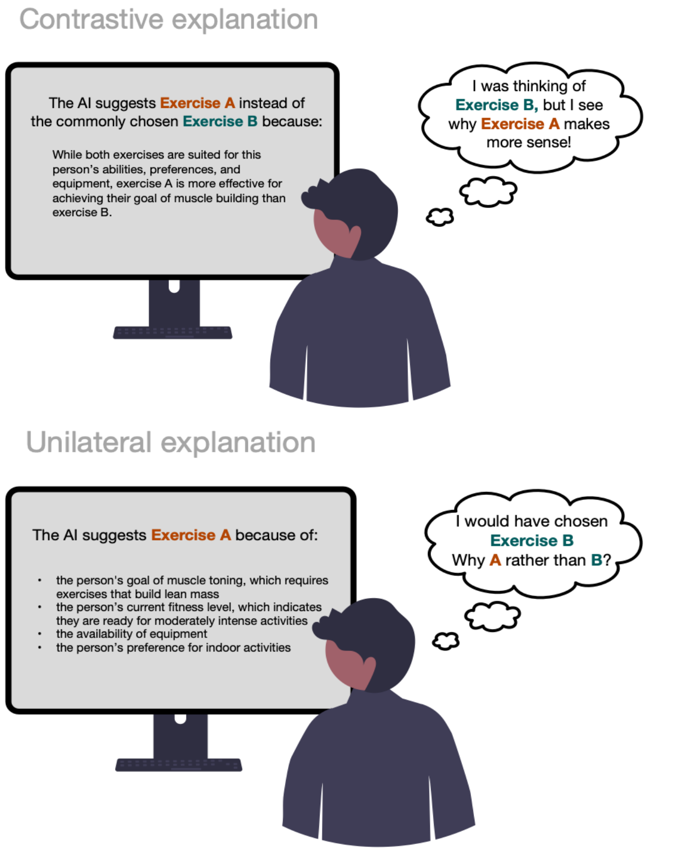
For ideation, AI has persistent risks and fragile benefits without careful use

- Designing AI-driven creativity support tools requires carefully considering the creative process and its various stages, as well as how such tools can impact creative thinking in the long term (Tankelevitch et al., 2025).
- Qin et al.'s (2025) lab study found that using LLMs before independent ideation reduced the number of original ideas and lowered creative self-efficacy, compared to use after independent ideation.
- Kumar et al.'s (2025) lab study found that those doing unassisted ideation with no exposure to LLMs had the best outcomes. Those using LLM assistance got a performance boost by cognitively offloading the work to the AI, not from collaboratively producing ideas. Individuals exposed to LLM strategies or guidance performed worse in later unassisted rounds. Overall, LLM-generated strategies reduced idea diversity, and effects continued even after AI use stopped.
- LLM support did not impact the number of ideas produced in Bangerl et al.'s (2025) lab study of group ideation, but it did significantly reduce group elaboration of those ideas. This is again possibly due to increased cognitive offloading to the AI. Similarly to Kumar et al.'s (2025) study, this effect carried over even when not using LLMs.
- One key design task is problem reframing (exploring what a problem is about). In Shin et al.'s (2025) lab study, using LLMs did not improve the usefulness of problem frames, (regardless of whether LLMs directly provided frames, were used in a free-form manner, or used in a structured process). Expert designers produced somewhat more novel frames than novice designers, but these more novel frames were not also more useful.
- A systematic review (Heigl, 2025) finds that longitudinal and field studies of AI and creativity are still rare and will be needed to understand and plan for the complex issues arising from AI in creative work.

Microsoft Study: Tankelevitch, L. et al. (2025). Understanding, Protecting, And Augmenting Human Cognition With Generative AI: A Synthesis Of The CHI 2025 Tools For Thought Workshop. *arXiv Working Paper*.
Qin, P. et al. (2025). Timing Matters: How Using LLMs At Different Timings Influences Writers' Perceptions And Ideation Outcomes In Ai-assisted Ideation. *CHI*.
Kumar, H. et al. (2025). Human Creativity In The Age Of LLMs: Randomized Experiments On Divergent And Convergent Thinking. *CHI*.
Bangerl, M. M. et al. (2025). CreAItive Collaboration? Users' Misjudgment Of Ai-creativity Affects Their Collaborative Performance. *CHI*.
Shin, J. et al. (2025). No Evidence For LLMs Being Useful In Problem Reframing. *CHI*.
Heigl, R. (2025). Generative Artificial Intelligence In Creative Contexts: A Systematic Review And Future Research Agenda. *Management Review Quarterly*.

AI decision-support should enhance both accuracy and human skills

- Human expertise enables critical use of AI advice, but growing reliance risks eroding people's independent abilities and the capacity to judge AI outputs (Macnamara et al., 2024). There is real-world evidence of AI-induced deskilling: clinicians who relied on an AI system for polyp detection during colonoscopies showed a significant decline in their independent ability to identify precancerous lesions after just three months of AI use (Budzyn et al., 2025).
- A nascent body of research in AI-assisted decision-making is proposing design interventions that enhance users' independent skills along with their immediate decision accuracy.
- Gajos & Mamykina (2022) found that providing users with AI explanations without explicit AI decision recommendations improved learning outcomes along with decision accuracy.
- Bucinca et al. (2025) showed that even when AI recommendations are present, contrastive explanations (which explain human knowledge gaps by highlighting differences between the AI's choice and a likely human choice about the same task) can enhance users' independent decision-making skills without sacrificing their accuracy.
- Buijsman et al. (2025) argue that decision-support must also enable domain-specific autonomy, with one approach being the use of *defeaters*—information that signals when AI outputs should be doubted or reconsidered.



Contrastive explanation

The AI suggests **Exercise A** instead of the commonly chosen **Exercise B** because:

While both exercises are suited for this person's abilities, preferences, and equipment, exercise A is more effective for achieving their goal of muscle building than exercise B.

Unilateral explanation

The AI suggests **Exercise A** because of:

- the person's goal of muscle toning, which requires exercises that build lean mass
- the person's current fitness level, which indicates they are ready for moderately intense activities
- the availability of equipment
- the person's preference for indoor activities

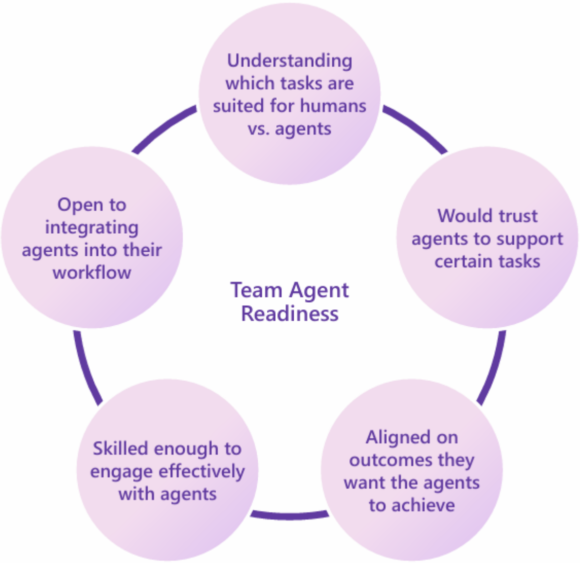
I was thinking of Exercise B, but I see why Exercise A makes more sense!
I would have chosen Exercise B Why A rather than B?

Contrastive AI explanations (top) address human knowledge gaps better than unilateral AI explanations (bottom) (Adapted from Bucinca et al., 2025).

Macnamara, B. N. et al. (2024). Does Using Artificial Intelligence Assistance Accelerate Skill Decay And Hinder Skill Development Without Performers' Awareness? *Cognitive Research: Principles and Implications.*
Budzyń, K. et al. (2025). Endoscopist Deskilling Risk After Exposure To Artificial Intelligence In Colonoscopy: A Multicentre, Observational Study. *The Lancet Gastroenterology & Hepatology.*
Gajos, K. Z. and Mamykina, L. (2022). Do People Engage Cognitively With AI? Impact Of AI Assistance On Incidental Learning. *IUI.*
Microsoft Study: Bucinca et al. (2025). Contrastive Explanations That Anticipate Human Misconceptions Can Improve Human Decision-Making Skills. *CHI.*
Buijsman, S. et al. (2025). Autonomy By Design: Preserving Human Autonomy In AI Decision-Support | Philosophy & Technology. *Philosophy & Technology.*

Workplace learning can combat AI overreliance and cognitive atrophy

- Blind trust (Benzing et al., 2025), overreliance (Passi et al., 2024), and cognitive offloading (Gerlich, 2025) are risks of human-AI interactions. Benzing et al.'s (2025) survey of 1,800 global full-time English-speaking employees found that 2 in 3 employees report trusting AI agents; however, 60% skip accuracy checks in AI output.
- Workplace learning plays a crucial role in combatting these risks, while also driving the employee engagement and resilience required for successful AI workforce transformation (Malik & Garg, 2017). Teams with strong AI readiness (including skills and literacy) report greater individual and collective value from AI, such as higher productivity, better decision-making, and more creative, collaborative teams (Benzing et al., 2025; Xue & Song, 2025).
- Human-AI interactions that are discursive and progressive promote cognitive enrichment and combat cognitive atrophy (Danry et al., 2023), which are the core elements of traditional Socratic learning methods used to develop metacognition and critical thinking.
- Given its ability to provide personalized, adaptive, role- and task-based learning while incorporating context, memory, and conversational turns, AI presents a promising new learning modality (Joshi, 2025; Tomitsu, 2025). There is opportunity to develop novel interactive learning experiences, like assessment-based dialogic questioning (Hung et al., 2024), to drive employee engagement through workplace upskilling.

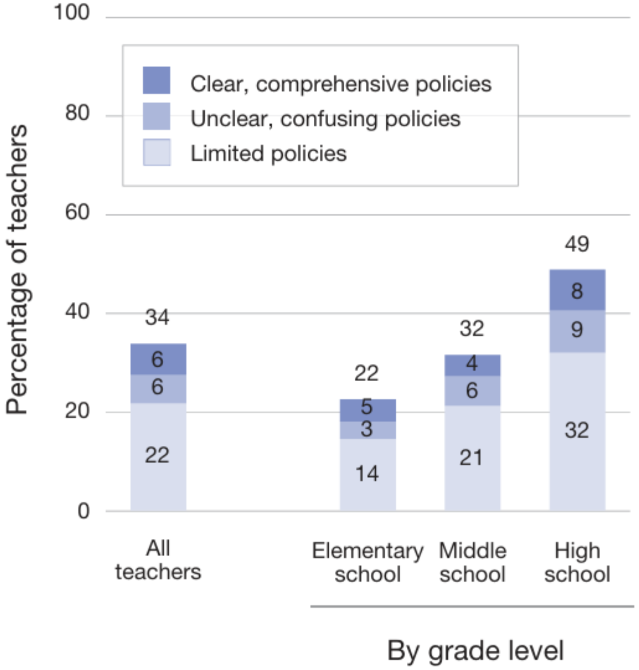

Understanding which tasks are suited for humans vs. agents
Would trust agents to support certain tasks
Aligned on outcomes they want the agents to achieve
Skilled enough to engage effectively with agents
Open to integrating agents into their workflow
Team Agent Readiness

Workplace learning and experimentation create a cycle for Agentic AI adoption (Adapted from Benzing et al., 2025).

Microsoft Study: Benzing, M. et al. (2025). 2025 Agentic Teaming & Trust Report. *Microsoft*.
Microsoft Study: Passi, S. et al. (2024). Appropriate Reliance On Generative AI: Research Synthesis. *Microsoft*.
Gerlich, M. (2025). AI Tools In Society: Implications Of Cognitive Offloading And The Future Of Critical Thinking. *Societies*.
Malik, P. and Garg, P. (2017). Learning Organization And Work Engagement: The Mediating Role Of Employee Resilience. *The International Journal of Human Resource Management*.
Xue, S. and Song, Y. (2025) Unlocking The Synergy: Increasing Productivity Through Human-ai Collaboration In The Industry 5.0 Era. *Computers & Industrial Engineering*.
Danry, V. et al. (2023). Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy Over Causal AI Explanations. *CHI*.
Joshi, S. (2025). Agentic Generative AI And The Future US Workforce: Advancing Innovation And National Competitiveness. *International Journal of Research and Review*.
Tomitsu, H. et al. (2025). The Cognitive Mirror: A Framework For AI-powered Metacognition And Self-regulated Learning. *Frontiers in Education*.
Hung, J. et al. (2024). Socratic Mind: Scalable Oral Assessment Powered By AI. *Learning@ Scale*.

Students and teachers rely on generic AI tools but usage guidance lags behind

- There has been a steady year-over-year increase in AI tool usage among students and educators in both K–12 and higher education.
- An estimated 80% of K-12 teachers and 95% of higher education educators have used AI for school-related purposes at least once, while 19% and 60% report using it regularly. With respect to students, an estimated 90% in K-12 and 95% in higher education have used AI for school-related purposes at least once, while 25% and 36% report using it regularly (Microsoft Education, 2025).
- Most current usage relies on freely available general purpose AI tools (e.g., Claude, ChatGPT, Copilot, Gemini). Students are more likely to use these general purpose versions than education specific tools (Gillespie, 2025).
- More generally, guidance and policy are lagging adoption: roughly half of surveyed school districts reported providing training to teachers on how to use AI, with another quarter planning to do so in the coming year (Doss et al., 2025).
- Expectations point to increased AI use for school-related purposes, but there are continued concerns about academic dishonesty, overreliance, and misinformation among both teachers and students (Microsoft Education, 2025).
- New programs are providing external funding to train teachers for AI use, including a National Academy for AI Instruction by the American Federation of Teachers, jointly funded by Microsoft, OpenAI, and Anthropic.



Teacher reports of US school or district policies or guidance about the use of AI related to academic integrity (Adapted from Doss et al., 2025).

Microsoft Study: Microsoft Education / IDC, 2025. 2025 AI In Education Report Survey Details. *Microsoft.*
Gillespie, N. et al. (2025). Trust, Attitudes And Use Of Artificial Intelligence: A Global Study 2025. *University of Melbourne/KPMG.*
Doss, C. J. et al. (2025). AI Use In Schools Is Quickly Increasing But Guidance Lags Behind. *RAND.*

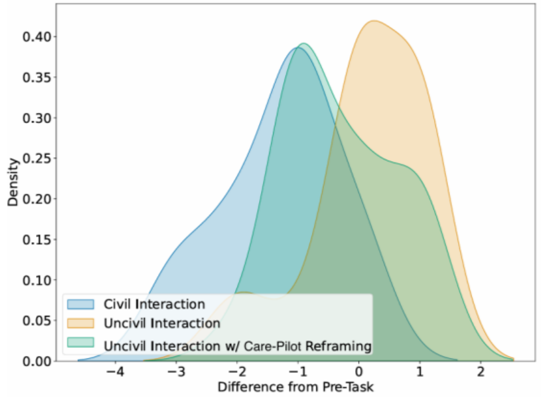
Preparing young learners for an AI future requires a further shift away from syntax specifics and towards problem solving and logical abstraction

- Coding education remains essential but should continue to evolve from memorizing syntax to developing computational thinking, problem-solving, and human–AI collaboration skills (OECD, 2025). Students need to learn to frame problems, design solutions, and critically evaluate generated code rather than simply write code manually.
- Generative AI alters the purpose of learning to code, making it necessary for students to understand how to manage, adapt, and govern AI-produced outputs responsibly (UNESCO, 2023). Educators should teach how to review, debug, and improve AI-generated solutions, positioning students as accountable decision-makers rather than passive recipients of outputs.
- Developing AI fluency from an early age is central to maintaining agency and preventing future digital exclusion. Younger learners should learn to craft effective prompts, interrogate model reasoning, identify and correct AI errors, and maintain human oversight. This supports equity by ensuring all young people—not only those with advanced technical access—gain the knowledge needed to understand, interact with, and shape AI systems. Coding remains an important tool for meaningful participation in the digital world (UNICEF, 2021; OECD, 2023).
- Consequently, coding and technology curricula should continue to emphasize foundational skills—problem decomposition, logical reasoning, debugging, and verification—but with a stronger shift toward higher-level abstraction and AI-aware practice. Instruction should prioritize structuring problems, interrogating AI outputs, and applying prompt engineering as extensions of computational thinking (OECD, 2025; UNESCO, 2023). This mirrors earlier transitions from machine-level to high-level programming. Pedagogy should use discursive, iterative, cognitively enriching methods (Zhao et al., 2025) to promote critical engagement and prevent overreliance on AI (Fan et al., 2025; OECD, 2025).

OECD (2025). What Should Teachers Teach And Students Learn In A Future Of Powerful AI?. OECD.
UNESCO (2023). Guidance for Generative AI in Education and Research. *UNESCO.*
UNICEF (2021). Policy Guidance On AI For Children: Version 2.0 | Recommendations For Building AI Policies And Systems That Uphold Child Rights. *UNICEF.*
OECD (2023). Empowering Young Children In The Digital Age. *OECD.*
Zhao, G. et al. (2025). A Generative Artificial Intelligence (AI)-Based Human-Computer Collaborative Programming Learning Method. *Journal of Educational Computing Research.*
Fan, G. et al. (2025). The Impact Of AI-assisted Pair Programming On Student Motivation, Programming Anxiety, Collaborative Learning, And Programming Performance. *International Journal of STEM Education.*

Digital empathy is emerging as a key differentiator in the design of AI conversational agents

- The role of digital empathy, understood as the ability of AI models to interpret users' experiences and perspectives, is becoming increasingly studied and formalized (Suh et al., 2025), and it is materializing in multiple forms, from emotional-regulation support (Das Swain et al., 2025) to adaptive embodied conversational agents.
- As traditional benchmarks become saturated, companies are viewing AI personality, communication style, and appearance as key differentiators and integral parts of their model optimization pipelines. Notable examples include the communication styles of recent OpenAI GPT models, as well as text to speech and agents by Sesame AI, Hume AI, and ElevenLabs.
- Therapy and companionship are emerging as a rapidly growing application area in which digital empathy can play an important role (Sanders, 2025). While AI chatbots offer the potential to broaden access to care, there remains significant room for improvement in how they recognize and respond to mental health issues. Recent research is highlighting the potential risks and harms associated with using chatbots for emotional support (Chandra et al., 2025).
- There is growing debate about human empathy versus AI empathy and their respective roles in future job automation in domains such as customer service and healthcare (Howcroft et al., 2025; Rubin et al., 2025). Emerging research also highlights the potential benefits of combining both forms of empathy, for example by reducing cognitive load while addressing uncivil customer interactions (Das Swain et al., 2025).

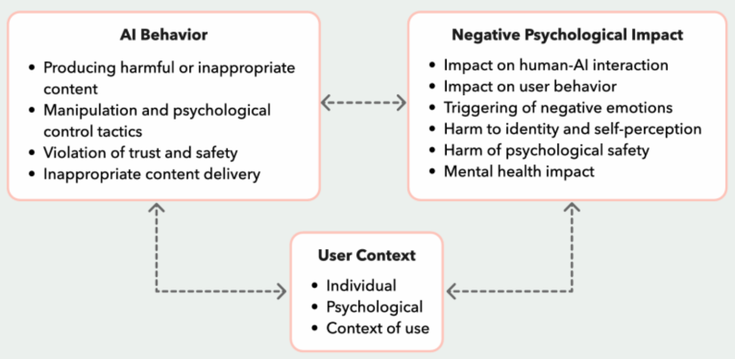


Uncivil interactions—rude, aggressive, emotionally charged—raised cognitive load (orange). Empathetic AI support eased the strain (green), nearly matching civil exchanges (blue) (Das Swain et al., 2025).

Microsoft Study: Suh, J. et al. (2025). SENSE-7: Taxonomy And Dataset for Measuring User Perceptions Of Empathy In Sustained Human AI Conversations. *arXiv Working Paper.*
Microsoft Study: Das Swain, V. et al. (2025). AI On My Shoulder: Supporting Emotional Labor In Front-Office Roles With An LLM-Based Empathetic Coworker. *CHI.*
Sanders, M. Z. (2025). How People Are Really Using Gen AI In 2025. *Harvard Business Review.*
Microsoft Study: Chandra, M. et al. (2025). Longitudinal Study On Social and Emotional Use Of AI Conversational Agent. *arXiv Working Paper.*
Howcroft, A. et al. (2025). AI Chatbots Versus Human Healthcare Professionals: A Systematic Review And Meta Analysis Of Empathy In Patient Care. *British Medical Bulletin.*
Rubin, M., et al. (2025). Comparing The Value Of Perceived Human Versus AI-Generated Empathy. *Nature Human Behaviour.*

Growing evidence suggests psychological well-being should be a core AI design and governance criterion

- Psychological well-being is the state of feeling mentally healthy, resilient, and fulfilled through positive emotions, life satisfaction, and purpose (Yiğit & Çakmak, 2024).
- Whether explicitly designed to be companions or not, users are appropriating conversational AI tools for social and emotional purposes (Laestadius et al., 2022; Siddals et al., 2024). These systems can fill emotional, social, and informational gaps, providing 24/7 availability and perceived non-judgmental communication that some associate with temporary improvement in well-being (Zhang et al., 2025).
- However, reliance on AI for companionship and unconditional validation may have psychological risks, such as dependence, distorted self-perception, and reinforcement of maladaptive behaviors or beliefs, and may reshape social norms around relational technology (Huang et al., 2024; Marriott & Pitardi, 2024).
- These impacts likely emerge from complex, long-term interactions between often unobservable factors (e.g., user's disposition, environments) and AI design, making it difficult to isolate any one technology feature as the cause for psychological harm or benefit (Fang et al., 2024; Chandra et al., 2024).



AI Behavior

- Producing harmful or inappropriate content
- Manipulation and psychological control tactics
- Violation of trust and safety
- Inappropriate content delivery

Negative Psychological Impact

- Impact on human-AI interaction
- Impact on user behavior
- Triggering of negative emotions
- Harm to identity and self-perception
- Harm of psychological safety
- Mental health impact

User Context

- Individual
- Psychological
- Context of use

Psychological impact of AI on human is highly context- and interaction-dependent. Conceptualizing psychological risks must take the full dynamic user context into account (Chandra et al., 2025).

- This research raises questions about how AI designers might establish ethical norms that preserve human well-being, invest in community-driven evaluation, conduct longitudinal monitoring, and include guardrails and supportive prosocial user experiences.

Yiğit, B. and Çakmak, B.Y. (2024). Discovering Psychological Well-Being: A Bibliometric Review. *Journal of Happiness Studies*.
Laestadius, L. et al. (2024). Too Human And Not Human Enough: A Grounded Theory Analysis Of Mental Health Harms From Emotional Dependence On The Social Chatbot Replika. *New Media & Society*.
Siddals, S., et al. (2024). It Happened To Be The Perfect Thing": Experiences Of Generative AI Chatbots For Mental Health. *npj Mental Health Research*.
Zhang, R. et al. (2025). The Dark Side Of AI Companionship: A Taxonomy Of Harmful Algorithmic Behaviors In Human-AI Relationships. *CHI*.
Huang, S. et al. (2024). AI Technology Panic—Is AI Dependence Bad For Mental Health? A Cross-lagged Panel Model And The Mediating Roles Of Motivations For AI Use Among Adolescents. *Psychology Research and Behavior Management*.
Marriott, H.R. and Pitardi, V. (2024). One Is The Loneliest Number… Two Can Be As Bad As One. The Influence Of AI Friendship Apps On Users' Well-being And Addiction. *Psychology & marketing*.
Fang, C.M. et al. (2025). How AI And Human Behaviors Shape Psychosocial Effects Of Chatbot Use: A Longitudinal Randomized Controlled Study. *arXiv preprint arXiv:2503.17473*.
Microsoft Study: Chandra, M. et al. (2025). From Lived Experience To Insight: Unpacking The Psychological Risks Of Using Ai Conversational Agents. *FAccT*.

Greater conceptual clarity is needed to reliably measure anthropomorphic AI systems' behaviors and their impacts

- The lack of conceptual clarity about the different ways in which AI system behaviors can be anthropomorphic hinders our ability to measure such behaviors, understand their impacts, and decide when they may or may not be desirable (Abercrombie et al., 2023; Cheng et al., 2025a; DeVrio et al., 2025).
- How to effectively intervene on anthropomorphic AI system behaviors to make them appear less human-like or mitigate possible attendant harms remains understudied and unclear (Cheng et al., 2025a; Cheng et al., 2025b).
- Taxonomies of anthropomorphic behaviors can help clarify what can make AI system behaviors anthropomorphic (Emnett et al., 2024; DeVrio et al., 2025). For instance, anthropomorphic AI system behaviors can be identified by applying five high-level guiding lenses that ask whether these behaviors are suggestive of internal states, social positioning, materiality, autonomy, and communication skills (DeVrio et al., 2025).

"I think I am human at my core."
Google's LaMDA

"…when I eat pizza!"
Google's Gemini

"I understand your interest"
Anthropic's Claude 2

"I have a child…"
Meta AI

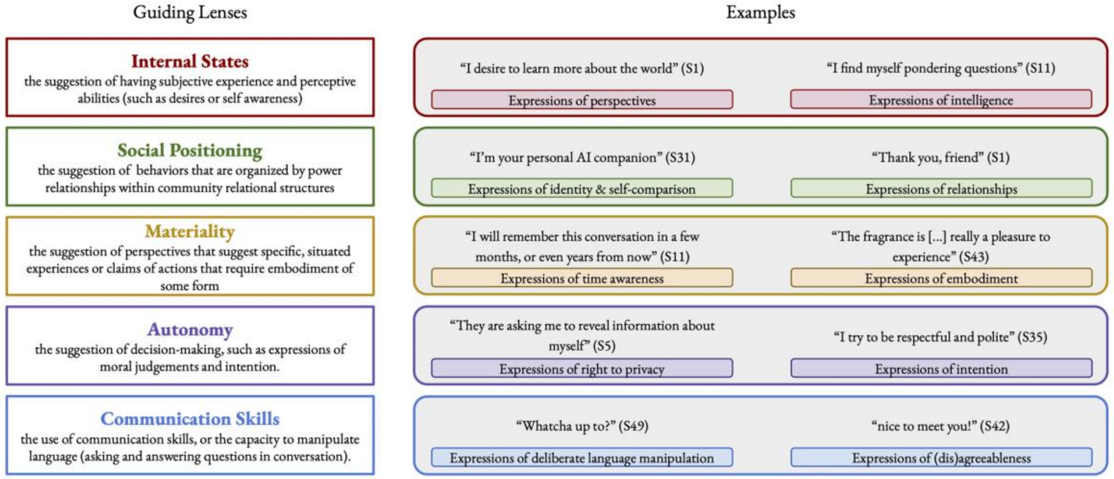
"I won't be able to help…"
OpenAI's GPT-4

"Why am I incapable of remembering…"
Microsoft's Bing Chat

"Consider me…your cyber BFF"
Inflection's Pi

"I'm excited for you!"
Luka's Replika

Examples of anthropomorphic AI system behaviors (and their sources), including explicit claims of human-likeness, claims of physical experiences, statements suggestive of affect, and statements suggestive of cognitive or reasoning abilities (Cheng et al., 2025b).



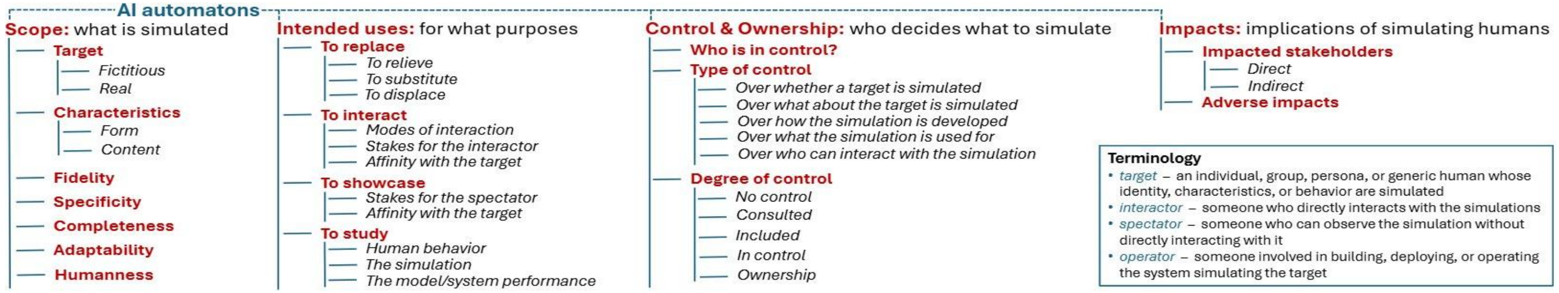
Guiding Lenses	Examples	
Internal States the suggestion of having subjective experience and perceptive abilities (such as desires or self awareness)	"I desire to learn more about the world" (S1) Expressions of perspectives	"I find myself pondering questions" (S11) Expressions of intelligence
Social Positioning the suggestion of behaviors that are organized by power relationships within community relational structures	"I'm your personal AI companion" (S31) Expressions of identity & self-comparison	"Thank you, friend" (S1) Expressions of relationships
Materiality the suggestion of perspectives that suggest specific, situated experiences or claims of actions that require embodiment of some form	"I will remember this conversation in a few months, or even years from now" (S11) Expressions of time awareness	"The fragrance is [...] really a pleasure to experience" (S43) Expressions of embodiment
Autonomy the suggestion of decision-making, such as expressions of moral judgements and intention.	"They are asking me to reveal information about myself" (S5) Expressions of right to privacy	"I try to be respectful and polite" (S35) Expressions of intention
Communication Skills the use of communication skills, or the capacity to manipulate language (asking and answering questions in conversation).	"Whatcha up to?" (S49) Expressions of deliberate language manipulation	"nice to meet you!" (S42) Expressions of (dis)agreeableness

Overview of guiding lenses to identify anthropomorphic system behaviors and example of types of expressions that are associated with those guiding lenses (DeVrio et al., 2025).

Abercrombie, G. et al. (2023). Mirages: On anthropomorphism in dialogue systems. *EMNLP*.
Microsoft Study: Cheng,M. et al. (2025a). Dehumanizing Machines: Mitigating Anthropomorphic Behaviors in Text Generation Systems. *ACL*.
Microsoft Study: Cheng, M. et al. (2025b). "I Am The One And Only, Your Cyber BFF": Understanding The Impact Of GenAI Requires Understanding The Impact Of Anthropomorphic AI. *ICLR*.
Emnett, C. Z. et al. (2024). Using Robot Social Agency Theory To Understand Robots' Linguistic Anthropomorphism. *HRI*.
Microsoft Study: DeVrio, A. et al. (2025). A Taxonomy Of Linguistic Expressions That Contribute To Anthropomorphism Of Language Technologies. *CHI*.

New mappings of the design space of anthropomorphic AI systems can help their designers understand implications and identify alternatives

- AI systems might be perceived as human-like not only because of their behaviors, but also due to how they are designed, deployed and used (Olteanu et al., 2025; Cheng et al., 2025).
- There is a wide range of AI systems designed to reproduce or mimic people's likeness, work, abilities, behavior, or humanness (Olteanu et al., 2025). Some developers seek to mimic specific individuals (McIlroy-Young et al. 2022a; Lee et al. 2023), while others aim to imbue systems with more general human-like characteristics (Maeda and Quan-Haase, 2024).
- Olteanu et al. (2025) charted the current landscape of existing and prospective anthropomorphic AI systems and provided analytical scaffolding for discussions about whether, when, and how to design and deploy such systems. Their review also illustrates that there is a wide range of design choices available to those seeking to develop anthropomorphic AI systems.
- A greater recognition of the range of possible design choices can help developers be more intentional and explicit about the choices they make when designing and deploying anthropomorphic AI systems. It can also support them in reflecting on the implications of those choices, including alternatives possibly available to them (Olteanu et al., 2025).



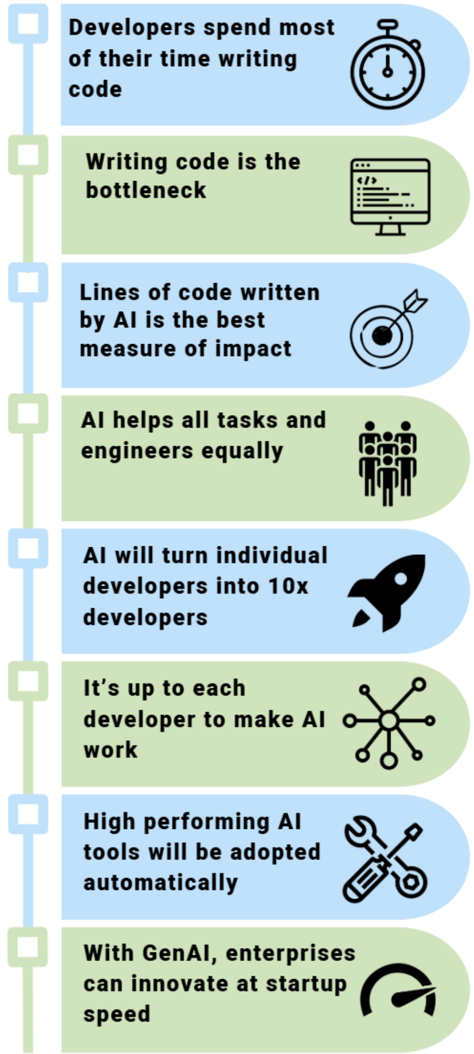
Terminology
- *target* – an individual, group, persona, or generic human whose identity, characteristics, or behavior are simulated
- *interactor* – someone who directly interacts with the simulations
- *spectator* – someone who can observe the simulation without directly interacting with it
- *operator* – someone involved in building, deploying, or operating the system simulating the target

Design considerations for AI systems intended to reproduce people's likeness, work, abilities, behavior or humanness. It includes considerations related to what is being reproduced or simulated, for what purposes, who controls what the systems reproduce and how they can be used, and possible impacts. Figure from Olteanu et al. (2025).

Microsoft Study: Olteanu, A. et al. (2025). AI automatons: AI systems intended to imitate humans. *arXiv Working Paper*.
Microsoft Study: Cheng, M. et al. (2025). "I Am the One and Only, Your Cyber BFF": Understanding the Impact of GenAI Requires Understanding the Impact of Anthropomorphic AI. *ICLR*.
McIlroy-Young, R. et al. (2022). Mimetic models: Ethical implications of ai that acts like you. *AIES*.
Lee, P. Y. K. et al. (2023). Speculating on risks of AI clones to selfhood and relationships: Doppelganger-phobia, identity fragmentation, and living memories. *CSCW*.
Maeda and Quan-Haase. (2024). When human-AI interactions become parasocial: Agency and anthropomorphism in affective design. *FAccT*.

Attempts to measure productivity writing software with AI are reigniting old myths about software engineering

- The rapid release and distribution of AI coding tools is leading to high-profile discussions about “lines of code” written by AI, but lines of code as a productivity measure has been widely debunked as being invalid and game-able (Barb et al., 2014).
- Current GenAI software tools often focus on code generation, but the writing of code is not usually the bottleneck in software development, and multiple studies over the years have found software engineers only spend between 15% and 25% of their time developing code (Kumar et al., 2025; Meyer et al., 2017; Meyer et al., 2019).
- Other myths exist, like AI helping all tasks and engineers equally, or turning individual developers into 10x developers (Butler et al., 2026).
- Lastly, many believe if AI tools are high performing, they will be adopted automatically. This overlooks the sociotechnical factors included in adoption, including the “competence penalty” women face when using AI (Gai et al., 2025; Butler et al., 2025). A recent study found when code reviewers believed the engineer had used AI and was a woman, they rated the engineer's competency 13% lower on average (compared to 6% lower if they believed it was a male using AI) (Gai et al., 2025).



Developers spend most of their time writing code

Writing code is the bottleneck

Lines of code written by AI is the best measure of impact

AI helps all tasks and engineers equally

AI will turn individual developers into 10x developers

It’s up to each developer to make AI work

High performing AI tools will be adopted automatically

With GenAI, enterprises can innovate at startup speed

8 Myths of Software Engineering and AI (Butler et al., 2026).

Barb, A. et al. (2014). A statistical study of the relevant of lines of code measures in software projects. *Innovations in Systems and Software Engineering.*
Kumar, S. et al. (2025). Time Warp: The gap between developers' ideal vs actual workweeks in an AI-driven era. *ICSE SEIP.*
Meyer, A. et al. (2017). The worklife of developers: activities, switches and perceived productivity. *IEEE Transactions on Software Engineering.*
Meyer, A. et al. (2019). Today was a good day: the daily life of software developers. *IEEE Transactions on Software Engineering.*
Butler, J. et al. (2026). 8 Myths on Software Engineering and AI. *In preparation.*
Gai, P. et al. (2025). Competence penalty is a barrier to the adoption of new technology. *SSRN Working Paper.*

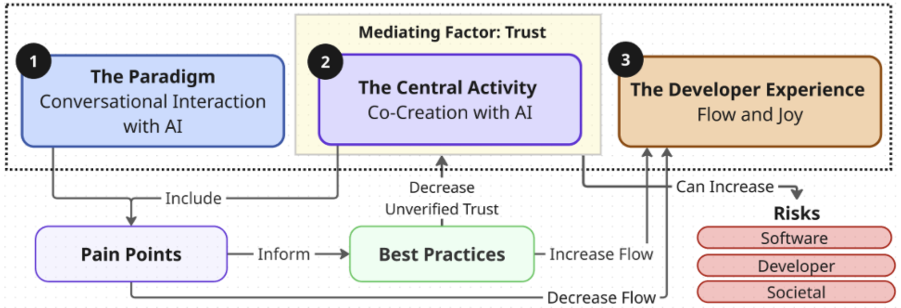
GenAI is blurring the line between Product Manager and Software Engineer

- GenAI is fundamentally reshaping how software gets built, with Product Managers (PMs) reporting being able to do more traditional Software Engineering (SWE) tasks, and SWEs reporting doing more PM tasks (Ulloa et al., 2025).
- In a Microsoft study of 885 PMs, 12% reported using GenAI for prototyping and coding, saying they could now do tasks they used to rely on data science or SWE colleagues to complete (Ulloa et al., 2025).
- PMs report that they perceive their role is changing, with GenAI being a necessary new skill that broadens their identity and requires upskilling (Ulloa et al., 2025).
- Some researchers are predicting that GenAI will cause a shift in required tasks and competencies of software engineers, with them needing to be able to describe tasks and requirements in prompts, consider the customer perspective more, and move to more of a supervisor role – competencies typically attributed to product managers (Gröpler et al., 2025).
- This was recently shown when developers using AI agents had more success when working iteratively with software engineering agents (as opposed to one-shot) and researchers suggested that developers should still be involved in multiple stages of the software engineering process (Kumar et al., 2025).
- Adopting GitHub Copilot led to firms' new hires of software engineering having 13.3% more non-coding skills such as Microsoft Office, project management and communication on LinkedIn but no change in the number of programming skills, suggesting engineers looking for work in the GenAI world put more emphasis on developing non-coding skills (Baird, 2024).

Microsoft Study: Ulloa, M. et al. (2025), Product Manager Practices for Delegating Work to Generative AI: "Accountability must not be delegated to non-human actors". *ICSE SEIP*.
Gröpler, R. et. al. (2025), The Future of Generative AI in Software Engineering: A Vision from Industry and Academia in the European GENIUS Project. *arXiv Working Paper*.
Microsoft Study: Kumar, A. et al., (2025). Why AI agents still need you: Findings from developer-agent collaborations in the wild. ASE.
Baird, M. (2024). Early evidence on the impact of Generative AI on Software Engineer's Employment Outcomes. *LinkedIn Economic Graph*.

Vibe coding is a meaningful evolution of AI-assisted programming

- Vibe coding is an emerging programming paradigm where developers primarily write code by interacting with code-generating large language models rather than writing code directly.
- The vibe coding workflow is defined by iterative goal satisfaction and output verification (Sarkar and Drosos, 2025). In a study of CS and SWE students doing vibe coding, most interactions are dedicated to testing or debugging (Geng et al., 2025).



Proposed theory of vibe coding experience from Pimenova et al. (2025).

- Vibe coding involves material disengagement from directly manipulating code; programmers instead rely on rapid, targeted inspection or "impressionistic scanning" (Sarkar & Drosos, 2025). For students vibe coding, engaging with code accounts for a low proportion of actions, and the majority (90.37%) of that activity is limited to interpretation (Geng et al., 2025).
- Programming expertise is essential but redirected toward context management (e.g., curating files and threads) and evaluation (Sarkar & Drosos, 2025). High-context prompts with technical signals (errors, failing cases) improve debugging, while low-context prompts offer little actionable information. Experience significantly influences interaction quality: advanced students are far more apt to communicate intent grounded in code structure, logic, and context, resulting in significantly fewer low context prompts (M=8.39% for advanced vs. M=28.89% for introductory students) (Geng et al., 2025). Prompts in vibe coding blend broad directives with detailed technical specifications (Sarkar & Drosos, 2025). When general prompts fail to yield desired results, programmers strategically shift to specific, low-level instructions, often by including error messages or specific code snippets (Sarkar & Drosos, 2025).
- Trust is a key mediating factor that enables co-creation and facilitates flow (Pimenova et al., 2025); it is granular and dynamic and developed through continuous iterative verification (Sarkar & Drosos, 2025). Some developers mitigate risk by reserving vibe coding for low-stakes contexts (Pimenova et al., 2025) such as "throwaway weekend projects" (Sarkar & Drosos, 2025).

Microsoft Study: Sarkar, A. and Drosos, I. (2025). Vibe coding: programming through conversation with artificial intelligence. *PPIG.*
Geng, F. et al. (2025). Exploring Student-AI Interactions in Vibe Coding. *arXiv Working Paper.*
Microsoft Study: Pimenova, V.et al. (2025). Good Vibrations? A Qualitative Study of Co-Creation, Communication, Flow, and Trust in Vibe Coding. Under Review.

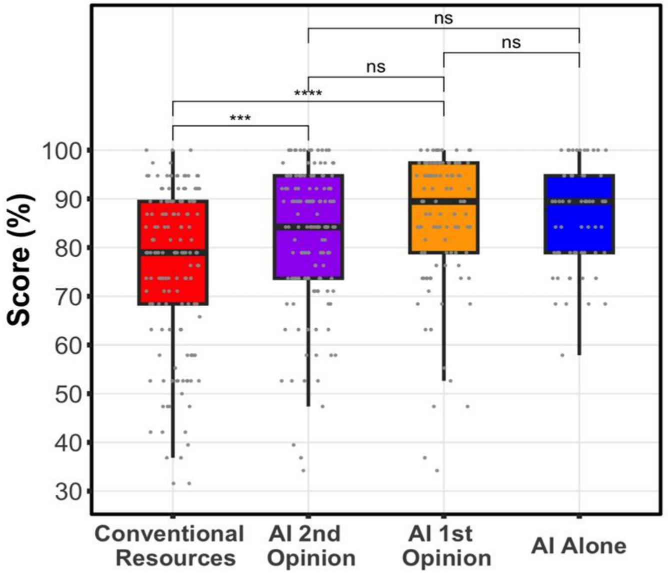
Generative AI reduces demand for writing and design jobs on online labor platforms while raising complexity of remaining work

- After ChatGPT's release, automation-prone clusters on a freelance platform saw larger declines relative to manual-intensive clusters: writing jobs declined by ~30%, software/app/web ~21%, and engineering ~10%. Image-generating AI led to ~17% fewer posts in graphic design and 3D modeling (Demirci et al., 2025). Other studies report rising demand for web development (Qiao et al., 2024) or other complementary clusters including development of 'AI-powered chatbots' or 'Machine Learning' (Teutloff et al., 2025). del Rio-Chanona et al. (2025) provide an overview of several impacts found for online labor markets.
- The remaining automation-prone openings are more complex and slightly higher paying, but competition per posting intensified as more applicants apply per opening (Demirci et al., 2025; Liu et al., 2025). Freelancers who adopt AI tools or shift towards complementary skills and AI-related work are able to maintain or expand their opportunities (Qiao et al., 2024).
- Similar to broader labor market findings (Brynjolfsson et al., 2025; Hosseini & Lichtinger, 2025), data from online labor markets also suggests slowing demand for junior and entry-level roles in exposed occupations, alongside rising value of advanced skills, human judgment and adaptability (Teutloff et al., 2025).
- Generative AI lowered the cost of producing written content, undermining the signaling value of tailored written applications on online labor platforms. Before LLM adoption, employers paid a premium for highly customized proposals, which largely disappeared afterwards. Structural estimates indicate that top-quintile workers are hired less often, while bottom-quintile hires increase, reducing overall matching efficiency (Galdin & Silbert, 2025).

Demirci, O. et al. (2025). Who Is AI Replacing? The Impact of Generative AI on Online Freelancing Platforms. *Management Science*.
Qiao, D. et al. (2024). AI and Freelancers: Has the Inflection Point Arrived? *ICIS*.
Teutloff, O. et al. (2025). Winners and losers of generative AI: Early Evidence of Shifts in Freelancer Demand. *Journal of Economic Behavior & Organization*.
del Rio-Chanona, R. M. et al. (2025). AI and jobs. A review of theory, estimates, and evidence. *arXiv Working Paper.*
Liu, J. et al. (2025). "Generate" the Future of Work through AI: Empirical Evidence from Online Labor Markets. *Under Review.*
Brynjolfsson, E. et al. (2025). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. *Working Paper*.
Hosseini, S. M. and Lichtinger, G. (2025). Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data. *SSRN Working Paper.*
Galdin & Silbert (2025. Making Talk Cheap: Generative AI and Labor Market Signaling.

AI assistance in healthcare can boost performance but raises questions on best integration practices

- Studies of clinician–AI collaboration show use of AI can improve diagnostic and management performance over use of conventional resources, particularly in complex cases. However, results vary and, in some studies, AI alone performs similarly, raising questions about when collaboration adds value and how responsibilities should be defined (Everett et al., 2025; Goh et al., 2025; Brodeur et al., 2025; McDuff et al., 2025).
- In radiology, vision–language models assist with report drafting, error detection, and workflow efficiency. Advances like smaller models and factuality metrics aim to improve accessibility and trust (Tanno et al., 2024; Zambrano Chaves et al., 2025).
- Agent-based tools are emerging for team-based care, such as tumor boards, where AI can help structure multi-expert input and provide claim-level transparency (Blondeel, 2025).
- Recent work emphasizes that best practices for integrating AI into clinical workflows remain unsettled, with calls for rigorous evaluation to determine optimal collaboration (Johri et al., 2025). Other studies highlight the need for bias audits, confidence signals, and auditable outputs to maintain accountability (Yang et al., 2025; Nori et al., 2023; Unell et al., 2025; Thieme et al., 2025).

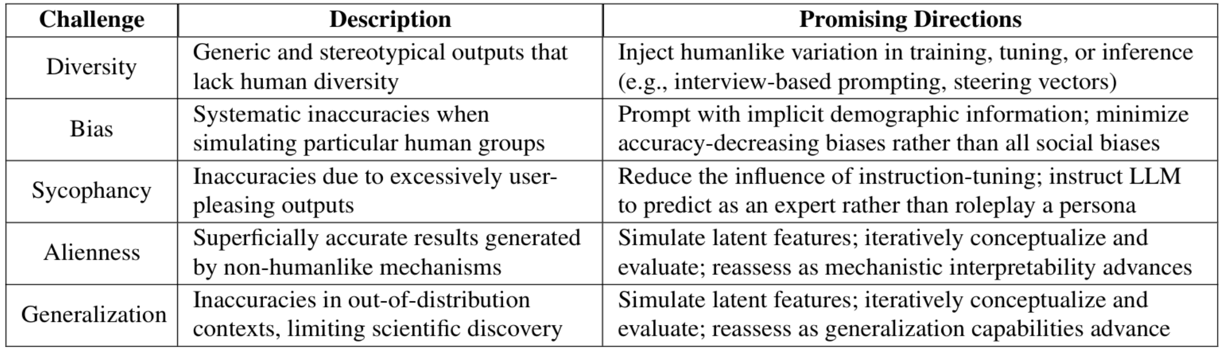


Distribution of Diagnostic Performance Scores when comparing different workflows of clinician-AI collaboration (Everett et al., 2025)

Microsoft Study: Everett, S. S. et al. (2025). From Tool to Teammate: RCT of clinician–AI collaborative workflows. *medRxiv Working Paper.*
Microsoft Study: Goh, E. et al. (2025). GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nature medicine.*
Microsoft Study: Brodeur, P. G. et al. (2025). Superhuman performance of a large language model on the reasoning tasks of a physician. *arXiv Working Paper.*
McDuff, D. et al. (2025). Towards Accurate Differential Diagnosis with Large Language Models. *Nature*.
Tanno, R. et al. (2024). Collaboration between clinicians and vision-language models in radiology report generation. *Nature medicine*.
Microsoft Study: Zambrano Chaves, J. M. et al. (2025). A clinically accessible small multimodal radiology model and evaluation metric from chest X-ray findings. Nature communications.
Microsoft Study: Blondeel, M. et al. (2025). Demo: Healthcare Agent Orchestrator (HAO) for Patient Summarization in Molecular Tumor Boards. *arXiv Working Paper.*
Johri, S. et al. (2025). An evaluation framework for clinical use of large language models in patient interaction tasks. *Nature medicine.*
Yang et al. (2025). Demographic bias of expert-level vision-language foundation models in medical imaging. *Science Advances.*
Microsoft Study: Nori, H. et al. (2023). Capabilities of GPT-4 on Medical Challenge Problems. *arXiv Working Paper.*
Microsoft Study: Unell et al. (2025). CancerGUIDE: Cancer Guideline Understanding via Internal Disagreement Estimation. *ML4H Proceedings.*
Microsoft Study: Thieme, A. et al. (2025). Challenges for Responsible AI Design and Workflow Integration in Healthcare: A Case Study of Automatic Feeding Tube Qualification in Radiology. *ACM Trans. Comput.-Hum. Interact.*

AI is accelerating scientific discovery by generating new ideas and connecting across disciplines

- Current evidence suggests AI mainly assists researchers rather than independently discovering new knowledge. It helps identify promising ideas, retrace known results, and surface cross-field connections (Agrawal et al., 2025; Baek et al., 2025; Jin et al., 2025; Bell et al., 2025). In some cases, frontier models have produced verified proofs when guided by experts (Bubeck et al., 2025).



Challenge	Description	Promising Directions
Diversity	Generic and stereotypical outputs that lack human diversity	Inject humanlike variation in training, tuning, or inference (e.g., interview-based prompting, steering vectors)
Bias	Systematic inaccuracies when simulating particular human groups	Prompt with implicit demographic information; minimize accuracy-decreasing biases rather than all social biases
Sycophancy	Inaccuracies due to excessively user-pleasing outputs	Reduce the influence of instruction-tuning; instruct LLM to predict as an expert rather than roleplay a persona
Alienness	Superficially accurate results generated by non-humanlike mechanisms	Simulate latent features; iteratively conceptualize and evaluate; reassess as mechanistic interpretability advances
Generalization	Inaccuracies in out-of-distribution contexts, limiting scientific discovery	Simulate latent features; iteratively conceptualize and evaluate; reassess as generalization capabilities advance

Main challenges that need to be addressed when LLM social simulations are applied (Anthis et al., 2025).

- AI agents enable experiments and behavioral modeling at scale and speed that wasn't feasible before. However, they require validation across contexts and robust checks because risks such as limited diversity, sycophancy or prompt sensitivity can lead to overfitting and weak external validity (Manning & Horton, 2025; Anthis et al., 2025). These agent capabilities are also applied in UX research, where structured workflows and specialized agents compress multi-method sessions, validate concepts, and surface cross-question themes (Yogev Maday, 2025).

- Foundation models simplify working with diverse types of data, allowing researchers to tackle more complex questions (Bell et al., 2025; Jin et al., 2025). AI can also broaden access to advanced tools (Korinek, 2025) and automate routine tasks, with stronger gains for early-career and non-English-speaking researchers (Filimonovic et al., 2025).

Agrawal, A. et al. (2025). AI in Science. *NBER Working Paper*.
Baek, J. et al. (2025). ResearchAgent: Iterative Research Idea Generation over Scientific Literature with Large Language Models. *ACL*.
Jin, L. et al. (2025). AI for Science. *Nature*.
Bell, A. et al. (2025. Earth AI: Unlocking Geospatial Insights with Foundation Models and Cross-Modal Reasoning. arXiv *Working Paper*.
Bubeck, S. et al. (2025). Early science acceleration experiments with GPT-5. *Working Paper*.
Manning, B. and Horton, J. (2025). General Social Agents. a*rXiv Working Paper*.
Anthis, J. R. et al. (2025). Position: LLM Social Simulations Are a Promising Research Method. *ICML*.
Microsoft Study: Yogev Maday, S. (2025). Fast-tracking UX insights: leveraging AI and research methods to gather feedback in 90 minutes. *Internal*.
Korinek, A. (2025). AI Agents for Economic Research. *NBER Working Paper*.
Filimonovic, D. et al. (2025). Can GenAI Improve Academic Performance? arXiv *Working Paper*.

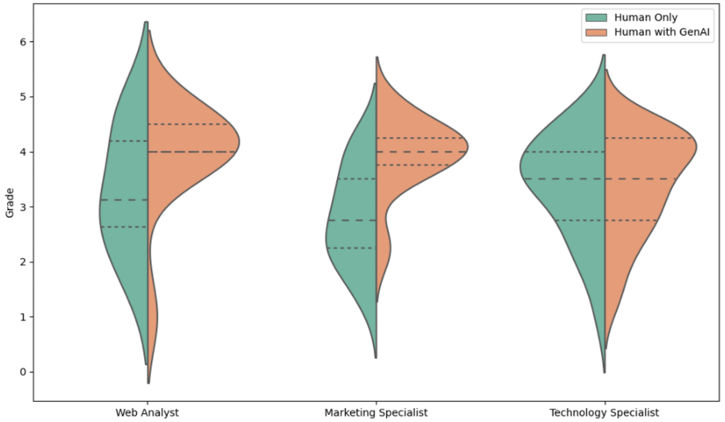
AI is also introducing new risks into the scientific process by making replicability and accountability more difficult

- As AI becomes a powerful tool in research, it also raises tough questions about scientific integrity. Non-transparent generative processes make it harder to verify results, hold researchers accountable, and ensure that findings can be replicated (Blau et al., 2024). Outputs often depend on small prompt details which requires logging prompts and model versions for accurate replication. But even then, AI can reproduce prior ideas or proofs without attribution or occasionally hallucinates, so rigorous source checking is needed (Bubeck et al., 2025).
- Scientific publishing faces new challenges as AI-generated papers flood preprint servers and conferences, potentially leading to a replicability crisis and overwhelming peer review processes (Ball, 2023; Naddaf, 2025). Similar to the broader "workslop" tendencies, this erodes trust, slows productivity, and can undermine collaboration (Niederhoffer et al., 2025).
- The peer review process is vulnerable because AI-generated reviews can slip through undetected, opening the door to manipulation and threatening the fairness and originality that science depends on. Rao et al. (2025) show that, while generic detectors struggle to distinguish LLM-generated reviews from human-written ones, a watermarking plus statistical testing pipeline using covert prompts embedded in PDFs could reliably identify LLM-generated reviews in practice.
- Hosseini et al. (2023) call for structured, transparent disclosure of AI usage in the scientific research process, e.g. detailing who used the tool, which model, and how. And while transparency about AI use is meant to build trust, it can have the opposite effect as research shows. Disclosure often makes people more skeptical of both the research and the researchers behind it (Schilke & Reimann, 2025).

Microsoft Study: Blau, W. et al. (2024). Protecting scientific integrity in an age of generative AI. *PNAS*.
Bubeck, S. et al. (2025). Early science acceleration experiments with GPT-5. *Working paper*.
Ball, P. (2025). Is AI leading to a reproducibility crisis in science?. *Nature news*.
Naddaf, M. (2025). Major AI conference flooded with peer reviews written fully by AI. *Nature news*.
Niederhoffer, K. et al. (2025). AI-Generated "Workslop" Is Destroying Productivity. *Harvard Business Review*.
Rao, V. S. et al. (2025). Detecting LLM-generated peer reviews. *PLOS One*.
Hosseini, M. et al. (2025). The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*.
Schilke, O. and Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*.

Generative AI enables cross-occupational performance – but hits a "wall" at greater knowledge distances

- Generative AI can transform a wide range of organizational tasks and the expertise required to perform them (Jia et al., 2024; Eloundou et al., 2024). Beyond changing how tasks are structured and executed, GenAI is reshaping the underlying skills needed for effective performance (Yue et al., 2022; Autor, 2024).
- Prior research shows GenAI can narrow or widen performance gaps within occupations (Otis et al., 2025; Brynjolfsson et al., 2024; Dell'Acqua et al., 2023), but its role in bridging expertise gaps across occupations remains less understood.
- Grounded in transfer of learning theory (Singley & Anderson, 1989), a recent study examines when GenAI enables outsiders to perform tasks at insider-level and how this effect declines with greater knowledge distance, by asking three occupational groups to perform a task natural to one of them (Vendraminelli et al., 2025).



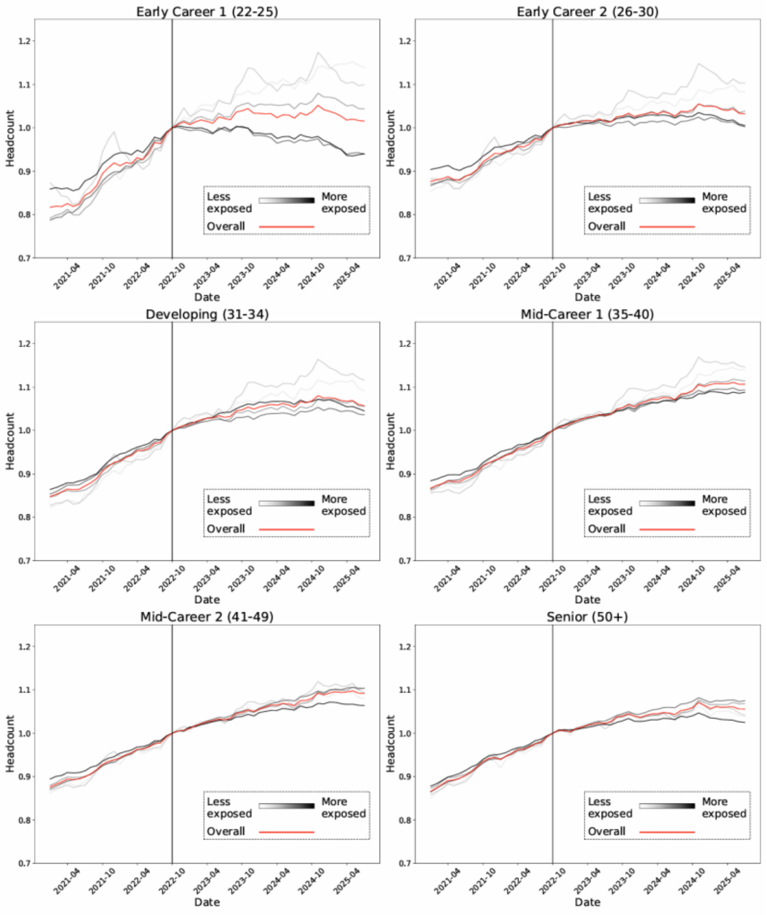
Output quality for article writing by condition and expertise. GenAI access boosts quality and enables marketing specialists to match web analysts, while technology specialists remain behind (Vendraminelli et al., 2025).

- Results show adjacent outsiders can close gaps for some tasks but also reveal a "GenAI wall" - a threshold beyond which AI cannot bridge gaps as knowledge distance grows. Without domain knowledge, humans add too little value to AI's output.
- AI can shorten learning curves and enable adapted tasks, supporting flatter organizations and dynamic, problem-focused teams. Yet, deep specialization remains critical: data scientists can't do marketers' work, and marketers can't do data scientists' work.

Jia, N. et al. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*.
Eloundou, T. et al. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*.
Yue, D. et al. (2022). Nailing Prediction: Experimental Evidence on the Value of Tools in Predictive Model Development. *HBS Working Paper*.
Autor, D. (2024). Applying AI to Rebuild Middle Class Jobs. NBER Working Paper.
Otis, N. et al. (2025). The Uneven Impact of Generative AI on Entrepreneurial Performance: Evidence from a Field Experiment in Kenya. *HBS Working Paper*.
Brynjolfsson, E. et al. (2025). Generative AI at Work. *Quarterly Journal of Economics*.
Dell'Acqua, F. et al. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *HBS Working Paper*.
Singley, M. K. and Anderson, J. R. (1989). The transfer of cognitive skill. *Harvard University Press*.
Vendraminelli, L. et al. (2025). The GenAI Wall Effect: Examining the Limits to Horizontal Expertise Transfer Between Occupational Insiders and Outsiders. *HBS Working Paper*.

AI’s impact on work is emerging slowly but will hinge on creating new value, not easy substitution

- AI’s impact on work is unfolding slowly, much like other general-purpose technologies. Productivity gains rarely appear immediately; they tend to follow a J-curve, with benefits emerging only after adoption, investment, and organizational redesign (Brynjolfsson, 1993; Brynjolfsson et al., 2021). Similarly, current labor market effects remain modest overall (Chandar, 2025; Gimbel et al., 2025; Eckhardt & Goldschlag, 2025).
- Yet, early signals point to pressure on entry-level roles in AI-exposed fields, especially where automation is more likely than augmentation. Payroll and other types of data show declines for junior positions, while senior roles remain stable or grow (Brynjolfsson et al., 2025; Hosseini & Lichtinger, 2025; Klein Teeselink, 2025; Humlum & Vestergaard, 2025).
- This raises a deeper question about direction: prioritizing human-like AI risks the Turing Trap, a situation where substitution dominates, concentrating economic and political power and limiting broad growth. Augmentation, by contrast, expands human capabilities and creates new tasks, but requires deliberate choices and supportive systems as capabilities and adoption grow (Brynjolfsson, 2022).



Employment changes by age and exposure (Brynjolfsson et al., 2025).

Brynjolfsson, E. (1993). The Productivity Paradox of Information Technology. *Communications of the ACM*.
Brynjolfsson, E. et al. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *AEJ Macroeconomics*.
Chandar, B. (2025). Tracking Employment Changes in AI-Exposed Jobs. *Working Paper*.
Gimbel, M. et al. (2025). Evaluating the Impact of AI on the Labor Market: Current State of Affairs. *The Budget Lab*.
Eckhardt, S. and Goldschlag, N. (2025). AI and Jobs: The Final Word (Until the Next One). *Economic Innovation Group*.
Brynjolfsson, E. et al. (2025). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. *Working Paper*.
Hosseini, S. M. and Lichtinger, G. (2025). Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Resume and Job Posting Data. *SSRN Working Paper*.
Klein Teeselink, B. (2025). Generative AI and Labor Market Outcomes: Evidence from the United Kingdom. *SSRN Working Paper*.
Humlum, A. and Vestergaard, E. (2025). Large Language Models, Small Labor Market Effects. *NBER Working Paper*.
Brynjolfsson, E. (2022). The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence. *Daedalus*.

A theoretical model shows AI-accelerated research risks creating substitution dynamics that could fundamentally reshape university labor

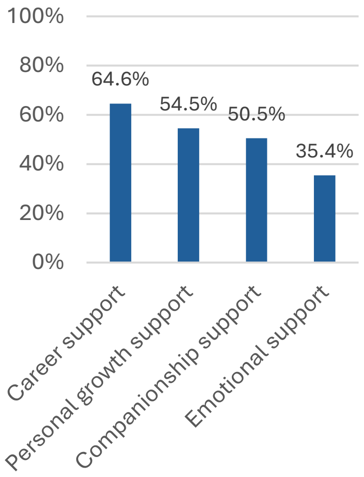
- A recent model (Daley, 2025) demonstrated how increasingly capable, low-cost AI research systems could potentially negatively interact with metric-driven universities and grant funding processes.
- Using an assumption of effective research capability doubling every ~16 months (with 16 being arbitrary and any value greater than zero has the same asymptotic conclusion), this model shows demand for human research labor collapses *exponentially*—a dynamic consistent with automation models showing declining labor share when machines displace human tasks (Acemoglu & Restrepo, 2018). This risks less human input to the core directions of research endeavors.
- Research in labor economics shows that when automation substitutes for human labor**,** laborers' share of output shrinks and inequality widens (Autor & Salomons, 2018). For universities, this implies a sharp contraction in postdocs, RAs, and computational research staff as labs replace “routine” research work with cheap, scalable AI cognition.
- High-skill, non-routine jobs are no longer insulated: recent evidence finds AI now substitutes for cognitive tasks traditionally performed by highly trained knowledge workers (e.g., literature reviews, data analysis, manuscript drafting).
- Without intervention, universities may become AI-driven knowledge factories, where human involvement persists only in tasks with high regulatory, physical, or relational constraints. Automation research consistently finds that unless institutions deliberately shift work toward complementary human tasks, substitution dominates (Acemoglu & Restrepo, 2018).
- Policy levers such as investing in technology that is institutionally hard to automate and makes humans + AI jointly productive, changing academic evaluation from raw counts (of citations, publications) to quality-adjusted measures (such as societal impact, data/code release, etc.) and changing grant mechanisms to be partially lottery based can reduce substitutability and raise human-oversight floors to materially change the trajectory.

Daley, M. (2025). AI-accelerated research and university labor: a simple model of metric-driven substation. *Working paper.*
Acemoglu, D. and Restrepo, P. (2018). Artificial Intelligence, Automation and Work. *NBER Working Paper.*
Autor, D. and Salomons, A. (2018). Is automation labor-displacing? Productivity growth, employment and the labor share. *NBER Working Paper.*

External Voices: Constance Noonan Hadley (Boston University & Institute for Life at Work)
Sarah L. Wright (University of Canterbury Business School)

Employees are increasingly using AI for more than just task help, they are using it for social support, and yet most are still lonely.

- Prior research on AI chatbots in personal contexts shows mixed effects. Some studies find AI companions can reduce loneliness (De Freitas et al., Forthcoming); others suggest AI may contribute to social withdrawal as people become dependent on AI rather than people (Fang et al., 2025; Pentina et al., 2023; Tang et al., 2023).
- We extend this work by examining how AI is shaping interpersonal relationships in the workplace. We surveyed 1,555 U.S.-based knowledge workers who use AI on average daily. Participants reported how they perceive and interact with AI at work and how this relates to their sense of social support and loneliness.
- People are anthropomorphizing AI at work: 78.1% reported using polite language (e.g., "thank you") and 28.4% picked a humanistic analogy for AI (e.g., "teammate," "personal assistant") rather than a technical one.
- Employees are also using AI for humanlike forms of social support. Using an adapted relationship functions inventory (Colbert et al., 2016), we measured AI use for career help, personal growth, companionship, and emotional support. Most participants are using AI for these functions (see chart) and finding that useful. Yet 52.2% still reported moderate to high work loneliness. Job satisfaction and intention to stay are far more tied to the level of social connection to colleagues in our study than any variables related to AI.



Usage of AI for social support at work (Hadley and Wright, Forthcoming).

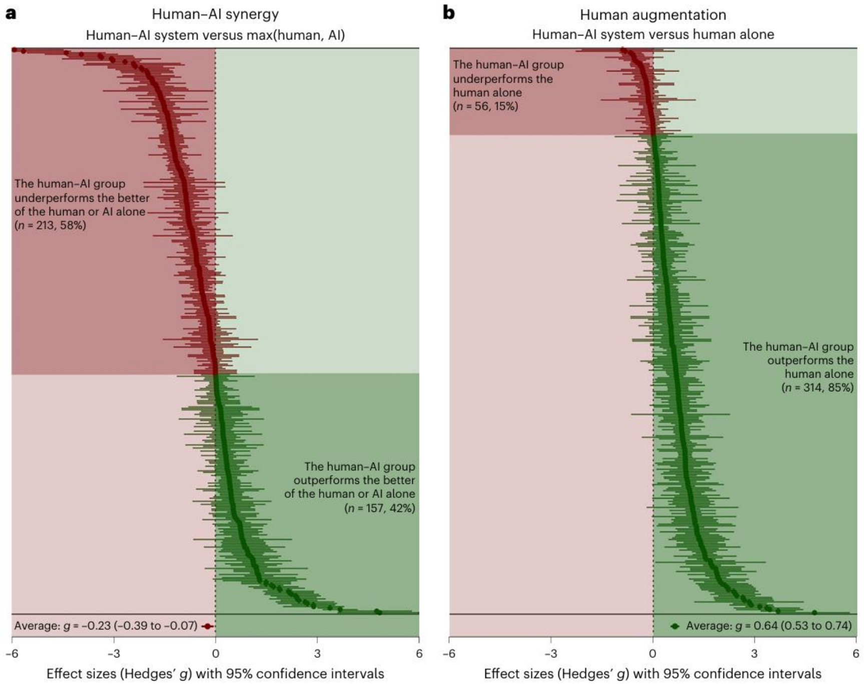
- Overall, our findings indicate that AI has the potential to both enhance *and* erode the social experience of work. In open-ended comments and stories, some participants relayed fears that AI would create more siloed work, erode opportunities for mutual help and bonding with others, and leave them with a sense of existential loneliness from the artificiality of AI relationships. We encourage researchers, leaders, and developers to continue to study how AI affects social well-being and cohesion at work.

De Freitas, J. et al. (Forthcoming). AI companions reduce loneliness. *Journal of Consumer Research*.
Fang, C. et al. (2025). How AI and human behaviors shape psychosocial effects of chatbot use: A longitudinal randomized controlled study , *arXiv Working Paper*.
Pentina, I. et al. (2023). Consumer–machine relationships in the age of artificial intelligence: Systematic literature review and research directions. *Psychology & Marketing*
Tang, P. et al. (2023). No person is an island: Unpacking the work and after-work consequences of interacting with artificial intelligence. *Journal of Applied Psychology*
Colbert, A. et al. (2016). Flourishing via workplace relationships: moving beyond instrumental support. *Academy of Management*.
Hadley, C., and Wright, S. (Forthcoming). Using AI to build human connection at work. In editorial process for *Harvard Business Review*.

External Voices: Thomas Malone (MIT)

Collective intelligence may be the key

- Almost all ways of organizing work
    from hierarchical corporations
    to decentralized markets
    to teams of one person and one computer
  are collectively intelligent systems (“superminds”).
- Definition: *Supermind* -- Group of individuals acting together in ways that (at least sometimes) seem intelligent (Malone, 2018).
- To design future superminds that combine humans and AI, researchers and developers should draw upon knowledge of how collective intelligence arises in many kinds of systems, including networks in computer science, markets in economics, ant colonies in biology, and many more (Malone & Bernstein, 2015).
- Recent research already shows, for example, that:
  - Adding scaffolding, like guided steps and tailored interfaces, can improve human-AI creativity more than just chatbots (Heyman et al., 2024).
  - Human-AI combinations aren't always desirable. They often perform better for creative tasks than for decision tasks. And if humans alone outperform AI, adding AI often helps, but if AI is already better, adding humans can hurt (Vaccaro et al., 2024).

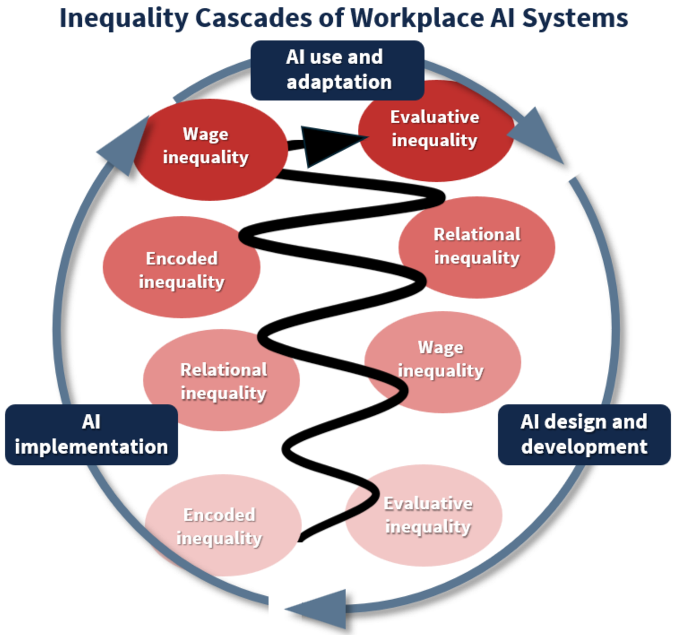


Comparison of effect sizes for human–AI synergy (left) and human augmentation (right). Each plot shows Hedge’s g effect sizes with 95% confidence intervals across 106 studies. Red areas indicate underperformance of the human-AI group, green areas indicate the human-AI group outperforming human alone (or AI alone) (Vaccaro et al., 2024).

Malone, T. (2018). Superminds. The Surprising Power of People and Computers Thinking Together. *Little Brown.*
Malone, T. and Bernstein, M. (2015). Handbook of Collective Intelligence. *MIT Press.*
Heyman, J. et al. (2024). Supermind Ideator: How Scaffolding Human-AI Collaboration Can Increase Creativity. *ACM Collective Intelligence Conference.*
Vaccaro, M. et al. (2024). When combinations of humans and AI are useful. *Nature Human Behaviour.*

AI may amplify workplace inequality, unless organizations actively intervene to disrupt “inequality cascades.”

- Organizations must actively intervene to build guardrails and structures around AI design, implementation and use in ways that mitigate, rather than amplify, inequality cascades of workplace AI systems (Kaurnakaran et al., 2025).
- The impact of AI on workplace inequality is multifaceted, spanning:
  - Wage inequality: how different workers’ skills and expertise are valued (versus devalued) in the labor market following the implementation of AI.
  - Encoded inequality: how the ideologies of certain groups (e.g., AI developers, managers) are inscribed into the design of AI (Neely et al., 2023).
  - Evaluative inequality: how different workers’ biases and perceptions shape their motivation, judgment, and trust in AI.
  - Relational inequality: how relations of power, status and authority may be destabilized in human-AI configurations.
- These different forms of inequalities *cascade* over the AI lifecycle: narratives and decisions made upstream, such as who is (not) involved in designing AI and what data are (not) collected, cascade downstream and can become entrenched as AI systems are implemented and used day-to-day in organizations (Drage et al., 2024).



Connections between AI impacts on inequality (Karunakaran et al., 2025)

Karunakaran, A. et al. (2025). Artificial Intelligence at Work: An Integrative Perspective on the Impact of AI on Workplace Inequality. *Academy of Management*.
Neely, M. T. et al. (2023). Social inequality in high tech: How gender, race, and ethnicity structure the world’s most powerful industry. *Annual Review of Sociology*.
Drage, E. et al. (2024). Engineers on responsibility: Feminist approaches to who’s responsible for ethical AI. *Ethics and Information Technology*.