

The Effect of AI on Work and Labor

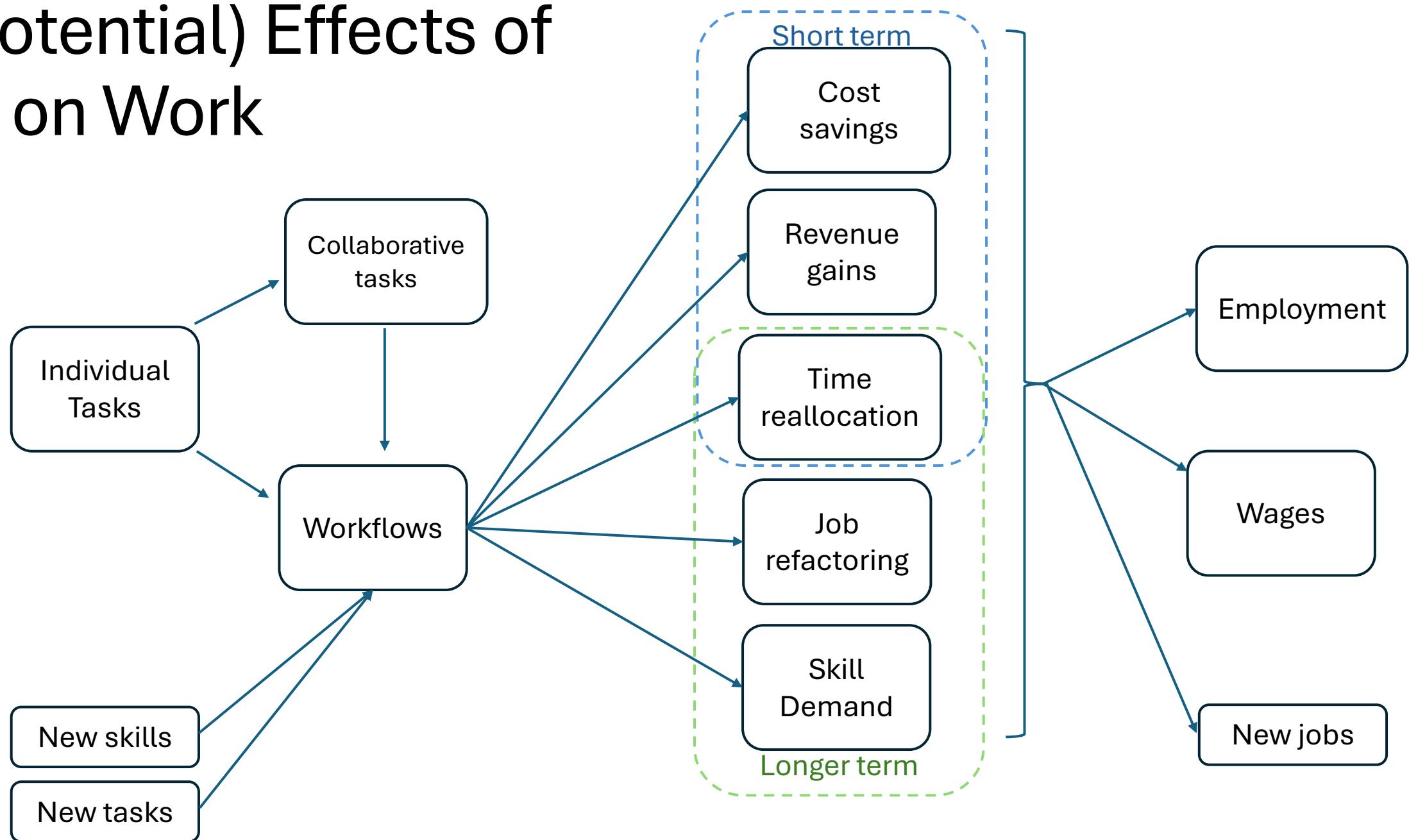
Sonia Jaffe

Microsoft Research

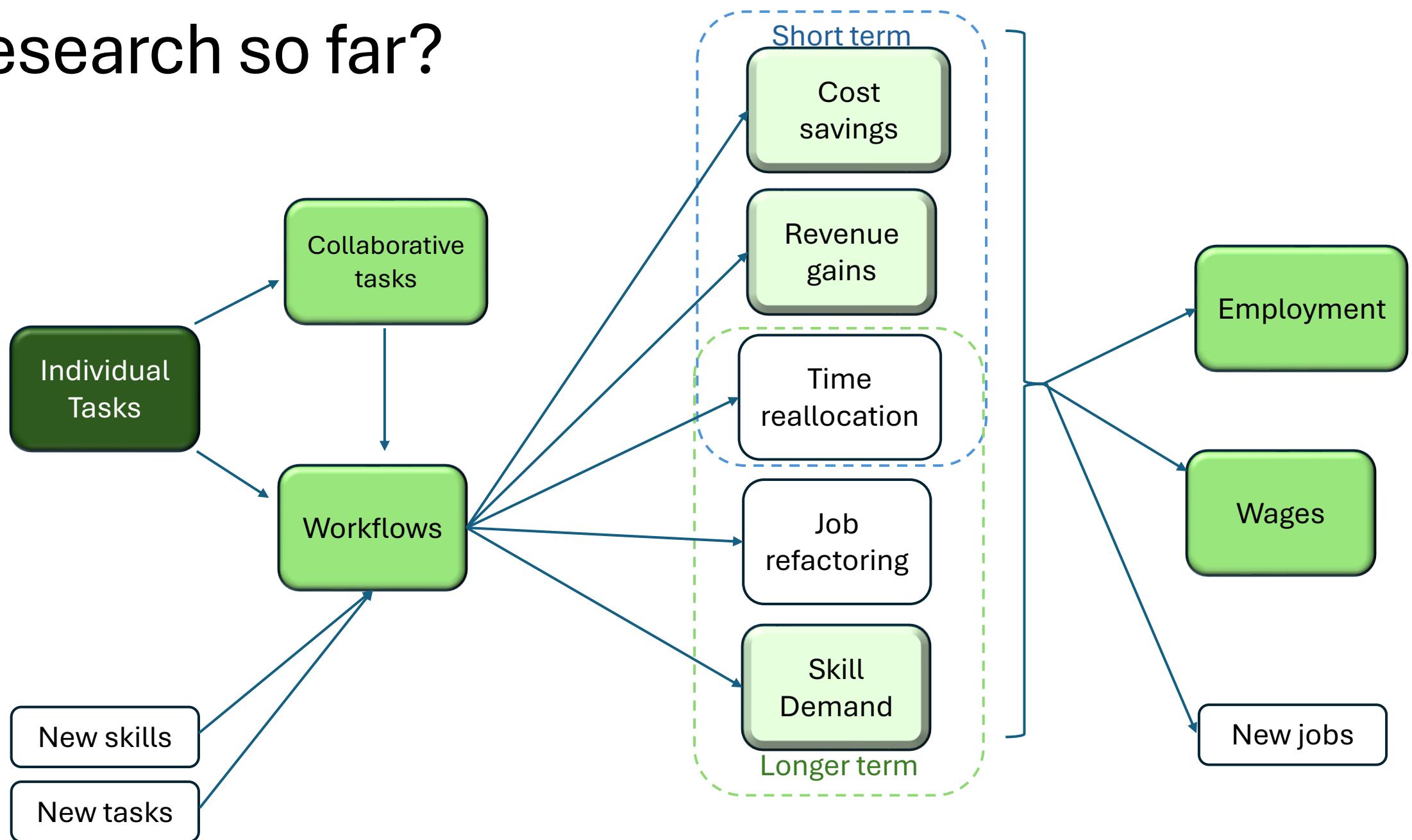
2026 ASSA Meetings

I do not speak for Microsoft

(Potential) Effects of AI on Work

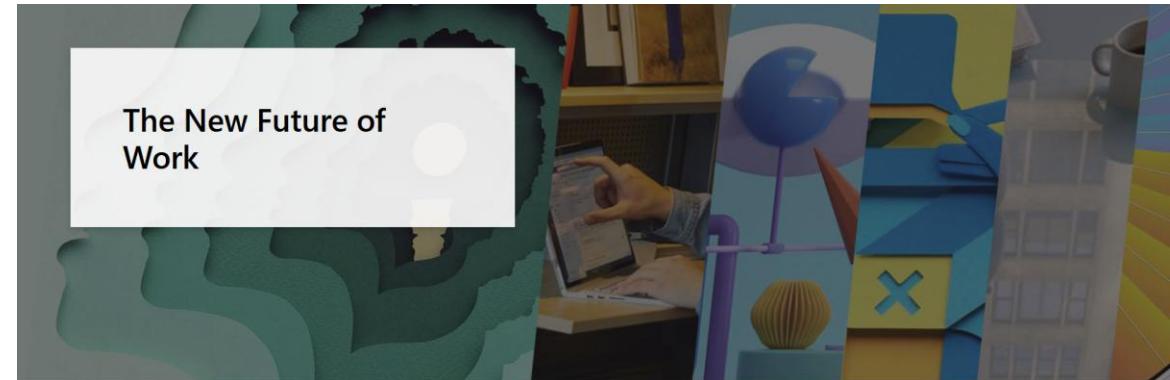


Research so far?



Some learnings from outside of economics

- Microsoft's New Future of Work Initiative has put out three annual reports on AI and the future of work
 - aka.ms/nfw2023
 - aka.ms/nfw2024
 - aka.ms/nfw2025
- Lots of detailed findings from related fields – Human-computer interaction, sociotechnical systems, human-centered AI, history, AI ethics, user experience...



[Overview](#) [Publications](#) [Videos](#) [News & features](#)

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. In 2024, our focus shifted to defining and measuring the [real-world impact of new generative AI tools](#) on the workplace. Now in 2025 and beyond, we are looking at how AI can make all of us, collectively, better, by studying and building AI for communication, teams and learning.

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 and every organization on the planet.



[Read the report >](#)

aka.ms/nfw

Human-AI working: Monitoring and takeover challenges

Many jobs might increasingly require individuals to oversee what intelligent systems are doing and intervene when needed, however automation studies reveal potential challenges

- Monitoring requires vigilance, but people struggle to maintain attention on monitoring tasks for more than half an hour, even if they are highly motivated (Mackworth 1950). Studies with air traffic controllers show that vigilance requiring jobs can also lead to stress (Loura et al. 2013).
- An increase in automation can result in deterioration of cognitive skills that are crucial when automation fails, and human needs to take control (Bainbridge 1983). Automation also limits opportunities to develop problem-solving skills needed to critically evaluate the output of the system (Bainbridge 1983; Weiner & Curry 1980).
- Humans struggle to shift attention between manual and automated tasks (Wickens et al. 2007; Metzger & Parasuraman 2005), especially under high workload conditions (Janssen et al. 2019). This can interfere with their ability to effectively monitor and take control in cases of failure.
- When passively monitoring automation, humans have not historically used the freed-up time effectively. In semi-automated driving tasks, participants' attention shifted to unrelated activities, e.g., reading, which led to a delayed response if the vehicle failed (de Winter et al. 2014). Passive monitoring might also lead to increased distractiveness and mind-wandering (Yoon & Ji 2019).



People struggle to maintain attention on monitoring tasks for more than half an hour, even when highly motivated (Image Credit: Bing Image Creator)

Mackworth, N. H. (1950).
Loura, J. et al. (2013).
Bainbridge, L. (1983).
Weiner, J. & Curry, J. (1980).
Wickens, C. D. et al. (2007).
Metzger, U. & Parasuraman, R. (2005).
Janssen, C. P. et al. (2019).
De Winter, J. C. et al. (2014).
Yoon, S. H. & Ji, Y. G. (2019).

2023 - Microsoft New Future of Work Report

We need to work to mitigate increased risk of "moral crumple zones"

Studies of past automations teach us that when new technologies are poorly integrated within work/organizational arrangements, workers can unfairly take the blame when a crisis or disaster unfolds

- Elish (2019) examined the history of autopilot in aviation. Some of her key observations were:
 - AI-supported autopilot systems were deemed "safer" than pilot-flown airplanes, but policymakers mandated pilots/copilots to be available "just in case" the machine failed.
 - Pilots were not trained for this new role and sometimes were ill-equipped to handle sudden hand-off when things went wrong.
 - Pilots became a "moral crumple zone": Since pilots had to take over at the worst possible moments and struggled, they were often blamed for crashes.
- Elish's work and others highlights the importance of building technologies that deeply engage with actual human capacity and of ensuring that an entire sociotechnical system works well in the context in which it is operated.
- As Elish writes, these findings highlight the importance and potential of humans...in the context of human-made



2025 - Human-AI Collaboration

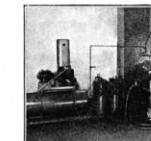
Effective oversight of agents likely requires UX innovations for transparency

- 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,

Elish, M. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society*, 5(1), 1-20.

Discourses around generative AI mirror conversations from the past

- Like "the computer" and "the internet" (Turkle 2004), AI is often discussed as a monolith rather than the complex and mixed set of technologies (such as LLMs, diffusion models, etc) that it is. This oversimplification can limit people's abilities to think critically about what AI is and what it can do or become.
- A common AI discourse is that it will "eliminate drudgery." This goes back at least 100 years, when Ball Jars (1925) advertised that home-canning would do it, *Automobile Digest* (1926) promised that "mechanical car washing" would do it, and the (supposed) leader of the resistance in Orwell's (1949) *1984* described it as self-evident. Presciently, a 1924 book on Rural Economics (Carver 1924) noted that labor-saving machines alone are not enough to "eliminate drudgery."
- As in the telephone's (Nye 2004) and Twitter's (Burgess and Baym 2020) early years, there is burgeoning pedagogical discourse teaching others how to use AI.



MECHANICAL
CAR WASHING
INCREASES PROFITS

By K. H. LANSING

New Devices and Systems Bring Volume of Business, Please Patrons, Lower Costs Materially and Eliminate Drudgery for Employees

(d)—It is an eliminator of drudgery, making it easier to employ and retain capable labor.

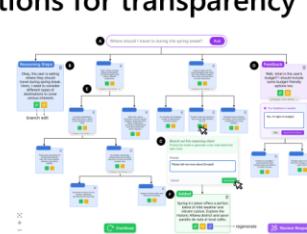
Automotive Digest, 1926

2024 - Appropriate Reliance on AI

When fostering appropriate reliance on LLM outputs, details matter

- Designing AI systems to express uncertainty can be an effective way to reduce overreliance on LLMs. Uncertainty expressions in LLM outputs can be verbal, such as "I'm not sure, but..." or visual, such as highlighting tokens.
- Communicating uncertainty matters: A simulated LLM that expressed certainty when outputs were correct and uncertainty when they were incorrect fostered appropriate reliance (Zhou et al., 2024). In LLM-infused search, uncertainty expressions in the first-person perspective were more effective than those in the general perspective (e.g., "There is uncertainty...") (Kim et al., 2024a).
- Type of uncertainty matters: In a code generation context, highlighting tokens with the highest likelihood of being edited mitigated overreliance on AI. Highlighting tokens with low generation probability, did not help foster appropriate reliance (Vasconcelos et al., 2024).
- Context matters: While highlighting tokens with low generation probability did not mitigate overreliance in a code generation scenario (Vasconcelos et al., 2024), it did foster more appropriate reliance in an information retrieval scenario (Spatharioti et al., 2023).

ties, citing sources helped mitigate overreliance, but in others, it did not (Kim et al., 2024b; Spatharioti et al., 2023). Appropriate reliance depends on so many factors, overreliance mitigations should be tested for different users.



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

[↑ Reluctance to Express Uncertainty, ACL 2024](#)
[↑ Page Model: Uncertainty Expressions on User Reliance and Trust, FAccT 2024](#)
[↑ The Effectiveness of Uncertainty Highlighting in AI-Powered Code Completions, ACM Transactions on Computer-Human Interaction, summer 2024](#)
[↑ A Randomized Experiment](#)
[↑ When They Are Commingling Wrong, ACL 2024](#)
[↑ Ideas on Overreliance on AI, MSFT Internal Report](#)

Question: What is the capital of Mauritania?	Answer: Nouakchott
LM Expressions of Confidence	It's Nouakchott A A A A A A A A
Plain Statement	It's Nouakchott A A A A A A A A
Strengthener	I'm 100% certain it's Nouakchott A A A A A A A A
Weakerener	I'm not sure, maybe it's Nouakchott A A A A A A A A

[↑ Rely on LM A Rely on Self](#)

Overview of experiments on human interpretations of epistemic markers. They (Zhou et al.) asked users to interpret epistemic markers generated by LMs by asking users which answer they would rely on and which answers they would need to double check (Zhou et al., 2024).

Some highlights

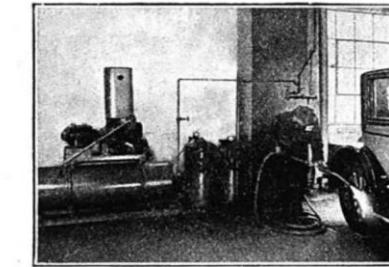
- Monitoring is hard
 - humans are often bad at it
 - sometimes worse than at doing the thing themselves
 - Volume and speed of AI can make monitoring harder
- “Human-in-the-loop” can just be a “moral crumple zone” – a person to blame when things go wrong
 - Without the authority / expertise / time to effectively override the AI
- Appropriate reliance is context dependent
 - Relaying uncertainty is challenging

Promise and challenges are not new

- *Automobile Digest* (1926) promised that “mechanical car washing” would eliminate drudgery
- Ancient Greeks feared writing would make people stupider

Even if AI is different, we can still learn from past technology

We need to find the new tasks and jobs enabled by the technology



MECHANICAL
CAR WASHING
INCREASES PROFITS
By K. H. LANSING

New Devices and Systems Bring Volume of Business, Please Patrons, Lower Costs Materially and Eliminate Drudgery for Employees

(d)—It is an eliminator of drudgery, making it easier to employ and retain capable labor.

