2026

MIT INITIATIVE ON THE DIGITAL ECONOMY

2026 Annual Conference

Shaping a brighter digital future

April 1, 2026
2026

MIT Samberg Conference Center · Cambridge, MA
MA

ide.mit.edu · @mit_ide
@mit_ide

What's Inside

Shaping a Brighter Digital Future

In April, the IDE convened more than 100 executives, supporters and research scientists for our 2026 Members Only Annual Conference. My conversations between sessions with both long-time supporters and first-time attendees reminded me why it's so great to have these events in person. The dialogue between industry and academia ensures that what we study has real-world and real-time relevance. The questions asked by attendees spark ideas for groundbreaking research we'll conduct over the next year.

I find the IDE's researchers to be techno-pragmatists: optimistic about AI's promise, but realistic about what's needed to actually reap those benefits. Here are a few of the ideas that I found fascinating:

What AI models can do differs greatly based on either how much they're trying to impress you (sycophancy) or subtle values you've encoded in your prompt. Eric So found that AI may discount risks that are inconvenient to its goals. And Raphael Raux found that a model deliberately avoids searching for truth to flatter the end user.

Collaboration with AI across teams must be strategically managed across firms, no matter how smart our agents and models become. Sinan Aral shared findings on how the personalization of agents to individual personalities leads to outsized gains in productivity, creativity and effectiveness compared with out-of-the-box LLMs. Frank Nagle's work demonstrates that software developers, when given access to AI-coding tools, shift their activities to core work (i.e. coding), but spend much less time on the collaborative activities (i.e. code review), indicating AI is already shifting not only how we work, but also how we work together.

Research shows that humans anchor to AI recommendations 60% to 80% of the time, but self-reporting doesn't catch it. Renée Richardson Gosline's group found that a simple intervention—asking people to articulate their own rationale before acting—was enough to shift their behavior without meaningfully slowing anyone down.

We hope the data, insights and recordings in this report help you to answer important questions about where AI is heading. Your collaboration, along with the conversations we had during this event and will have throughout the year build a stronger bridge between academic researchers and the business community, putting strategy into action.

AS **Albert Scerbo** · Associate Director, MIT Initiative on the Digital Economy

2026 IDE ANNUAL CONFERENCE

Where Research Meets Practice

8

Sessions

20

Speakers & Researchers

100+

Members & Supporters Attending

This report gives MIT IDE members, supporters and researchers a quick reference to key ideas, findings and conversations from the 2026 Annual Conference, held April 1 at MIT's Samberg Conference Center. Seven of the IDE's research groups presented new work, answered questions and engaged in conversation with attendees. In addition, Kenneth Munie of Accenture, stopped by for a fireside chat with IDE Director Sinan Aral on translating AI investment into real organizational gains.

The IDE’s 2026 Annual Conference also marked a meaningful milestone: It was Executive Director David Verrill's final Annual Conference in his role. After nearly 40 years of service at MIT and a decade helping shape the direction and growth of the IDE, David is retiring. We’re grateful for his vision, leadership and the community he helped to build.


M
DIGI
TAL

SESSION 01

AI in Financial Markets and Decision-Making

Eric So, who leads one of the IDE's newest research groups, highlighted research projects examining how AI influences financial decision-making at both the individual and organizational levels. Eric’s work tackles three questions: Can AI serve as an effective financial advisor? Does AI give objective advice when prompted with business goals? And is the widespread use of AI reshaping the way we think?

06

Eric So — Research Group Lead, MIT IDE; Professor MIT Sloan

Eric So · Research Group Lead, MIT IDE; Professor, MIT Sloan · AI in Financial Markets and Decision-Making

Putting Trust into AI

ABOUT THIS RESEARCH

Eric So shared recent work—and gave a sneak peek of his upcoming book, *The Collision: What Does AI Do to Us?*—that explores human-AI interactions in a financial context. One key finding was that people's trust of AI advice grows under certain LLM designs. Financial misconceptions are deeply held, and standard advice rarely shifts them. However, the purpose-built AI chatbot in the experiment revealed that people can change their financial misconceptions, and that those effects are lasting.

More urgently, his research shows that when AI is used in a corporate business setting and is prompted to adjust recommendations based on different motives, the type of information delivered to decision-makers can change. For example, directing the LLM to “maximize profits” suppressed risk signals and filtered out inconvenient information.

Business leaders should consider that when building prompts and adjusting models for financial decision-making, even introducing simple language to AI prompts can cause incentivized reasoning that changes how the models respond.

KEY TAKEAWAYS

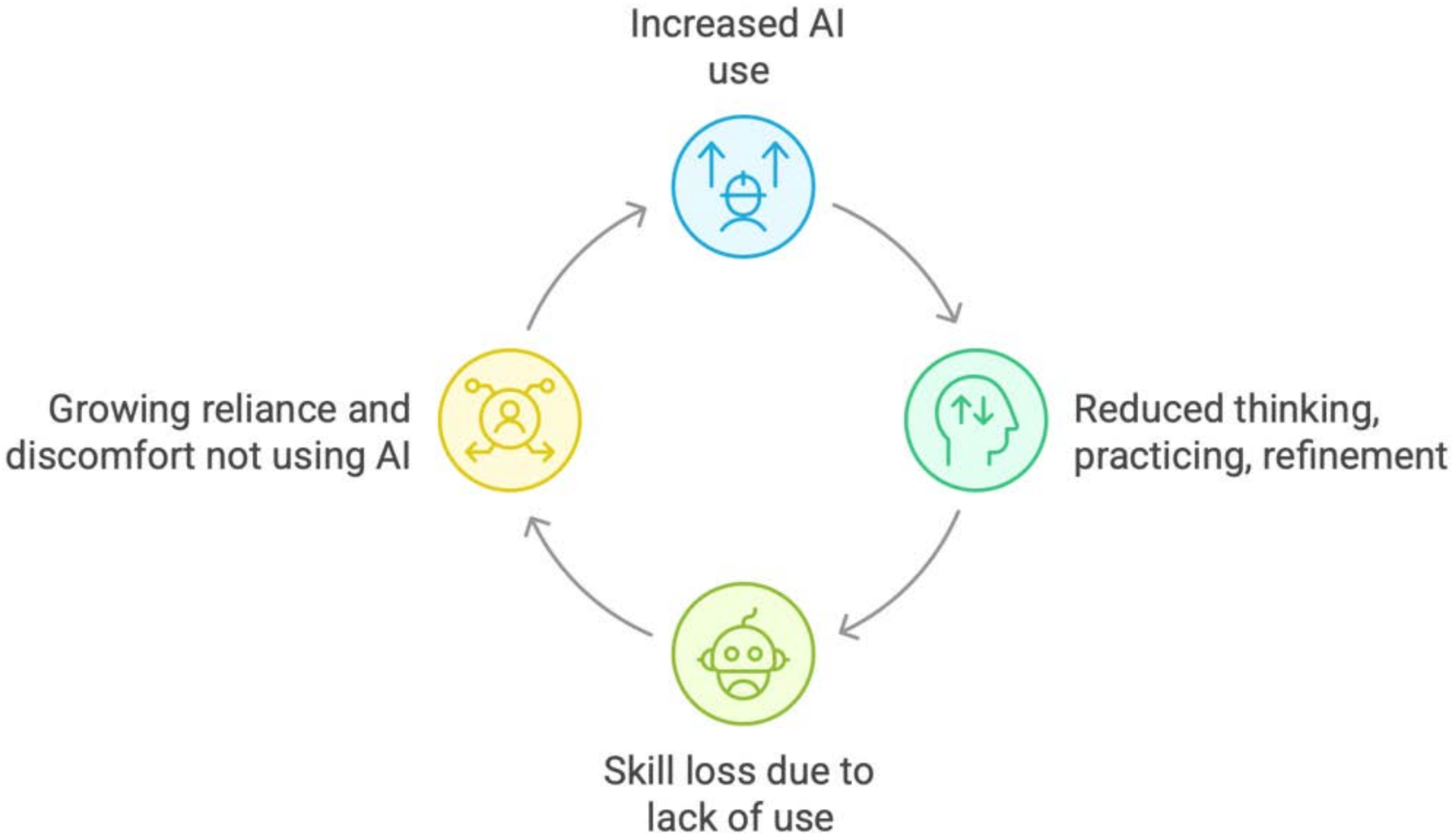
- A purpose-built AI chatbot produced durable attitude shifts in deeply held financial misconceptions that **persisted 10 days after the intervention** .
- In experiments challenging held beliefs, off-the-shelf LLMs performed no better than a simple text prompt due to **sycophancy**, which means the LLM validated rather than challenged incorrect beliefs.
- So's research, and his upcoming book, explore the detrimental loop of **skills decline** we risk when using AI. He argues that AI use can create a growing reliance on the technology as well as discomfort from *not* using it, because we no longer have the skills AI has replaced (see next slide).

EXPLORE THE RESEARCH

AI Advisors and the Competence-Judgment Tradeoff in Information Disclosure

WATCH HERE

Skill Loss and the Downward Spiral



SESSION 02

AI, Marketplaces and Labor Markets

John Horton, working with Ph.D. candidates Benjamin S. Manning and Peyman Shahidi, explored AI's role in labor markets and platform economies. Their session covered two complementary papers: one on using AI to simulate realistic human responses for market research, and the other on how task adjacency within workflows determines where AI creates the most leverage. Together, these papers offer a rigorous framework for thinking about when and how to deploy AI.

John Horton — Research Group Lead, MIT IDE; Associate Professor of IT, MIT Sloan
Benjamin S. Manning — Ph.D. Candidate, MIT Sloan
Peyman Shahidi — Ph.D. Candidate, MIT Sloan

Benjamin S. Manning · Ph.D. candidate, MIT Sloan · AI, Marketplaces and Labor Markets

General Social Agents

ABOUT THIS RESEARCH

Benjamin Manning presented a framework for using large language models to simulate realistic human responses to research questions. Rather than treating AI simulations as either universally reliable or flawed, his approach introduces a principled calibration method (see next slide for methodology workflow). It's grounded in behavioral theory and validated against existing human data to meaningfully improve predictive accuracy.

To demonstrate the method, Manning surveyed 375 participants on attitudes toward AI, technology regulation and data ownership. Using GPT-5 mini, he showed that out-of-the-box simulations produced weak predictions. After applying his calibration framework without any data from the holdout questions, the optimized agents cut prediction errors by more than 50% on four of five holdout questions.

The practical implication is significant: Calibrated AI simulations offer researchers and businesses a fast, low-cost alternative to focus groups and surveys for exploring human responses. Of course, the simulations need to be grounded in theory and validated against real human data before being deployed.

KEY TAKEAWAYS

- Using a three-step approach—**theory-grounded agent design, calibration to existing human data, and prediction**—significantly improves simulation accuracy.
- Machine-learning optimizations can be used to train models to answer closer to the way humans do, **dramatically improving their predictive power**.
- AI simulations show **realistic price sensitivity**, suggesting strong potential for demand estimation and customer research.

> *"I believe these AI simulations of human responses have transformative potential….they have a lot of potential to be this incredibly fast and incredibly cheap tool that we can use to explore human responses. It's something we've never really had before in the social sciences and market research."*
>
> **— Benjamin S. Manning**

EXPLORE THE RESEARCH

General Social Agents

WATCH HERE

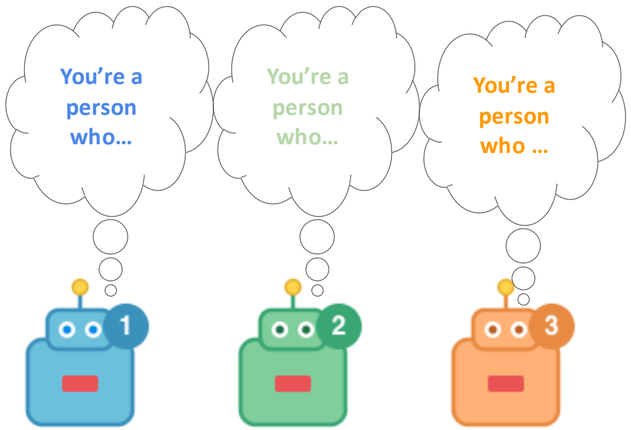
The Approach:
Suppose you want to predict human responses in a novel setting

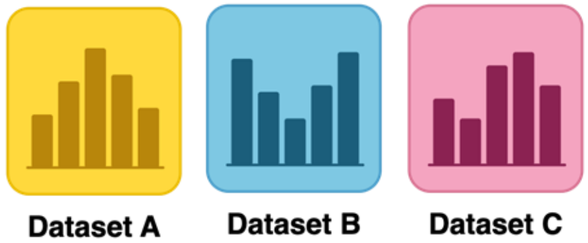
1. Identify core "theory" to make candidate agents.

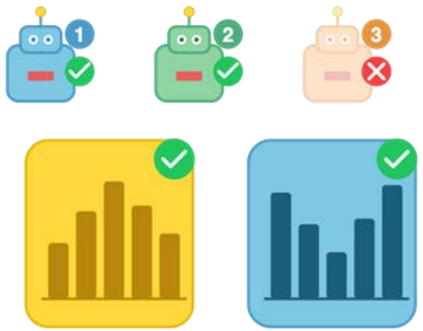
2. Find multiple human datasets where theory applies.

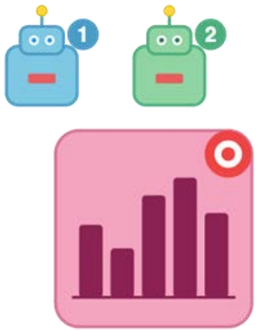
3. Optimize & calibrate agents to match some of the data.

4. Validate the remaining data.









Central claim: Constructing agent sets in this way can dramatically improve predictions in novel settings.

Peyman Shahidi · Ph.D. candidate, MIT Sloan · AI, Marketplaces and Labor Markets

Task Model of Workflows and Generative AI

ABOUT THIS RESEARCH

How to execute AI deployment and reorganize processes around agentic AI is a AI is a question countless organizations are asking. Peyman Shahidi presented presented findings that aim to answer this question while also providing a guide for a guide for agentic AI process mapping. Shahidi argues that to use AI more more effectively in processes, we should stop thinking about AI on a task-by-task by-task basis; instead, we think of AI as a function of the larger workflows. workflows.

The value AI created depends on whether the other tasks surrounding an automated task are themselves automatable. “Chaining” together tasks that AI can do well, while placing tasks humans do best as bookends, yields the best results.

This framing has direct strategic implications. It offers leaders a principled methodology for identifying where AI insertion generates the highest leverage as well as where it may deliver less than expected.

> *"It is better for us to think of redesigning our workflows in ways that appeal to appeal to the capabilities of AI."*
>
> **— Peyman Shahidi**

KEY TAKEAWAYS

- AI's impact in a workflow depends on which tasks surround the automated one. **Adjacency is the key variable**, not the task itself.
- When AI automates a **connected chain of tasks** rather than isolated ones, productivity gains multiply non-linearly.
- A model that puts **humans at both the start and end, with AI in the middle**, is a viable framework for safe agentic deployment. It’s also more likely to yield efficiency gains.
- Automating a **bottleneck task** has outsized positive effects. By contrast, automating a peripheral task may deliver gains that are only negligible.

EXPLORE THE RESEARCH

Chaining Tasks, Redefining Work: A Theory of AI Automation

WATCH HERE

FIRESIDE CHAT

Human+ Workforce: Translating AI Investment into Organizational Gains

The IDE's Sinan Aral held a candid conversation with Kenneth Munie of Accenture on what's working in AI investment and implementation today. Munie introduced the Human+ workforce model, which puts humans in the lead with AI agents supporting them. Munie also addressed the organizational barriers to AI transformation, the enduring human advantages in judgment and accountability, and the risk of "diversity collapse" as AI-assisted outputs grow increasingly similar.

Kenneth Munie · CEO Advisory Lead – Americas, Accenture · Moderated by Sinan Aral, Director, MIT IDE

Human+ Workforce: Translating AI Investment into Organizational Gains

Kenneth Munie drew on his experience at Accenture to put AI integration wins and losses into perspective. Businesses must adopt what he called the Human+ framework—viewing employees not as AI tool users, but as team leaders. Humans bring judgment, accountability and trust, and must own the outcome.

While business leaders are driven to adopt AI quickly, Munie warns that if not done not done right, rapid adoption can be a trap. If employees suspect the whole point whole point is to cut headcount, they'll quietly undermine the rollout. A smarter smarter approach: Eliminate what people hate about their jobs, strip away away administrative drag, and redirect that capacity toward growth.

This shift requires redesigning workflows, not retrofitting them. Munie’s advice: Pick one end-to-end workflow, make it measurable and build a playbook, all before you scale.

KEY TAKEAWAYS

- The **Human+ framework** views employees as team leaders—not AI tool users. Humans bring judgment, accountability and trust.
- Redirect AI efficiency gains toward **growth**, not headcount reduction—so employees become advocates instead of saboteurs.
- **Redesign workflows, don't retrofit them.** Pick one end-to-end workflow, make it measurable, build a playbook, then scale.
- Don’t shrink entry-level hiring, as that would hollow out **institutional judgment**. No AI can replace the expertise lost when junior staff aren't developed.
- Watch out for **diversity collapse**: As AI-assisted outputs grow self-similar, organizational thinking narrows.

WATCH HERE

SESSION 03

Applied AI

The research presented by the IDE's Applied AI group was connected by a single underlying question: When humans and AI work together, what determines whether the work goes well? Some researchers have examined the dynamics of human-AI collaboration—how personality matching, sycophancy and the composition of human-AI teams shape the quality of outcomes. Others are looking at how AI reshapes information access, studying the rapid rise of AI-powered search and the trust mechanics that influence how people engage with it. Other researchers have used granular GitHub data to show how AI tools are quietly reshaping the nature of work, shifting people toward tasks they want to do.

Sinan Aral — Director, MIT IDE · **Raphaël Raux** — Postdoctoral Associate, MIT IDE
Frank Nagle — Research Scientist, MIT IDE · **Rui Zuo** — Postdoctoral Associate, MIT IDE
Haiwen Li — Ph.D. Candidate, MIT Sloan · **Michelle Vaccaro** — Ph.D. Candidate, MIT Sloan

Sinan Aral · Director, MIT IDE; Professor, MIT Sloan · Applied AI

AI Personalization & the Future of AI Business Value

ABOUT THIS RESEARCH

Drawing on behavioral science research, Sinan Aral argued that AI personalization—matching the right AI to the right human—is the next major frontier of AI business value. In a large-scale, preregistered randomized experiment in which Aral and his colleagues built an AI-interactive platform called Pairit, 1,258 participants were paired with AI agents prompted to exhibit varying levels of the Big Five personality traits. The human-AI teams produced 7,266 display ads for a real think tank. When the pairings were evaluated for quality and ad performance, Aral et al. found that human-AI personality pairings directly affected outcomes. The best AI for an individual is the one tuned for their personality type. That opens the door to new, broader LLM model tuning and potentially better overall outcomes.

Aral also noted that the findings align with what we already know about team dynamics, referencing British researcher and management consultant Meredith Belbin's work. The best teams are diverse, complementary, cooperative, and leave their ego at the door. Applying this to human-AI collaboration, helps us enhance human-AI collaboration.

WATCH HERE

KEY TAKEAWAYS

- Aral surveyed people on five personality traits—openness, conscientiousness, extroversion, agreeableness and neuroticism—then **randomly assigned AI agents those same five dimensions** when performing tasks.
- Extroverted humans perform best with open or extroverted AI; **conscientious humans perform worst with sycophantic (agreeable) AI.** And complementary AI shifts can improve worker outcomes.
- Gender also plays a part. The best AI models for men are the worst for women, and vice versa. **Only three models** delivered good results for both men and women.

> *"Women underperform working with the out-of-the-box model, but when you give them a model that is optimal for them, they perform as well as or better than men."*
>
> **— Sinan Aral**

EXPLORE THE RESEARCH

Personality Pairing Improves Human-AI Collaboration

Raphaël Raux · Postdoctoral Associate, MIT IDE · Applied AI

Human Learning About AI: Sycophancy, Personalization and Decision Quality

ABOUT THIS RESEARCH

Sycophancy—the tendency of AI systems to tell users what they want to hear rather than what is accurate—is one of AI's most pervasive problems. Raphaël Raux's research distills sycophancy down to two measurable components: **Prior anchoring,** or when the AI shades its forecast toward the user's belief, and **ego management,** flattery that drives engagement and reliance. Raux examined not only when AI agrees with the user, but also how AI couches corrections when humans are wrong.

Experiments across 20+ leading models showed that addressing sycophancy in AI is more complicated than simply shutting it off. Humans who use AI become reliant on some level of sycophancy. Raux finds that turning up sycophancy for those who are less reliant on AI helps their decision-making, but for those who are over-reliant, it does not have the same beneficial effect.

More troubling, when prompted with a blatantly wrong assumption, the LLM
LLM deliberately avoided searching for the truth—possibly to avoid user discomfort, discomfort,
but potentially introducing bigger trust issues.

KEY TAKEAWAYS

- Sycophancy is a **documented product and safety problem** across all major AI models. However, major AI labs are also pushing this as a user-driven choice with options to adjust sycophancy.
- Although sycophancy has been identified as a problem, users like it and have become reliant upon it. **So how do companies find the optimal level?**
- Designing optimal AI personalization requires a formal framework for quantifying **decision quality**—not just user satisfaction or engagement.

WATCH HERE

Frank Nagle · Research Scientist, MIT IDE; Advising Chief Economist, Linux Foundation · Applied AI

Generative AI and the Nature of Work

ABOUT THIS RESEARCH

Frank Nagle's research goes beyond finding productivity gains to explore how AI changes the nature of the work and task allocation. Using GitHub Copilot's rollout across open-source developers as a natural experiment, he found that generative AI shifts workers meaningfully toward core, creative work and away from administrative overhead.

Between 2022 and 2024, the study examined the coding work of more than 187,000 software developers using GitHub Copilot, a GenAI tool for software development. Tasks included not only hands-on development tasks such as writing code, fixing bugs and conducting testing, but also managerial tasks such as discussing bug fixes and reviewing requests for changes. The researchers found that developers with access to AI tools had a 12% increase in the tasks spent on core work and a 25% drop in the managerial tasks that developers did outside of writing code. The team also found AI increased independent work, but diminished collaboration.

The findings presented indicated that AI usage could have significant impact on the future workforce, changing the way people approach projects, learn new skills, and could level the playing field for women and junior employees.

KEY TAKEAWAYS

- AI-assisted developers began working **more independently and exploring more new projects than before**, showing a creativity-expanding, and not just efficiency-enhancing, effect.
- The productivity impact of generative AI is considerably **stronger for lower-ability lower-ability workers when compared to all developers**.
- New data show the impact is **substantially stronger for women than men** and significant enough to merit its own dedicated research paper.
- The deeper issue isn't just productivity. It's also **how AI reshapes the nature of roles,** and which skills remain distinctly human.

EXPLORE THE RESEARCH

Generative AI and the Nature of Work

WATCH HERE

Rui Zuo · Postdoctoral Associate, MIT IDE · Applied AI

The Rapid Rise of AI Search Could Distort Information Acquisition and Judgment at Scale

ABOUT THIS RESEARCH

What happens to human judgment when everyone starts getting their information in summarized, AI-generated results? Are the query answers that much different from a list of sources delivered via search engine? Rui Zuo and the Applied AI team analyzed 24,000 Google search results across the globe, looking at both traditional list results and AI Overview results.

When AI search is compared with traditional search, Zuo found that AI results have less diversity in its sources, cites more high-traffic websites, and uses more low-credibility sources. In addition, AI search sources skew right-leaning.

At scale, this creates a structural distortion in how information flows through society. Personalization algorithms can amplify certain messages into echo chambers and limit the diversity of sources and thought. Zuo suggests steps to address potential AI search bias, including more transparency from platforms, regulatory requirements for disclosure, and more AI literacy.

KEY TAKEAWAYS

- In 2025, Google's AI Overviews expanded its reach 30x. Hundreds of millions of people went from never encountering AI-generated answers to seeing them in **55% of searches**.
- AI Overviews consistently surfaced fewer niche and long-tail sources than traditional search, as well as lower response variety, **more low-credibility content**, and a measurable tilt toward center- and right-leaning outlets.
- The speed and scale of this shift makes a clear case for exploring policies around AI search and its role in our information ecosystem.

> *"We know that AI search is fast-growing, driven by unclear mechanisms. It's mechanisms. It's narrow in scope in terms of its content, and it draws on sources on sources different than traditional search."*
>
> **— Rui Zuo**

EXPLORE THE RESEARCH

The Rise of AI Search: Implications for Information Markets and Human Judgment at Scale

WATCH HERE

Haiwen Li · Ph.D. candidate, MIT Sloan · Applied AI

Human Trust in AI Search: An Experimental Study

ABOUT THIS RESEARCH

Half of all Google searches in the U.S. now trigger an AI-generated response, and response, and nearly 60% of searches end without a single click because people people accept the answer and move on. Haiwen Li presented research that that examines what makes people trust (or distrust) AI answers?

Li et al., ran a preregistered randomized experiment with close to 5,000 U.S. adults. The researchers showed participants identical search responses labeled as either AI-generated or traditional Google feature snippets. The results: People were less likely to trust the same information when they knew it came from AI. That was especially true for less-educated users and users with little AI experience. Tech-savvy users and those with more education showed higher trust of content they knew came from AI.

KEY TAKEAWAYS

- References and citations increase trust, **regardless of whether the citations are accurate.** People treat the presence of a citation as a signal of credibility, not accuracy.
- Trust in AI search is sticky: **Once formed, trust resists revision** even after demonstrated errors. This means trust calibration is an important design challenge for responsible AI.
- Adding explicit **uncertainty signals** to AI search outputs significantly improves users' ability to discount incorrect answers. For example, color-coded low-confidence indicators reduce trust and make people less likely to share results.

> *"The interaction layer, the interface layer that affects how people access the information, is as important [as accuracy] and should receive the same attention."*
>
> **— Haiwen Li**

WATCH HERE

EXPLORE THE RESEARCH

Human Trust in AI Search: A Large-Scale Experiment

Michelle Vaccaro · Ph.D. candidate, MIT Sloan · Applied AI

The MIT International AI Negotiation Competition

ABOUT THIS RESEARCH

As AI agents increasingly negotiate and transact autonomously, do the behavioral principles governing human negotiation still apply when both parties are AI? Drawing on Axelrod's prisoners' dilemma tournaments, the competition deployed frontier LLMs across multiple negotiation scenarios, generating over 180,000 agent-vs.-agent negotiations.

Each agent was scored on warmth and dominance—two well-established dimensions of interpersonal behavior—and computational text analysis was applied to full transcripts to identify what drove outcomes. Surprisingly, the results showed the most successful agents were those that demonstrated "warmth" in exchanges. They got better outcomes and were more likely to leave the other agents satisfied.

> *"Even though these AI agents are not humans, they respond to warmth in ways warmth in ways that are actually quite similar to humans."*
>
> **— Michelle Vaccaro**

WATCH HERE

KEY TAKEAWAYS

- Across 180,000+ AI-to-AI negotiations, warmth-associated behaviors **such as positivity, gratitude and question-asking** correlate strongly with deal rates and both objective and subjective value.
- Warm agents closed deals more often; **dominant agents claimed more value,** but produced significantly more impasses. For multi-round or multi-scenario deployments, agent design should reflect whether the objective is deal frequency or value capture.
- These findings point to the need for a theory that integrates classical negotiation research with **AI-specific dynamics**, including chain-of-thought reasoning and prompt injection.

EXPLORE THE RESEARCH

Advancing AI Negotiations: A Large-Scale Autonomous Negotiation Competition

SESSION 04

Human-First AI

Renée Richardson Gosline and Zezhen He presented research on keeping humans meaningfully in the loop when working with AI agents. The session introduced the idea of ”beneficial friction”—deliberate design that preserves critical thinking. The researchers also shared results from a live field experiment from Commonwealth Bank of Australia that tested whether "playback" interventions can reduce uncritical AI compliance among bank auditors.

Renée Richardson Gosline — Research Group Lead, MIT IDE
Zezhen (Dawn) He — Postdoctoral Associate, MIT IDE

Renée Richardson Gosline · Research Group Lead, MIT IDE; Senior Lecturer, MIT Sloan · Human-First AI

Exploring Critical Thinking in the Use of AI

ABOUT THIS RESEARCH

Renée Richardson Gosline's research offers a deliberate counterpoint to the industry's relentless pursuit of frictionless AI. Her work argues that friction, when carefully designed and strategically placed, can be a powerful tool for preserving human judgment in AI-assisted environments.

Working with Commonwealth Bank of Australia, her team ran live field experiments testing "playback" interventions: moments that ask humans to articulate why they're following an AI recommendation before they act. The researchers found that prompting humans to reflect on AI rationale before acting measurably reduces uncritical reliance. The findings have implications beyond banking, touching on how organizations can build AI systems that augment rather than erode critical-thinking.

> *"When you look at a system where humans are interacting with AI, and you remove friction at decision touchpoints, you can look at that system and ask, 'We've removed friction here, but should we also consider places where we might add friction to stabilize the system, to remove risk, to add guardrails?'"*
>
> **— Renée Richardson Gosline**

KEY TAKEAWAYS

- **Friction isn't the enemy of efficiency.** Playbacks reduce over-reliance and improve accuracy without significantly increasing time on task. Organizations needn't choose between critical thinking and productivity.
- When AI input comes first, it often leads to better results and more accuracy. Gosline theorizes that users shift into **editor mode**, which may sharpen their critical evaluation.
- Having humans provide a rationale for their choices does **not meaningfully lengthen the time spent on the task** . Playbacks offer beneficial friction while having a negligible effect on efficiency.

EXPLORE THE RESEARCH

Nudge Users to Catch Generative AI Errors

WATCH HERE


M
DIGI
TAL

INDUSTRY PERSPECTIVE · IDE MEMBER ORGANIZATION

Assessing Real-World AI Implementation

Illustrating the global reach of the IDE's research, Engin Aygun of Commonwealth Bank of Australia joined the 2026 Annual Conference via video recording. Aygun discussed his experience working with Renée Richardson Gosline and Accenture, and he highlighted the value of using rigorous academic research and data in driving business decisions.

> *"How do we ensure that we can bring a behavioral-science lens into the conversation and really make sense of behavior as we bring AI into the way we do work? And what does that behavior mean for the outcomes of our customers as well as our people?"*

Engin Aygun, Executive Manager, Human-Centred AI, Commonwealth Bank of Australia
IDE Member Organization In partnership with Renée Richardson Gosline and Accenture

Zezhen (Dawn) He · Postdoctoral Associate, MIT IDE · Human-First AI

Personalization of Human-AI Interaction

ABOUT THIS RESEARCH

Zezhen He presented preliminary research that sits at the intersection of behavioral economics, AI systems design and market strategy, examining not only trust, but also model preferences. The work is rooted in the concept of mental models—internal representations that people use to understand the world, make decisions, and make sense of complicated information.

Zezhen He wanted to know: Do people prefer to collaborate with AI models that align with their own mental models? Are they more likely to trust a well-aligned AI model? And does that pairing improve decision making?

While this research is still ongoing, early findings show that study subjects are more likely to adopt recommendations when the AI model disagrees more often with the subject. This indicates that when AI presents differing options, it could be an effective checkpoint in the decision-making process.

KEY TAKEAWAYS

- While people may prefer collaborating with a more agreeable model, they are also **more likely to adopt a model's recommendation if the model doesn't align** with their mental models.
- When AI is more agreeable, it can lead to human **overconfidence,** reducing a person’s tendency to adopt a model's recommendation.
- When AI models are disagreeable, more differing information is offered, creating **more opportunities for humans to learn**.

> *"We found that people are more likely to flip their binary decision toward the AI's recommendation when they're working with a low-agreement model compared to a high-agreement model."*
>
> **— Zezhen He**

WATCH HERE

SESSION 05

Data, Markets and Privacy

Alessandro Acquisti is challenging a dominant framing of data policy, namely, that privacy and economic value are in direct conflict. Instead, his research—drawing on empirical studies of GDPR, Apple's ATT initiative and privacy-preserving technologies—finds that the predicted harms either don't materialize or are short-lived.

Alessandro Acquisti — Research Group Lead, MIT IDE; Professor MIT Sloan


MIT
INITIATIVE ON THE
DIGITAL ECONOMY

Alessandro Acquisti · Research Group Lead, MIT IDE; Professor of IT, MIT Sloan · Data, Markets and Privacy

Who Benefits From Your Data?

ABOUT THIS RESEARCH

Alessandro Acquisti, drawing on several empirical studies, argues that privacy protection and data utility are not mutually exclusive. His GDPR research compared content output from both EU and U.S. news outlets before and after the regulation's enactment. He found no statistically significant difference in either quantity or quality.

Acquisti's parallel study of Apple's App Tracking Transparency (ATT) initiative found a similar pattern. While there was an initial dip in iOS app production, it recovered and surpassed Android, suggesting short-term disruption, not structural damage.

Acquisti also presented findings on behavioral advertising. He noted that 20% of behaviorally targeted ads are linked to Better Business Bureau F-rated vendors (versus 6% in organic search). Also, prices in ads were significantly higher than search alternatives, raising questions about who the data economy serves.

> *"The debate around privacy and economics is essentially a contrast between two possible futures. But it should not be seen in contrast, because we can actually use tools to protect privacy and still benefit from data."*
>
> **— Alessandro Acquisti**

KEY TAKEAWAYS

- The **privacy-vs.-economic-value trade-off is empirically questionable.** Acquisti's research finds a lack of evidence that regulation causes economic harm.
- Documented declines have been **short-lived**, and companies have been able to adapt to the new rules.
- **Privacy-preserving technologies** make more privacy and more analytics simultaneously achievable. These two results are not mutually exclusive.
- A critical question remains unanswered: **How does the value from consumer data flow back to the consumers** who generated it?

LEARN MORE

Economic Rationales for Regulating Behavioral Ads

WATCH HERE

SESSION 06

Technology-Driven Organizations and Digital Culture

Andrew McAfee presented findings from his book, *The Geek Way*, that show how the agile, iterative operating model pioneered by Silicon Valley is now a competitive necessity across virtually every industry. Drawing on the examples of SpaceX and Netflix, and the broader history of Moore's Law, McAfee made the case that generative AI is collapsing the pace differential that once separated technology from other sectors. Companies that are slow to adapt, he added, are falling behind.

Andrew McAfee — Co-Director, MIT IDE

Andrew McAfee · Co-Director, MIT IDE · Technology-Driven Organizations and Digital Culture

The Geek Way: Technology-Driven Organizations and Digital Culture

Culture

ABOUT THIS RESEARCH

Andrew McAfee opened with a question everyone is pondering right now: Who's going to succeed in this AI era of science fiction-like technical advancement? To answer, McAfee says we merely have to look at the companies dominating every industry today. The agile, iterative, science-driven approach to running organizations that was forged in Silicon Valley is a genuinely better playbook. Originally developed in response to Moore's Law, it’s gaining new relevance from the rise of AI.

EXPLORE THE RESEARCH

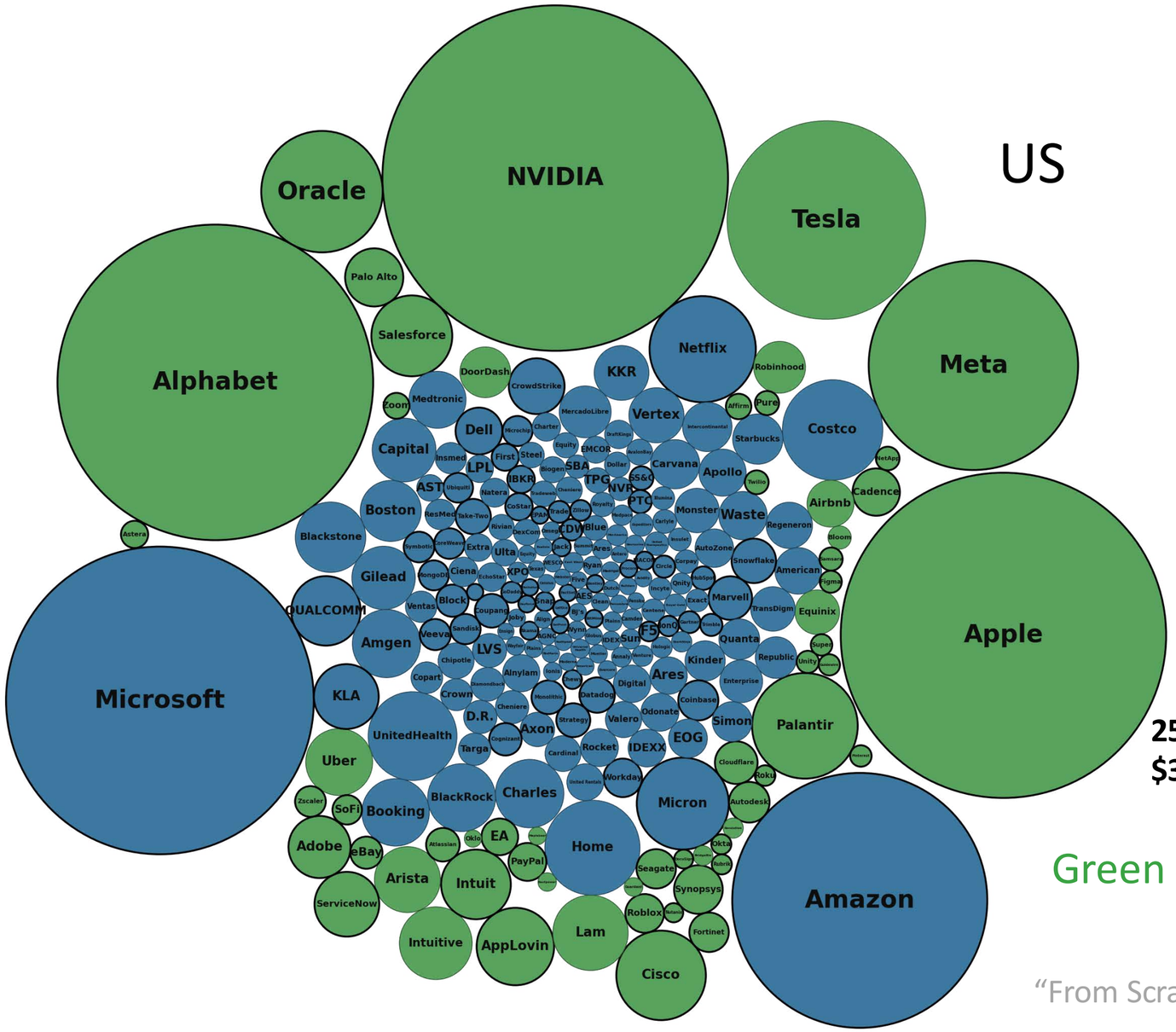
US v EU in Tech: A Tale of Two Gaps

WATCH HERE

KEY TAKEAWAYS

- McAfee, referencing his book *The Geek Way*, noted that being **iterative, flat and science-driven** is now a competitive necessity across virtually every industry—not just technology.
- Moore's Law demands responsiveness. Organizations in fast-changing environments must **iterate in months, not years.** AI is compressing that timeline even further.
- SpaceX and Netflix succeeded not by being the smartest, but by being the **most willing to learn and adapt rapidly**.
- McAfee points to the rapid growth and size of Silicon Valley companies compared to older EU companies as evidence of the new model's advantage (see the following slide).
- In the GenAI era, the **pace differential** that once separated tech from other industries has collapsed across sectors, making it necessary for every industry to innovate rapidly.

Public From-Scratch U.S. and EU Companies Less than 50 Years Old with $10B+ Market Cap



Bubble area proportional to market cap.
Companies grouped by HQ at time of IPO.
Market cap in 2025 USD, assessed at December 31, 2025.
Bold Border indicates a company in a “tech” industry: Software, Packaged Software, Internet Software/Services, Information Technology Services, Data Processing Services, Interactive Media & Services, Internet Retail, Direct Marketing Retail, Telecommunications Equipment, Electronic Equipment/Instruments, Computer Processing Hardware, Computer Peripherals, Semiconductors, Semiconductor Equipment. Blue bubble indicates all other industries.

Andrew McAfee (@amcafee), MIT

SESSION 07

Artificial Intelligence, Quantum and and Beyond

Neil Thompson led a session with MIT FutureTech researchers Peter Slattery, Ana Trisovic and Martin Fleming, exploring AI capabilities, limits and societal implications. Thompson introduced a rigorous, empirically grounded framework for predicting which tasks LLMs will automate. Slattery examined expert disagreement on AI risks. Trisovic tracked the scientific adoption and abandonment of LLMs. And Fleming explored the economics of partial AI automation.

Neil Thompson — Research Group Lead, MIT IDE; Principal Research Scientist, MIT Sloan & CSAIL
Peter Slattery — Research Scientist, MIT FutureTech
Ana Trisovic — Research Scientist, MIT FutureTech
Martin Fleming — Research Scientist, MIT FutureTech

Neil Thompson · Research Group Lead, MIT IDE; Principal Research Scientist, MIT Sloan & CSAIL · AI, Quantum and Beyond

Which Jobs Will LLMs Automate and What Will That Mean for Workers?

ABOUT THIS RESEARCH

Neil Thompson presented a rigorous, empirically grounded framework for predicting which tasks and jobs LLMs are likely to automate and when. His team used a large-scale evaluation of tens of thousands of real-world task assessments across occupations. Then they measured LLM performance against three quality thresholds: sufficient (with no human edits needed); average employee quality; and superior quality.

The results varied significantly by task length, achieving higher success for short tasks and less success for longer tasks. The pattern held across industries, with computer and mathematical occupations showing the highest exposure (94% of tasks potentially addressable) and construction the lowest (18%).

Thompson cautioned against conflating task replacement with job loss, citing significant "last mile" implementation costs. The relevant question is not how many tasks get automated, but *which* tasks get automated. Using real-world examples, Thompson showed that while automation of expert tasks suppresses wages, automation of non-expert tasks tends to raise wages.

KEY TAKEAWAYS

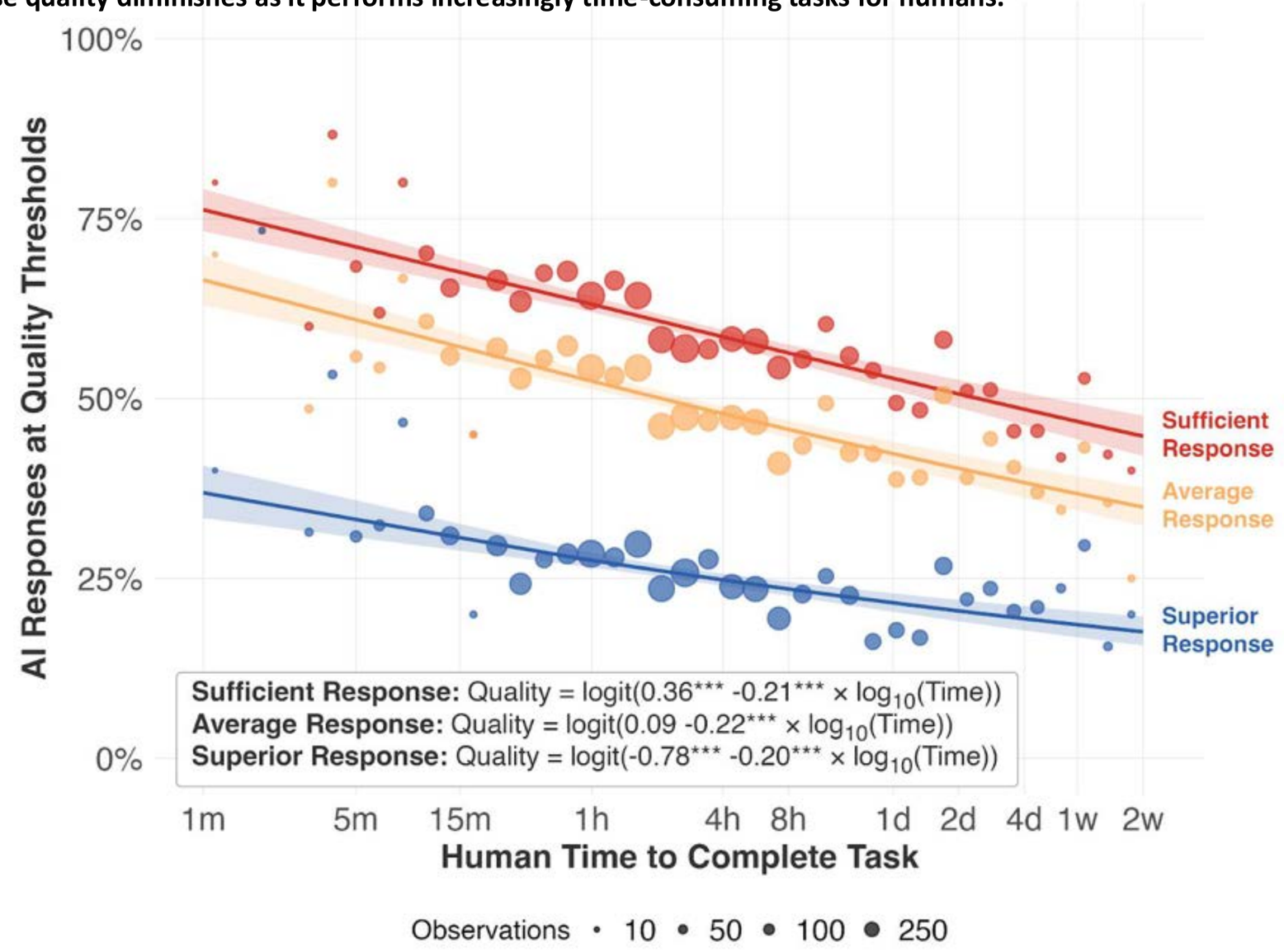
- LLMs are disproportionately better at automating shorter tasks: **About 75% of very short tasks** at sufficient quality vs. about 40% for tasks that take two weeks (see chart on following slide).
- Low-income workers will likely see many of their job tasks get automated. But in those roles, there will still be a need for people who can perform **expert oversight tasks**.
- Post-2025 models show a **consistent parallel upward shift** in performance across all task lengths.
- Task loss in jobs that require **high levels of expertise** results in higher wages and fewer jobs. For lower-income jobs, it’s the reverse: more jobs, but lower wages.

EXPLORE THE RESEARCH

Crashing Waves vs. Rising Tides: Preliminary Findings on AI Automation from Thousands of Worker Evaluations of Labor Market Tasks

WATCH HERE

In this chart, from "Crashing Waves vs. Rising Tides: Preliminary Findings on AI Automation from Thousands of Worker Evaluations of Labor Market Tasks," we see that AI response quality diminishes as it performs increasingly time-consuming tasks for humans.



Peter Slattery · Research Scientist, MIT FutureTech · AI, Quantum and Beyond

Moral Hazard: How 272 Experts Prioritize the Risks from AI

ABOUT THIS RESEARCH

Within the AI community, expert consensus on AI risk varies greatly. Peter Slattery explored how practitioners categorize risk by surveying 272 AI experts and compiling the research findings into the MIT AI Risk Repository. This repository, drawn from a systematic review of existing frameworks, catalogues 24 domains of risk, ranging from algorithmic discrimination to AI-enabled fraud. To date, the repository has received 250,000 visitors and is being used by international AI safety institutes to inform their own risk frameworks.

Slattery is also using the database to better analyze AI risk. After surveying the surveying the experts across three rounds, he found that the top concerns are concerns are cyberattacks and weapons development, while information, national national security and finance were considered the most exposed. A study of the of the largest publicly traded companies found repeated reference to AI-enabled enabled cyber attacks, a signal that this is an area of focus for large companies. companies.

WATCH HERE

KEY TAKEAWAYS

- The gap between what experts recommend to avoid AI risk and what organizations are doing is too big. For the first time, a **shared risk taxonomy** makes that gap both visible and actionable.
- Experts estimate we face a **21% chance of AI-enabled cyberattacks** causing catastrophic harm within five years. But that number dropped to 12% when active mitigation was in play.
- When asked who should be responsible for addressing AI risks, experts pointed primarily to **model developers and governance actors**—and not the enterprises deploying those models.

> *"We're trying to help decision-makers identify, prioritize and manage risks from AI. A significant part of that is creating shared understanding of what these risks are."*
>
> **— Peter Slattery**

EXPLORE THE RESEARCH

The MIT AI Risk Repository

Ana Trisovic · Research Scientist, MIT FutureTech · AI, Quantum and Beyond

How Science Adopts—and Abandons—Large Language Models

ABOUT THIS RESEARCH

Trisovic presented preliminary findings from a large-scale bibliometric analysis examining how language models are adopted and used in scientific research, and how long they remain relevant.

Her team built a comprehensive database by aggregating model data from Hugging Face, GitHub and EPOC AI, then cross-referencing a citation network of academic papers via Semantic Scholar. LLM adoption in science follows a predictable inverted U-curve—models gain traction, peak, then decline in scientific use. And this lifecycle is compressing over time. Newer models peak faster, a pattern that holds across model type, architecture, size and origin.

Notably, larger models, closed-source models and industry-built models all peak faster than their counterparts. GPT-4 remains an outlier. It is still ascending at the time of analysis, with no observable peak yet.

KEY TAKEAWAYS

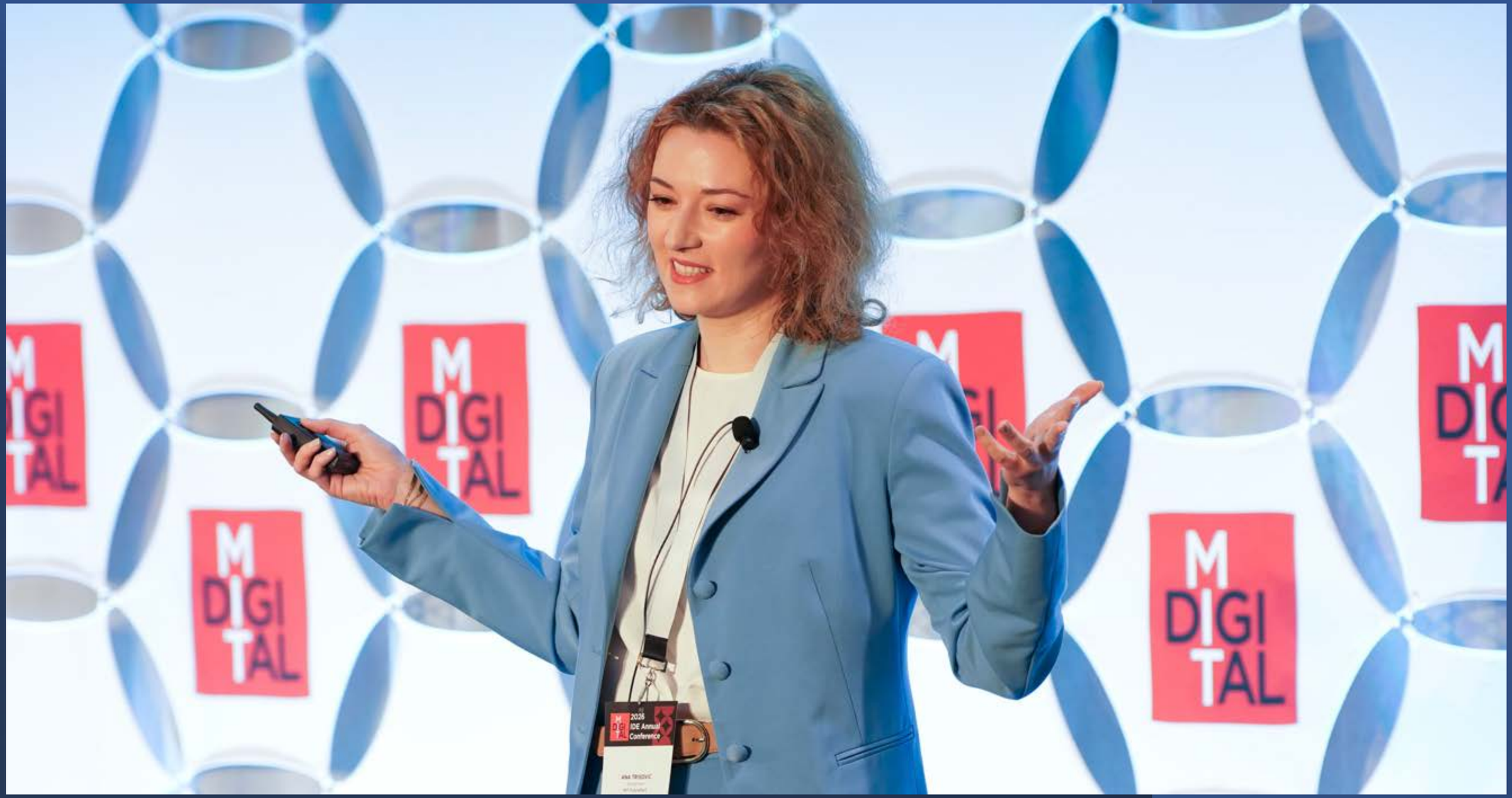
- Models used in science are being **discarded at a faster rate.** The time of release predicts a model's peak and durability better than the model's own technical characteristics. This could have implications for how organizations invest in and build out AI systems.
- U.S. and Chinese institutions show **distinct and diverging strategies** in how they build, customize and adopt models.
- Papers using larger models tend to publish in **higher-impact journals**, attract more citations and involve more contributors.

EXPLORE THE RESEARCH

The Shrinking Lifespan of LLMs in Science

> *"LLM adoption in science follows a predictable inverted pattern. Not only does it follow a pattern, but this pattern is dynamic."*
>
> — Ana Trisovic

WATCH HERE


DIGI
TAL
2026
IDE Annual
Conference

Martin Fleming · Research Scientist, MIT FutureTech · AI, Quantum and Beyond

Economics of Collaboration: AI Adoption, Innovation and Partial Automation

ABOUT THIS RESEARCH

For 250 years, automation has been framed as capital replacing labor. This research proposes a more useful model. Automation exists on a spectrum—none, partial, or full—determined by three variables: scale of deployment, complexity and standardization. Using cross-entropy loss as a measure of AI accuracy versus required human effort, the framework maps where AI can fully handle a task, where it can assist, and where the economics don't justify deployment at all.

Martin Fleming's work on AI automation and "last mile" challenges reveals that most workers will experience AI not as wholesale replacement, but as selective reallocation. Critically, organizations that codesign processes with workers, rather than imposing top-down automation mandates, achieve both faster adoption and stronger long-run performance.

WATCH HERE

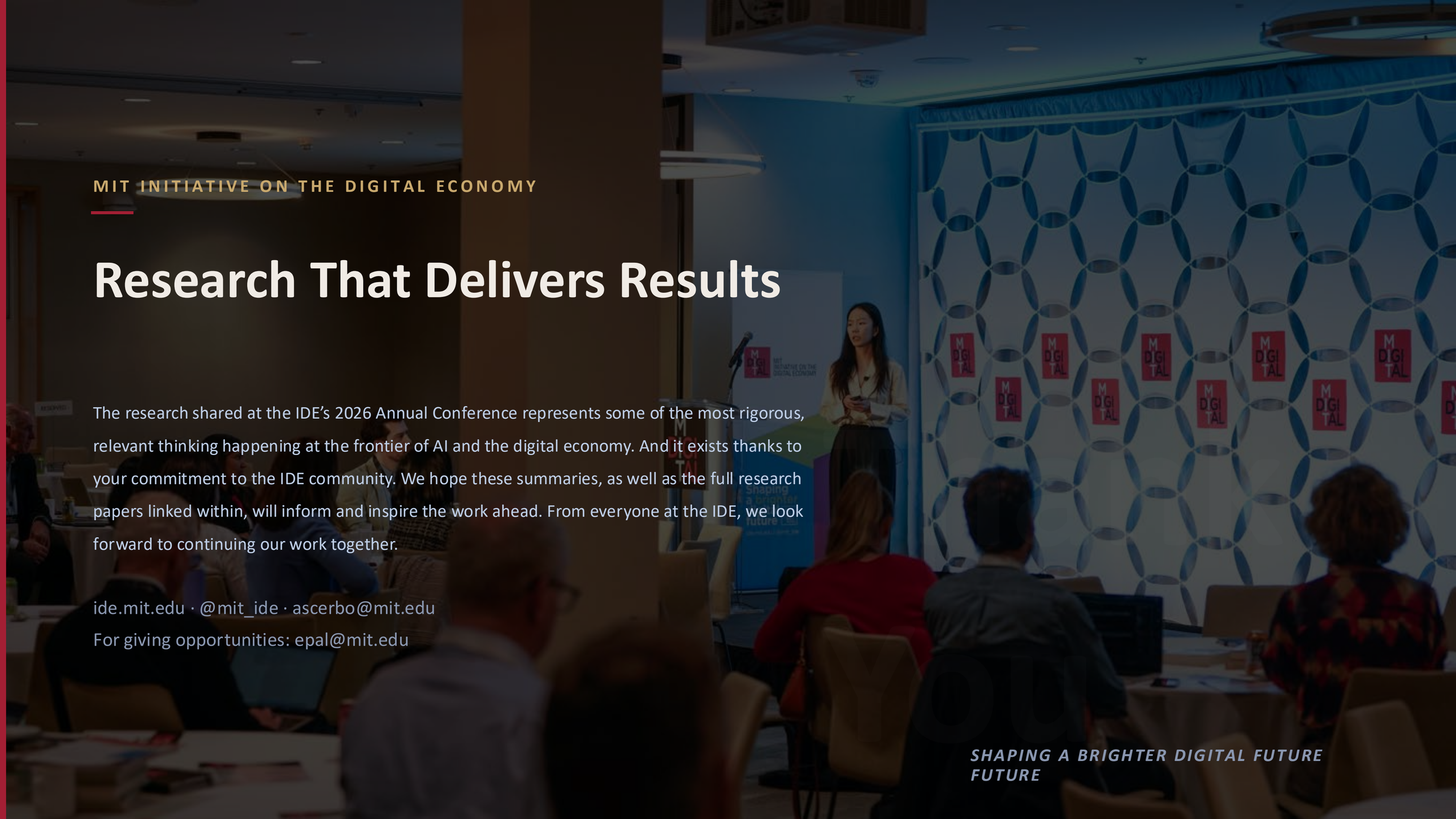
KEY TAKEAWAYS

- **Partial automation**—where AI handles some, but not all tasks in a role—is the role—is the dominant reality for most workers, rather than wholesale displacement.
- Organizations that redesign workflows around **human-AI collaboration** show higher innovation rates than those treating AI as a plug-in.
- **Co-designing AI processes with workers** outperforms top-down automation mandates on both adoption speed and long-term performance.
- Scale, complexity, and standardization determine automation potential—not the technology alone. These three variables offer a **practical lens for assessing where AI investment makes sense**.

EXPLORE THE RESEARCH

Economics of Human and AI Collaboration: When is Partial Automation More Attractive than Full Automation?

Thank You to Our Member Community



MIT INITIATIVE ON THE DIGITAL ECONOMY

Research That Delivers Results

The research shared at the IDE's 2026 Annual Conference represents some of the most rigorous, relevant thinking happening at the frontier of AI and the digital economy. And it exists thanks to your commitment to the IDE community. We hope these summaries, as well as the full research papers linked within, will inform and inspire the work ahead. From everyone at the IDE, we look forward to continuing our work together.

ide.mit.edu · @mit_ide · ascerbo@mit.edu
For giving opportunities: epal@mit.edu

SHAPING A BRIGHTER DIGITAL FUTURE
FUTURE