Returns to Intelligence

Suproteem K. Sarkar and Luke Melas-Kyriazi

April 14, 2026 (Preprint)

Please click here for the latest version

When AI improves, how does usage respond? We document a Jevons-like effect in AI demand: Improved capabilities increase usage, and shift AI use toward more complex tasks. We study firms that use the coding platform Cursor. After improvements in model capabilities in late 2025, workers sent 44% more agent messages per week on the platform. The initial increase was larger for lower-complexity messages, though usage shifted over time to higher-complexity messages. The increase was stronger in smaller, private, and newer firms, as well as in tasks with more cross-system dependencies like architecture and deployment work. The evidence is consistent with higher latent AI demand from firms that are organized more flexibly and can scale production more effectively. We discuss implications for investment by model providers, and consider how future usage growth relates to both competition between firms and downstream demand.

We thank Mohit Agarwal, Erik Brynjolfsson, Bharat Chandar, Jiafeng Chen, Andrey Fradkin, Adam Hofmann, Spencer Yongwook Kwon, Dennis Lee, Yueran Ma, Amil Merchant, Karl Oskar Schulz, Mirac Suzgun, Jordan Topoleski, and seminar participants at Stanford and the University of Chicago for helpful discussions. Sarkar: University of Chicago, Booth School of Business (suproteem@uchicago.edu); Melas-Kyriazi: Cursor (luke@cursor.com).

Global spending on AI infrastructure is expected to approach $1 trillion in 2028 (Patkar et al., 2025). Returns to these investments depend not only on technological progress, but also on how downstream usage responds to improved AI capabilities.

History indicates that technological progress can increase resource use. When steam engines became more coal-efficient, consumption of coal rose rather than fell—cheaper steam power enabled new applications like textile mills, railways, and steamships (Jevons, 1865; Sorrell, 2009). Similar effects have been observed for efficiency gains in lighting (Fouquet and Pearson, 2006). A key driver of these effects is latent downstream demand. In other markets with fewer potential new varieties, like vehicles and climate control, more efficient fuel or heating has not increased usage enough to offset the resource savings (Small and Dender, 2007; Aydin et al., 2017).

How AI progress affects AI usage depends on the adaptability of firms and the latent demand for new applications. Advances in AI have reduced the cost of completing some tasks with AI (Erol et al., 2026). If AI usage remains constant—firms do not modify their production processes or find new product applications—such efficiency gains may portend lower AI spending. However, if AI usage increases enough—firms use AI more for existing tasks or find applications for new tasks—AI spending may rise.

This paper studies how AI progress relates to AI usage, and how this relationship varies across firms and tasks. We analyze data on firms that use the coding platform Cursor. We find that capability improvements in late 2025 were followed by increased AI usage—workers sent 44% more messages per week on the platform. In addition to using more AI, workers shifted AI requests to target more complex tasks. The usage responses were concentrated in smaller, private, and newer firms, as well as in tasks with more cross-system dependencies. We discuss how these effects, coupled with the economics of adopting firms, may inform forecasts of AI usage.

Background The recent growth of AI capabilities, usage, and revenue has been driven in large part by coding agents. Improvements in coding agent capabilities have been made possible by abundant coding data on the internet and cheap-to-generate verifiable rewards for reinforcement learning. Coding agents have also become a focus area of AI labs because coding agents accelerate the development of AI itself, by speeding up the process of building training infrastructure and running experiments. Some labs believe that these advances could give rise to recursive self-improvement, where the rate of progress accelerates as agents facilitate the research process.

The model release cycles of the AI labs create natural discontinuities in the capabilities of models available to the public. In particular, the recent releases of Claude Opus 4.5 on November 24, 2025 and GPT-5.2 on December 11, 2025 were widely acknowledged as step-function increases in model intelligence and usefulness, particularly for coding. Andrej Karpathy, a founding member of OpenAI and former Director of AI at Tesla, described the impact of these releases: "I don't think I've typed like a line of code probably since December, basically, which is an extremely large change" (Karpathy, 2026). This sentiment, now common among software engineers, suggests that this moment could be ideal to study how AI usage patterns change in response to improved model capabilities.

We study usage patterns on Cursor, a platform for agent-assisted software development. Because Cursor is used by firms of varying sizes and sectors, it provides an ideal setting for observing how different firms respond to changes in AI capabilities. We use data on agent interactions, user query classifications, and firm characteristics. The breadth and granularity of these data allow us not only to measure the volume of agent usage before and after major model upgrades, but also to analyze how the characteristics of agent interactions evolve.

Increased usage and shift to more complex tasks Following major capability increases in late 2025—including the introduction of Claude Opus 4.5 and GPT-5.2—we document a sharp increase in AI usage. The average weekly messages per user rose by 44%. This spike began in the month following the new model releases, and persisted over the subsequent months. Developers initially used agents to complete more tasks of similar scope as before, but over time developers shifted toward using agents for more complex tasks.

Organizational flexibility predicts larger responses We analyze how the increase in AI usage varies across firm characteristics. Our analysis finds that smaller, private, and newer firms experienced larger increases in agent usage. For example, weekly agent messages per user rose more in the smallest employee tercile than in the largest (52% vs 38%), more in private firms than in public ones (46% vs 40%), and more in the youngest firm age tercile than in the oldest (47% vs 37%).

We also analyze variation across task categories. Among the larger increases in agent usage are tasks like architecture design and DevOps, which require reasoning and integration across systems. There was also larger growth in messages related to learning and understanding code.

Downstream demand and the distribution of surplus In addition to production-side factors like firm size and ownership, responses in AI usage should also depend on market- and demand-side factors.

Market competition may drive up AI usage by creating a technological arms race. For example, if one hedge fund uses more AI to develop trading strategies, rival funds have a stronger incentive to use AI tools. Consistent with this view, we find that financial firms were in the top half of sectors by increases in usage. If the size of an adopting market remains fixed, greater demand for AI may increase the margins of model providers without proportionally increasing the margins of AI-using firms.

Demand for new varieties may drive up AI usage by creating new applications. Consistent with this view, we find that firms in the media and advertising sector, which create content, had the highest increase in usage. If more capable models eventually enable more innovation in biology, materials science, and engineering, increased AI capabilities may increase product varieties that use these innovations. All in all, if the size of the total market increases, margins may rise for both model providers and AI-using firms.

Limitations Our evidence is observational, and is not the product of a randomized experiment. However, the late-2025 model releases have been widely viewed as major improvements in coding capabilities (Karpathy, 2026; Alphabet Inc., 2026). By observing field data on agent use, we observe a more natural distribution of responses to improved model capabilities than would be possible in a lab experiment. We consider a sample of workers who were already experienced agent users, in a period with no major platform scaffold improvements, which suggests that these estimates can be interpreted as responses to improved model capabilities.

Our main outcome is usage intensity, measured by messages per user, rather than spending or total compute. Although messages per user do not map one-for-one into inference spending, public reports of AI provider revenues indicate that spending also increased over this period (Anthropic, 2026). In addition, the share of users in the sample that were active on the platform did not fall during this period, so the results are unlikely to reflect attrition of less active users.

Finally, our setting is professional software development on Cursor during an early phase of AI diffusion. This makes our setting informative about production workflows that may be constrained by model capabilities. However, although the sectoral patterns we document are suggestive of broader market forces like competition and latent demand, they are not dispositive evidence about such effects.

Related literature This paper contributes to the growing literature on demand for frontier models and responses to capability improvements. Demirer et al. (2025) document patterns in pricing, supply, and demand across models and providers. Andrews and Farboodi (2025) find that markets respond to major model releases. Amodei (2024) discusses the "marginal returns to intelligence," and the conditions through which capability gains translate into economic gains. There is a growing literature on static evaluations across model capabilities (Hendrycks et al., 2021; Srivastava et al., 2022; Liang et al., 2023; Kwa et al., 2025; Patwardhan et al., 2025). Our paper uses message-level data from professional software developers to study how capability improvements change the level of AI usage and the content and complexity of requests.

Second, this paper relates to the literature on how improvements in AI systems may change production tasks. Jahani et al. (2024) show that as model capabilities improve, users learn to issue more ambitious and better-targeted prompts. Anthropic (2025) studies the tasks developers perform with AI, and Sarkar (2026) finds that agents shift worker effort to abstract, higher-order tasks. Hampole et al. (2025) find changes in task composition related to AI exposure. The broader task-based literature argues that new technologies can reshape the composition of work (Autor et al., 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2019). A growing body of literature studies AI in software development (Cui et al., 2024; Peng et al., 2023; Hoffmann et al., 2024; He et al., 2025; Chen and Stratton, 2026; Albarran, 2025; Chen et al., 2025; Becker et al., 2025; Shen and Tamkin, 2026). Our results also complement the argument from Cunningham (2026) that stronger language models are valuable in knowledge-creating settings, where the objective is not only to retrieve known information, but also to generate new ideas.

Third, our paper relates to the literature on general purpose technologies, organizational complements, and heterogeneous AI adoption and usage across firms (David, 1990; Bresnahan and Trajtenberg, 1995; Brynjolfsson and Hitt, 2000; Garicano, 2000; Bresnahan et al., 2002; Bloom et al., 2012, 2014; Brynjolfsson et al., 2019, 2021; Crouzet and Eberly, 2021; Crouzet et al., 2022; Brand et al., 2024; Agrawal et al., 2024; Bonney et al., 2024; Bick et al., 2024). With respect to organizational complements, Crouzet et al. (2025) argue that younger firms have organizational styles that enable innovation, Eisfeldt et al. (2023) find higher AI-related returns for firms with more data assets, and Acharya et al. (2026) find larger effects of AI on innovation for more focused firms.

On the extensive margin, McElheran et al. (2024) find more frequent AI adoption in younger and larger firms, Yotzov et al. (2026) in younger firms, and Baslandze et al. (2026) in larger firms. Babina et al. (2024) find that AI-related growth is stronger for larger firms. Technology investments can become more effective with scale (Hsieh and Rossi-Hansberg, 2023; Lashkari et al., 2024, 2026). Our paper studies responses on the intensive margin—conditional on using AI for coding, we find smaller and younger firms had greater usage increases after model capabilities improved. We do not measure how efficiently usage converts into revenue across adopting firms, so we do not produce directly comparable comparative statics to Babina et al. (2024) with respect to firm size.

Fourth, this paper relates to the literature on rebound effects and latent demand. Efficiency improvements do not necessarily reduce resource use; they can instead increase total usage by expanding the range of economically viable applications (Jevons, 1865; Sorrell, 2009; Fouquet and Pearson, 2006; Small and Dender, 2007; Aydin et al., 2017). Babina et al. (2024) find that AI investment is associated with more product innovation. Our evidence is consistent with a Jevons-like effect: More capable models may raise, rather than reduce, total inference demand by making new applications worthwhile. While our evidence is informative of inference demand, it does not speak to labor demand. A growing literature studies competition and market structure in AI, and their implications for welfare (Korinek and Vipra, 2024; Azoulay et al., 2024; Athey and Scott Morton, 2025).

Relative to these literatures, our paper makes three contributions. First, we provide direct evidence on how firms respond to improvements in frontier model capabilities. Second, we show that these capability shocks change not only the level of AI usage, but also the composition of messages, which shift toward more complex tasks. Third, we document heterogeneity across firms, tasks, and markets, and discuss how these results may relate to organizational complements, production processes, market structures, and downstream demand.

1. Framework and Data

We present a brief organizing framework and detail the data sources used for analysis.

1.1. Organizing Framework

We consider how firms decide to use AI. In our stylized framework, capability improvements and cost reductions increase usage, especially for firms that can effectively reorganize production processes. Easier tasks become capability-saturated earlier than harder tasks, and capability improvements shift the distribution of exposed tasks toward greater complexity. We include reduced-form discussions of how downstream demand drives the value of additional AI use, and how the distribution of surplus depends on market structure.

Setup A firm decides how intensively to use AI across task categories j. Let θ denote model capabilities, which raise the probability $\pi_j(\theta)$ that AI successfully completes a task in category j.[1] The value of a successfully completed task is v_j. Let c_j denote the marginal cost of AI usage, and let ϕ denote organizational flexibility, which lowers the cost of adopting workflows around higher AI use.[2]

For each task category j, the firm chooses usage intensity x_j to maximize expected payoff

$$\max_{x_j \geq 0} x_j[\pi_j(\theta)v_j - c_j] - \frac{x_j^2}{2\phi}$$

which is increasing in the expected value of the task $\pi_j(\theta)v_j$ and decreasing in marginal cost c_j and workflow adjustment cost $x_j^2/2\phi$. The first order condition yields usage

$$x_j^*(\theta, \phi) = \phi[\pi_j(\theta)v_j - c_j]_+$$

Usage AI becomes more cost-effective when usage costs decrease and success probabilities increase. Both these forces increase usage, i.e., $\frac{\partial x^*}{\partial c_j} \leq 0$, $\frac{\partial x^*}{\partial \theta} \geq 0$. For example, suppose that using legacy models to generate dashboards was costly, as it required multiple retries and manual corrections. If capability improvements increased how reliably an agent can produce a dashboard from one request, this task may now become more worthwhile. This could lead a developer tools-producing firm to generate new dashboards for its customers.

[1] Monotonicity of task success probabilities in θ is a simplifying assumption. Prior work has shown that model task performance differs in ways people do not expect (Dell'Acqua et al., 2026; Vafa et al., 2024; Morris et al., 2026), and more directly that model updates can change performance across tasks in different directions (Chen et al., 2023; Kirk et al., 2024).

[2] Constant c_j is a simplifying assumption. Pricing of new models may relate to capabilities, improved capabilities may change token consumption per request (Chen et al., 2026), and pricing menus may be more complex than modeled.

Complexity Easier tasks may reach capability saturation earlier than harder tasks. For example, suppose success probabilities follow sigmoids $p_e(\theta) = \sigma(\theta - \tau_e)$ for easier tasks and $p_h(\theta) = \sigma(\theta - \tau_h)$ for harder tasks, with $\tau_h > \tau_e$. Then for $\theta > (\tau_e + \tau_h)/2$, $p_h'(\theta) > p_e'(\theta)$, so marginal capability gains are larger for harder tasks. As model capabilities improve, AI can become marginally more worthwhile to use for complex tasks. This effect is stronger if the value of successful task completion also increases with complexity.

Organizational flexibility More capable models promote more usage, especially when firms can reorganize production processes, i.e., $\frac{\partial^2 x^*}{\partial\theta\, \partial\phi} \geq 0$. If smaller, private, and newer firms reorganize more effectively, their usage may respond more to capability improvements.

Downstream demand Usage increases more when firms have more potential downstream applications that raise the value of use, i.e., when v_j is large enough for more tasks. For firms that can adjust output to meet downstream demand, for example by developing new products or varieties, usage increases may be higher.

Distribution of surplus Gains from improved model capabilities may accrue across model providers, adopting firms, and consumers. In markets with smaller total growth opportunities and higher competition, the gains from usage may be competed away between adopting firms, and accrue to model providers and consumers. In markets with larger total growth opportunities, adopting firms may capture a larger share of surplus.

1.2. Data

We study firms that use the coding platform Cursor. We use a random sample of 500 firms, uniformly drawn from a subset of firms that joined the platform before July 1, 2025. We also restrict the sample to users who joined before July 1, 2025. We match these firms to sector, total employees, founding year, and publicly traded status as of the end of 2025 using an internal Cursor database. 149/500 firms are publicly traded, and founding year is matched for 478/500 firms. The mean total employee count is 9,094, and the median is 1,559.

We link these firms to usage data over time. We analyze data on the number of messages sent per week on the platform between the weeks of July 7, 2025 and March 16, 2026. We also analyze data on message complexity and task category through a platform classifier; this data is available starting the week of December 15, 2025. This classifier runs transiently when the agent is in use, and does not store message content. The complexity classification considers the work an engineer would have to do to complete a task. These labels correspond to tasks that are likely Trivial (line-level context), Low (file-level context), Medium (multi-file context), and High (cross-codebase context) in a typical codebase.

We use these data to build firm-by-week panels of coding agent usage. For the base panel, the observations in each firm-week row correspond to the average number of messages across users active in that firm and week. For analysis that uses complexity or task classifications, we additionally use the average number of messages for each classification across users active in that firm and week.

2. Capability Improvements

After improvements to model capabilities in late 2025, AI usage increased and shifted toward more complex tasks. The largest percent increase was in messages classified as high complexity.

2.1. Increased Usage

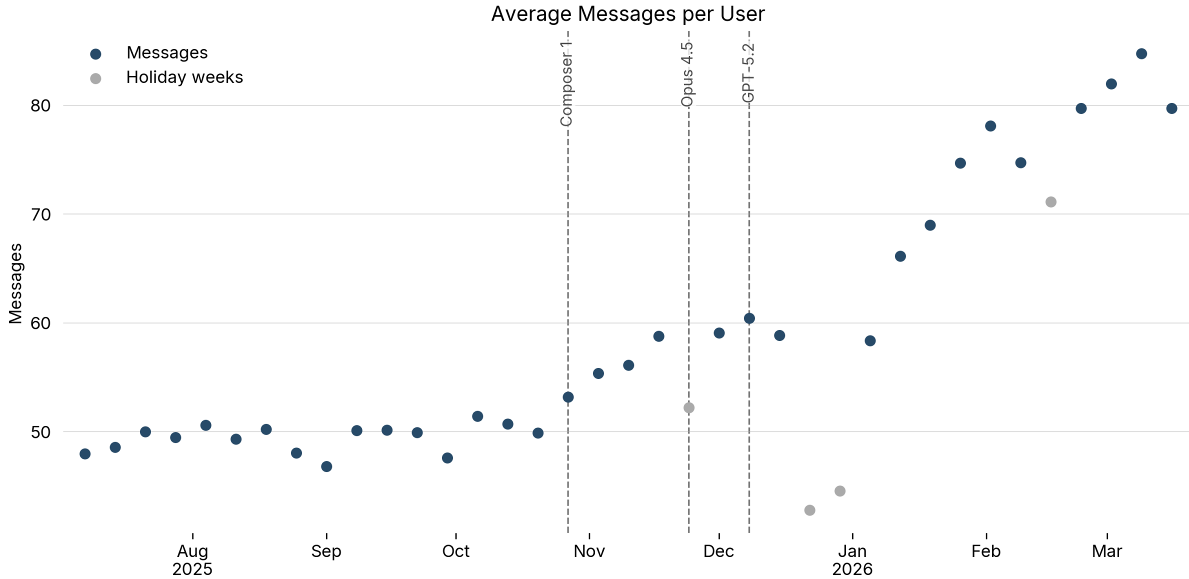
Figure 1 plots the average number of messages per user over time. We include vertical lines for the releases of Opus 4.5 and GPT-5.2, which have been acknowledged as step changes in model capabilities. We also include a vertical line for the release of Composer 1, which was a model that reduced the cost for a typical query. These releases may have increased viable AI applications by enabling new use cases and reducing the costs of existing use cases.

The average messages per week rose from 52 in 2025 to 75 in 2026, excluding holiday weeks. On average, this reflected a 44% increase in messages per user on the platform. This growth in usage suggests that models may have reached a threshold of capability that enabled more frequent use and new use cases.

2.2. Shift to More Complex Tasks

Figure 2 plots the average number of messages sent over time, split by complexity classification. Initially, growth was larger in low- and medium-complexity messages. After 4–6 weeks, growth was larger in high-complexity messages. Over the available

Figure 1: Usage increased after late-2025 model releases.



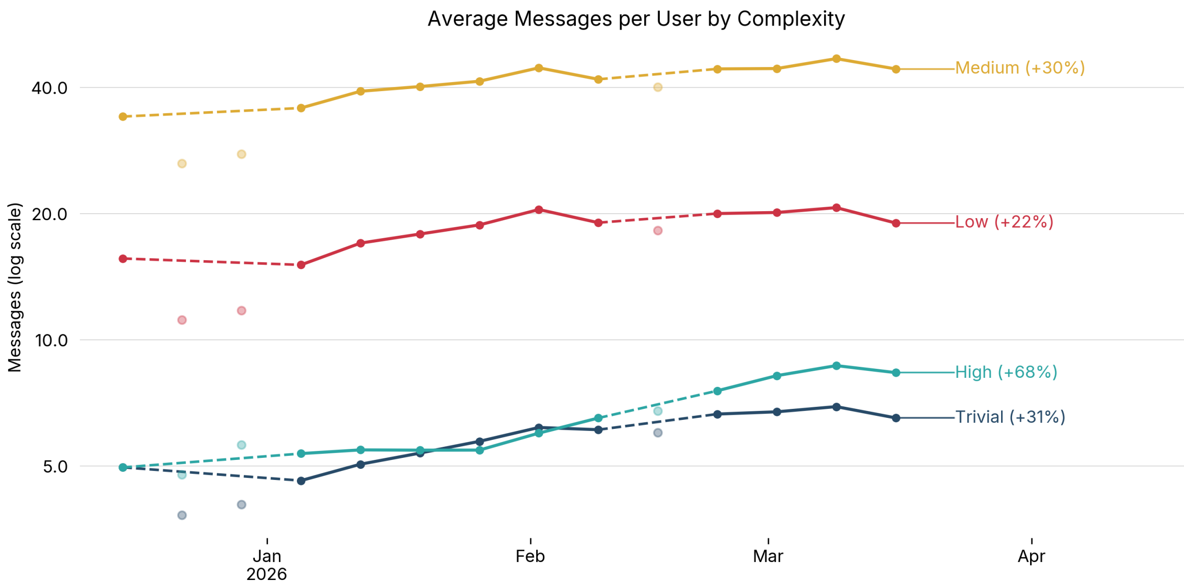
Notes – This figure plots weekly messages per user, averaged across firms in the sample. Vertical lines correspond to three model releases in late 2025. Users send fewer messages in holiday weeks, and data from these weeks are indicated in the figure in gray.

sample period, the largest increase was in high-complexity messages, at 68%. By contrast, low-complexity messages increased by 22% and medium-complexity messages increased by 30% over the same time period.

Improved model capabilities allow workers to perform new and challenging tasks, but the shift in the distribution of task difficulty lags the general increase in usage. This lag may be partially explained by the time it takes users to naturally discover new capabilities over the course of their work, and it may also be explained by firms having to reorient workflows around new model capabilities. For example, in order for developers to fully delegate tasks that last hours or days to their agents, they may have to switch from running agents locally to running them in cloud-hosted environments, which requires changing their development workflows.

Another finding is that the average number of weekly messages classified as trivial increased by 31%. This increase may reflect a qualitative change in how people use their coding tools: Some developers' primary mode of implementation may have switched from manual editing to conversing with an agent. The first reaction of these developers to seeing a small issue in their code may now be to ask the agent to fix it rather than fixing it manually, even when the fix is trivial.

Figure 2: Usage shifted to more complex tasks.



Notes – This figure plots weekly messages per user, split by complexity classifications, averaged across firms. The figure is restricted to weeks in the conversation classification subsample, which contains data on message classifications. The vertical axis has a log scale so relative changes are easier to compare. Users typically send fewer messages in holiday weeks, and data from these weeks are shown in muted colors and overlaid with dashed lines.

3. Heterogeneity Across Firms, Tasks, and Markets

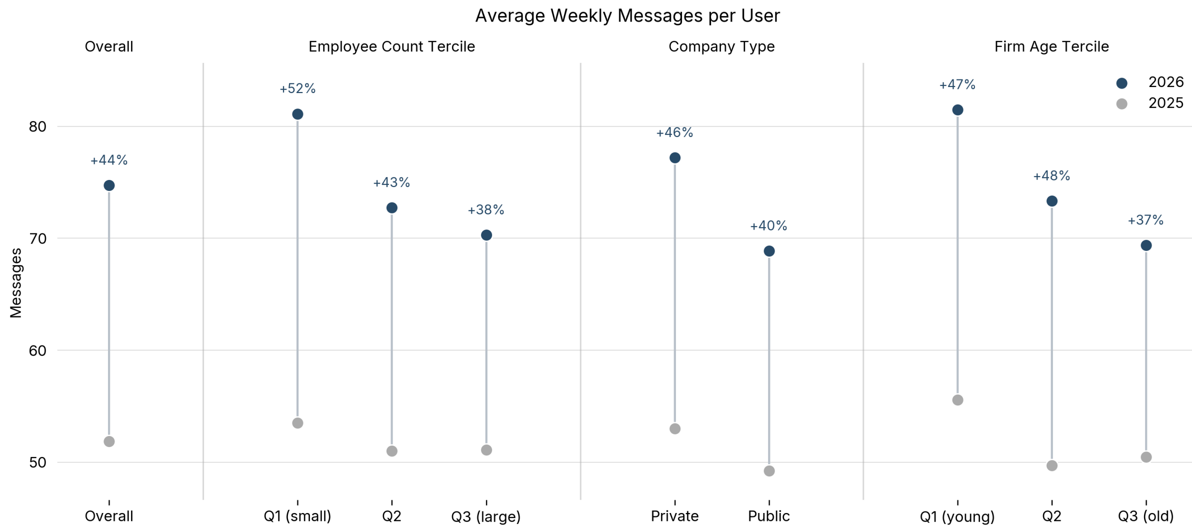
We find that usage increases are stronger in smaller, private, and newer firms. They are also stronger in tasks with more cross-system dependencies. The largest increases were in media, software, and financial firms.

3.1. Firm Characteristics

Figure 3 plots changes in agent usage across firms of varying size, ownership, and age. We find that responses to improved model capabilities are heterogeneous across firms, with disproportionate increases in usage across smaller, private, and newer firms. Table A1 reports regression results that interact a post-period indicator with firm characteristics, and finds similar results.

First, we compare firms by size, dividing the sample into terciles based on the number of total employees. The median number of employees was 582 for the smallest tercile, 1,559 for the middle tercile, and 9,712 for the largest tercile. Smaller and mid-sized

Figure 3: Stronger increases for smaller, private, and newer firms.



Notes – This figure plots the mean messages per user in 2026 versus 2025, as well as percent increases, averaged across firms in the sample. The left-most panel corresponds to the overall increase. The second, third, and fourth panels split the sample by firm characteristics. We exclude data from the four holiday weeks.

firms saw larger relative increases in usage. Average weekly agent messages per user rose by 52% in the smallest tercile, 43% in the middle tercile, and 38% in the largest tercile.

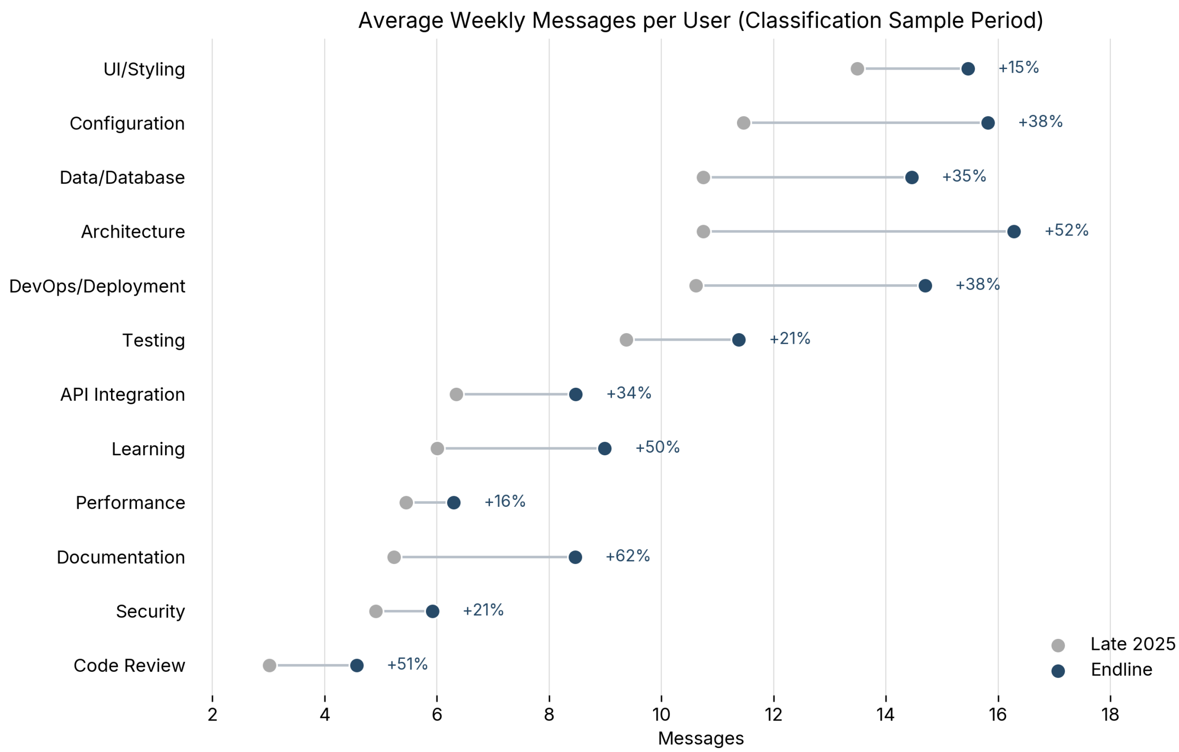
Second, we consider ownership structure. Private firms experienced a larger increase (46% vs 40%) in agent usage than public companies despite already having higher usage before the capability improvements. This difference could reflect faster decision cycles, higher willingness to experiment, and competitive pressures at private companies.

Third, we consider firm age. The median firm age in years was 11 for the youngest tercile, 15 for the middle tercile, and 28 for the oldest tercile. Average weekly agent messages per user rose by 47% in the youngest tercile, 48% in the middle tercile, and 37% in the oldest tercile. Younger firms experienced larger increases in usage than older firms, and the middle tercile experienced similar percent increases to the younger tercile despite a lower baseline level.

3.2. Task Characteristics

Figure 4 plots changes in agent usage across task categories. The classifier assigns up to two task categories per message, and the figure considers all messages that had a given

Figure 4: Stronger increases for tasks with more dependencies.

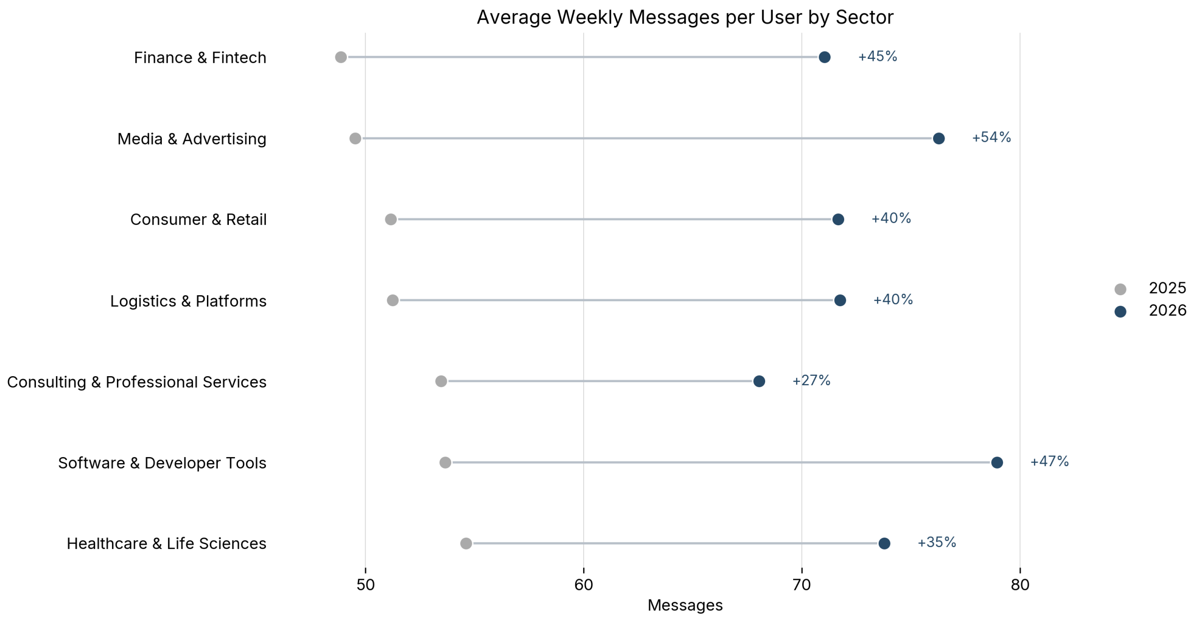


Notes – This figure plots the mean messages per user, as well as percent increases, in the first and final weeks of the conversation classification subsample. Each row splits the sample by task category classification.

classification, so the sum of messages across categories is larger than the average number of messages in a given week. Between the first and final weeks of the conversation classification sample, the largest increase was in documentation (62%). Other large increases were in architecture (52%), code review (51%), learning (50%), DevOps and deployment (38%), and data and database (35%). By contrast, more self-contained tasks such as UI/Styling (15%) saw smaller increases. The growth in tasks involving documentation and learning is notable; this may reflect the degree to which more capable coding agents help firms gather and integrate knowledge.

The usage growth in tasks with cross-system dependencies, like architecture and deployment, suggests that firms are using agents for harder-to-verify tasks. Improved capabilities may lead to higher trust in model outputs, which may make workers more likely to use models for these tasks. Improved capabilities may also allow workers to write more ambitious tests than before, which may shift verifiability thresholds. This new

Figure 5: Stronger increases in media, software, and finance.



Notes – This figure plots the mean messages per user in 2026 versus 2025, as well as percent increases, averaged across firms in the sample. We report changes for each sector. We exclude data from the four holiday weeks.

task distribution could also create more technical debt if firms do not spend as much effort verifying model-generated outputs.

3.3. Market Characteristics

Figure 5 plots changes in agent usage across sectors. Usage increased in every sector, but the largest relative gains were in Media & Advertising (54%), Software & Developer Tools (47%), and Finance & Fintech (45%). There were more moderate increases in Consumer & Retail (40%), Logistics & Platforms (40%), Healthcare & Life Sciences (35%), and Consulting & Professional Services (27%). Usage increases do not relate one-to-one with initial usage: Consulting and healthcare already had relatively high baseline message rates in 2025, but had smaller increases than media, software, and finance.

One interpretation is that these sectoral differences reflect market structure. In finance, for example, capability gains may create an arms-race dynamic: If some firms can use AI to form trading strategies, rival firms may have stronger incentives to adopt similar tools. In software and media, by contrast, more capable agents may raise usage by expanding the set of viable products and content varieties that firms can produce. In markets where final output is more bounded, more capable models may eventually lower the cost of existing workflows more than they induce increases in total usage. This evidence is suggestive—we cannot rule out other explanations, like differences in technical skill or process adaptability across sectors.

4. Discussion

Our paper finds that improved AI capabilities are associated with increased AI usage. After capability improvements in late 2025, users of the coding platform Cursor sent more messages and used AI for more complex tasks. These effects were concentrated in smaller, private, and newer firms, which suggests the importance of organizational structure. They were also concentrated in tasks with cross-system dependencies, which suggests that more complex tasks may continue to drive increased AI usage as model capabilities improve. Sector-level evidence suggests that AI usage increases may be sensitive to competition and the potential for more final output.

Investment by model providers The private return to investing in additional model capabilities may vary across use cases. In consumer-facing chat applications, users may already be at a point where additional model capabilities have small effects on total usage. By contrast, production applications appear more capability-bound. Tasks such as architecture design, debugging, and database work benefit from models that are reliable over more complex design requirements. In these settings, capability improvements may expand the set of tasks for which AI is a viable input.

This distinction also suggests that model capabilities and scaffolding are complements. Better base models increase the value of tools, such as retrieval and memory, that help firms apply intelligence to their workflows. Conversely, improvements in scaffolding may be important in use cases where firms are more constrained by frictions in applying model outputs to production. The optimal investment frontier is therefore unlikely to be one-dimensional. Some downstream usage is enabled by more intelligent models, while some is enabled by integrating existing intelligence into better systems.

Competition and the distribution of surplus The distribution of surplus between model providers and model users is likely to depend on downstream market structure. In more competitive markets, firms may increase AI usage, but the gains may be competed away through lower prices and higher consumer expectations. In that case, some of the surplus from better models may accrue to model providers and end users rather than to adopting firms. By contrast, when more capable AI enables new products or entry into previously infeasible activities, adopting firms may accrue more of the surplus.

Taken together, our findings suggest that the economics of AI should be framed not only around substitution, but also around expansion. More capable models can increase usage by making AI applicable to new tasks. We present early evidence on how the magnitude of such increases relates to economic forces.

References

Acemoglu, Daron and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171. 4

_ **and Pascual Restrepo**, "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, 2019, *33* (2), 3–30. 4

Acharya, Viral V., Xuehua Shao, and Zhaoxia Xu, "AI and Productivity: The Role of Innovation," March 2026. Working Paper. 4

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb, "Artificial intelligence adoption and system-wide change," *Journal of Economics & Management Strategy*, 2024, *33* (2), 327–337. 4

Albarran, Nik, "What New Data Tells Us About the Rise of Agentic AI in Engineering Workflows," Jellyfish Blog August 2025. Accessed: 2026-02-16. 4

Alphabet Inc., "2025 Q4 Earnings Call," February 2026. Alphabet Investor Relations. 3

Amodei, Dario, "Machines of Loving Grace: How AI Could Transform the World for the Better," October 2024. Essay. 4

Andrews, Isaiah and Maryam Farboodi, "Do Markets Believe in Transformative AI?," NBER Working Paper 34243, National Bureau of Economic Research September 2025. 4

Anthropic, "Anthropic Economic Index: AI's Impact on Software Development," April 2025. 4

_ , "Anthropic raises $30 billion in Series G funding at $380 billion post-money valuation," February 2026. 3

Athey, Susan and Fiona Scott Morton, "Artificial Intelligence, Competition, and Welfare," NBER Working Paper 34444, National Bureau of Economic Research November 2025. 5

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333. 4

Aydin, Erdal, Nils Kok, and Dirk Brounen, “Energy Efficiency and Household Behavior: The Rebound Effect in the Residential Sector,” *RAND Journal of Economics*, 2017, *48* (3), 749–782. 1, 5

Azoulay, Pierre, Joshua L. Krieger, and Abhishek Nagaraj, “Old Moats for New Models: Openness, Control, and Competition in Generative AI,” NBER Working Paper 32474, National Bureau of Economic Research May 2024. Revised July 2024. 5

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson, “Artificial Intelligence, Firm Growth, and Product Innovation,” *Journal of Financial Economics*, 2024, *151*, 103745. 5

Baslandze, Salomé, Zachary Edwards, John Graham, Ty McClure, Brent H. Meyer, Michael Sparks, Sonya R. Waddell, and Daniel Weitz, “Artificial Intelligence, Productivity, and the Workforce: Evidence from Corporate Executives,” Working Paper 34984, National Bureau of Economic Research 2026. 5

Becker, Joel, Nate Rush, Elizabeth Barnes, and David Rein, “Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity,” July 2025. 4

Bick, Alexander, Adam Blandin, and David J. Deming, “The Rapid Adoption of Generative AI,” NBER Working Paper 32966, National Bureau of Economic Research September 2024. 4

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen, “The Distinct Effects of Information Technology and Communication Technology on Firm Organization,” *Management Science*, 2014, *60* (12), 2859–2885. 4

_ , Raffaella Sadun, and John Van Reenen, “Americans Do IT Better: US Multinationals and the Productivity Miracle,” *American Economic Review*, 2012, *102* (1), 167–201. 4

Bonney, Kathryn, Cory Breaux, Cathy Buffington, Emin Dinlersoz, Lucia S. Foster, Nathan Goldschlag, John C. Haltiwanger, Zachary Kroff, and Keith Savage, “Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey,” NBER Working Paper 32319, National Bureau of Economic Research April 2024. 4

Brand, James M., Mert Demirer, Connor Finucane, and Avner A. Kreps, “Firm Productivity and Learning with Digital Technologies: Evidence from Cloud Computing,” NBER Working Paper 32938, National Bureau of Economic Research September 2024. Revised December 2025. 4

Bresnahan, Timothy F. and Manuel Trajtenberg, "General Purpose Technologies 'Engines of Growth'?," *Journal of Econometrics*, 1995, *65* (1), 83–108. 4

__ **, Erik Brynjolfsson, and Lorin M. Hitt**, "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *The Quarterly Journal of Economics*, 2002, *117* (1), 339–376. 4

Brynjolfsson, Erik and Lorin M. Hitt, "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *Journal of Economic Perspectives*, 2000, *14* (4), 23–48. 4

__ **, Daniel Rock, and Chad Syverson**, "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, Chicago: University of Chicago Press, 2019, pp. 23–60. 4

__ **,** __ **, and** __ , "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, January 2021, *13* (1), 333–372. 4

Chen, Fiona and James Stratton, "Artificial Intelligence in the Firm," January 2026. 4

Chen, Lingjiao, Chi Zhang, Yeye He, Ion Stoica, Matei Zaharia, and James Zou, "The Price Reversal Phenomenon: When Cheaper Reasoning Models End Up Costing More," 2026. 6

__ **, Matei Zaharia, and James Zou**, "How is ChatGPT's Behavior Changing over Time?," 2023. 6

Chen, Valerie, Ameet Talwalkar, Robert Brennan, and Graham Neubig, "Code with Me or for Me? How Increasing AI Automation Transforms Developer Workflows," *arXiv preprint arXiv:2507.08149*, 2025. 4

Crouzet, Nicolas and Janice C. Eberly, "Rents and Intangible Capital: A Q+ Framework," Working Paper 28988, National Bureau of Economic Research 2021. 4

__ **,** __ **, Andrea L. Eisfeldt, and Dimitris Papanikolaou**, "The Economics of Intangible Capital," *Journal of Economic Perspectives*, 2022, *36* (3), 29–52. 4

—, **Zhiguo He, Victor Lyonnet, and Yueran Ma**, "Why Don't Old Firms Do New Things?," November 2025. Working Paper. 4

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz, "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers," *Available at SSRN 4945566*, 2024. 4

Cunningham, Tom, "Knowledge-Creating LLMs," January 2026. Blog post. 4

David, Paul A, "The dynamo and the computer: an historical perspective on the modern productivity paradox," *The American Economic Review*, 1990, *80* (2), 355–361. 4

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," *Organization Science*, 2026. 6

Demirer, Mert, Andrey Fradkin, Nadav Tadelis, and Sida Peng, "The Emerging Market for Intelligence: Pricing, Supply, and Demand for LLMs," NBER Working Paper 34608, National Bureau of Economic Research December 2025. 4

Eisfeldt, Andrea L., Gregor Schubert, and Miao Ben Zhang, "Generative AI and Firm Values," Working Paper 31222, National Bureau of Economic Research 2023. 4

Erol, Mehmet Hamza, Batu El, Mirac Suzgun, Mert Yuksekgonul, and James Zou, "Cost-of-Pass: An Economic Framework for Evaluating Language Models," 2026. 1

Fouquet, Roger and Peter J. G. Pearson, "Seven Centuries of Energy Services: The Price and Use of Light in the United Kingdom (1300–2000)," *The Energy Journal*, 2006, *27* (1), 139–178. 1, 5

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, *108* (5), 874–904. 4

Hampole, Menaka, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller, "Artificial Intelligence and the Labor Market," 2025. Working paper. 4

He, Hao, Courtney Miller, Shyam Agarwal, Christian Kästner, and Bogdan Vasilescu, "Speed at the Cost of Quality: How Cursor AI Increases Short-Term Velocity and

Long-Term Complexity in Open-Source Projects," *arXiv preprint arXiv:2511.04427*, November 2025. 4

Hendrycks, Dan, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt, "Measuring Massive Multitask Language Understanding," in "International Conference on Learning Representations" 2021. 4

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu, "Generative AI and the Nature of Work," 2024. 4

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The Industrial Revolution in Services," *Journal of Political Economy Macroeconomics*, 2023, *1* (1). 5

Jahani, Eaman, Benjamin S. Manning, Joe Zhang, Hong-Yi TuYe, Mohammed Alsobay, Christos Nicolaides, Siddharth Suri, and David Holtz, "Prompt Adaptation as a Dynamic Complement in Generative AI Systems," *arXiv preprint arXiv:2407.14333*, 2024. 4

Jevons, William Stanley, *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*, London: Macmillan and Co., 1865. 1, 5

Karpathy, Andrej, "Interview on the *No Priors* Podcast," March 2026. Reported by Jason Ma, *Fortune*, March 21, 2026. 2, 3

Kirk, Robert, Ishita Mediratta, Christoforos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and Roberta Raileanu, "Understanding the Effects of RLHF on LLM Generalisation and Diversity," in "Proceedings of the 12th International Conference on Learning Representations (ICLR)" 2024. 6

Korinek, Anton and Jai Vipra, "Concentrating Intelligence: Scaling and Market Structure in Artificial Intelligence," NBER Working Paper 33139, National Bureau of Economic Research November 2024. 5

Kwa, Thomas, Ben West, Joel Becker, Amy Deng, Katharyn Garcia, Max Hasin, Sami Jawhar, Megan Kinniment, Nate Rush, Sydney Von Arx, Ryan Bloom, Thomas Broadley, Haoxing Du, Brian Goodrich, Nikola Jurkovic, Luke Harold Miles, Seraphina Nix, Tao Lin, Neev Parikh, David Rein, Lucas Jun Koba Sato, Hjalmar Wijk, Daniel M. Ziegler, Elizabeth Barnes, and Lawrence Chan, "Measuring AI

Ability to Complete Long Software Tasks," *arXiv preprint arXiv:2503.14499*, 2025. NeurIPS 2025. 4

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard, "Information Technology and Returns to Scale," *American Economic Review*, 2024, *114* (6), 1769–1815. 5

_ , **Wensu Li, Christina Qiu, and Neil Thompson**, "AI and Scale: A Quantitative Task-Based Theory of Automation," March 2026. 5

Liang, Percy, Rishi Bommasani, Tony Lee et al., "Holistic Evaluation of Language Models," *Transactions on Machine Learning Research*, 2023. 4

McElheran, Kristina, J. Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI Adoption in America: Who, What, and Where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415. 5

Morris, Meredith Ringel, Dan Altman, Haydn Belfield, Arthur Goemans, Hasan Iqbal, Ryan Burnell, Iason Gabriel, Samuel Albanie, and Allan Dafoe, "Characterizing Model Jaggedness Supports Safety and Usability," 2026. 6

Patkar, Vishwas, Lindsay A. Tyler, Carolyn L. Campbell, Vishwanath Tirupattur, Stephen C. Byrd, James Egan, Kyle B. Yellin, and Catherine Liu, "Bridging a $1.5tr Data Center Financing Gap," North America Insight, Morgan Stanley Research July 2025. 1

Patwardhan, Tejal, Rachel Dias, Elizabeth Proehl, Grace Kim, Michele Wang, Olivia Watkins, Simón Posada Fishman, Marwan Aljubeh, Phoebe Thacker, Laurance Fauconnet, Natalie S. Kim, Patrick Chao, Samuel Miserendino, Gildas Chabot, David Li, Michael Sharman, Alexandra Barr, Amelia Glaese, and Jerry Tworek, "GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks," *arXiv preprint arXiv:2510.04374*, 2025. 4

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer, "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," February 2023. 4

Sarkar, Suproteem K., "AI Agents and Higher-Order Work," *SSRN Electronic Journal*, March 2026. 4

Shen, Judy Hanwen and Alex Tamkin, "How AI Impacts Skill Formation," *arXiv preprint arXiv:2601.20245*, January 2026. 4

Small, Kenneth A. and Kurt Van Dender, “Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect,” *The Energy Journal*, 2007, *28* (1), 25–52. 1, 5

Sorrell, Steve, “Jevons’ Paradox Revisited: The Evidence for Backfire from Improved Energy Efficiency,” *Energy Policy*, 2009, *37* (4), 1456–1469. 1, 5

Srivastava, Aarohi, Abhinav Rastogi, Abhishek Rao et al., “Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models,” *arXiv preprint arXiv:2206.04615*, 2022. 4

Vafa, Keyon, Ashesh Rambachan, and Sendhil Mullainathan, “Do Large Language Models Perform the Way People Expect? Measuring the Human Generalization Function,” in “Proceedings of the 41st International Conference on Machine Learning,” Vol. 235 of *Proceedings of Machine Learning Research* PMLR 2024, pp. 48919–48937. 6

Yotzov, Ivan, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Steven J. Davis, Kevin M. Foster, Aaron Jalca, Brent H. Meyer, Paul Mizen, Michael A. Navarrete, Pawel Smietanka, Gregory Thwaites, and Ben Zhe Wang, “Firm Data on AI,” Working Paper 34836, National Bureau of Economic Research February 2026. 5

A. Appendix

Results across firm characteristics Table A1 reports regression results following

$$\text{Log(Messages)}_{f,t} = \alpha_f + \delta_t + \beta \times 1(\text{Post})_t \times X_f + \varepsilon_{f,t} \quad (1)$$

across firms f and weeks t. $1(\text{Post})_t$ is an indicator for week t falling in 2026, and X_f is a firm-level characteristic. Employee count and years since founding are cross-sectionally percentile ranked. The four holiday weeks are excluded from the regression.

Table A1: Usage increases by firm characteristics.

	Log Messages			
	(1)	(2)	(3)	(4)
Post	0.34***			
	(0.01)			
Post × Employee Count (rank)		-0.14***		
		(0.05)		
Post × Private			0.05**	
			(0.02)	
Post × Years Since Founding (rank)				-0.09**
				(0.04)
Firm FE	✓	✓	✓	✓
Week FE		✓	✓	✓
Observations	16,500	16,500	16,500	15,774
Firms	500	500	500	478
R^2	0.626	0.677	0.676	0.672
Within R^2	0.318	0.005	0.002	0.003

Notes – This table reports regression results following Equation (1). Standard errors are clustered by firm.

Results across sectors Table A2 reports regression results following

$$\text{Log(Messages)}_{f,t} = \alpha_f + \delta_t + \beta \times 1(\text{Post})_t \times S_f + \varepsilon_{f,t} \tag{2}$$

across firms f and weeks t. $1(\text{Post})_t$ is an indicator for week t falling in 2026, and S_f is a sector indicator. The baseline sector is Consulting & Professional Services. The four holiday weeks are excluded from the regression.

Table A2: Usage increases across sectors.

	Log Messages	
	(1)	(2)
Post	0.23***	
	(0.03)	
Post × Healthcare & Life Sciences	0.07	0.07
	(0.04)	(0.04)
Post × Software & Developer Tools	0.13***	0.13***
	(0.04)	(0.04)
Post × Finance & Fintech	0.13***	0.13***
	(0.04)	(0.04)
Post × Consumer & Retail	0.10*	0.10*
	(0.05)	(0.05)
Post × Media & Advertising	0.16***	0.16***
	(0.05)	(0.05)
Post × Logistics & Platforms	0.10**	0.10**
	(0.04)	(0.04)
Firm FE	✓	✓
Week FE		✓
Observations	16,500	16,500
Firms	500	500
R^2	0.628	0.677
Within R^2	0.322	0.007

Notes – This table reports regression results following Equation (2). The baseline sector is Consulting & Professional Services. Standard errors are clustered by firm.