

AI at Work: Workforce Case Studies and Experiments from Google Workspace

Scott Strand, Nacho Cigliutti, Zanna Iscenko, and Fabien Curto Millet

Google AI & Economy Program

Abstract

This white paper presents empirical findings from a series of studies conducted by the Google Workspace team from 2024-2025, evaluating the real-world impact of generative AI on knowledge-worker productivity.

Drawing from over 23,000 participants, we synthesize findings from three distinct studies: an in-person qualitative study on user workflows ("Exploratory study on Archetype Journeys in Workspace"); a randomized experiment measuring task success and time savings ("AI Advantage for Small Businesses Study"); and a randomized trial evaluating an AI form-creation feature ("Help Me Create" Study). This paper builds upon the growing body of research on the subject of AI and labor productivity by experimentally evaluating the impact of generative AI tools embedded within widely-used office software. Complementing prior research focusing on single occupations, single firms, specific tasks, or relying on non-experimental data, we analyze the effects of AI on workplace productivity across a diverse population of workers spanning a range of occupations.

The results show that AI assistance consistently improves task success rates and reduces time on task, especially in text-intensive tasks such as drafting emails, summarizing documents, and creating forms. For instance, in our large-scale trial, users with access to the AI form-creation tool spent 27% less time editing their forms, largely driven by a 68% reduction in time spent adding questions. Similarly, our lab experiment found that for text-heavy tasks, AI assistance increased task success rates by up to 25% and reduced completion time by up to 32%. Notably, the small business study revealed a significant "leveling" effect, where AI disproportionately boosted the performance of less proficient users, narrowing the skill gap. Across studies, AI tools consistently boosted user satisfaction and confidence. As Gemini in Workspace undergoes a major evolution in 2026, bringing significant new functionality and more powerful integrations, this research serves as a starting point for understanding the impact of an even more expansive set of AI capabilities in collaboration and productivity apps.

1 Introduction

The rise of generative AI, particularly large language models (LLMs), is transforming how knowledge work is done (Hartley et al. (2024), Bick, Blandin, and Deming (2025), Bonney

et al. (2024), DORA (2025)). AI-powered tools are now embedded into the daily workflows of millions of professionals, helping them to write code, draft emails, synthesize complex documents, and much more. Understanding this impact is critical, not just for advancing product design, but also for informing discussions about the future of work.

This paper provides a synthesis of Google Workspace’s recent research exploring how embedded AI tools influence knowledge worker productivity. We combine findings from three experiments conducted by the Google Workspace team between 2024 and 2025. These include controlled usability testing, a randomized experiment with hundreds of small business users, and a large-scale A/B experiment of “Help Me Create” in Google Forms.

A growing body of research has explored this topic, generally finding that AI enhances productivity, though the specifics of these findings depend on the methodology. Studies have ranged from high-level macroeconomic perspectives to analyses of single companies or workers in specialized domains, like software development. To understand the effects of AI on the future of work more broadly, our research looks beyond single domains to measure the impact on tasks that stretch across knowledge worker professions, such as writing emails, preparing presentations, or creating surveys, where AI is used with widely-used office productivity software.

While some studies have examined knowledge workers outside of software engineering, they have often been limited to users at a single firm, focused on one type of task (e.g. management consulting, customer service, financial analysis), or measured performance on abstract tasks rather than those actually performed on the job.

Our goal with this paper is to make a contribution toward filling this gap, providing further evidence of AI’s benefits for real-world workers across a range of tasks across occupational categories. By combining qualitative usability analysis, randomized experiments, and a large-scale field trial in production software, these studies examine a range of realistic, on-the-job tasks performed by knowledge workers across a variety of professions. We aim not only to quantify where AI has a causal effect on productivity growth (and where there are opportunities to improve), but also to explore how design choices can amplify its benefits. In doing so, we aim to inform customers, companies, researchers, and developers hoping to better understand AI at work and how users can benefit.

2 Gemini-Based AI Tools in Google Workspace

Knowledge workers operate within dense digital ecosystems. They manage projects through email, develop plans in documents, and analyze performance in spreadsheets - all while using a fragmented landscape of productivity applications. As a result, productivity can often suffer from context-switching, manual repetition of low-value tasks, and the cognitive load of starting from blank pages. Gemini in Workspace is designed to address these frictions by serving as an embedded, context-aware AI assistant that augments user capabilities across applications, streamlines workflows, and accelerates the path from idea to output.

Gemini-based tools are natively integrated across Google Workspace applications such as Gmail, Docs, Sheets, Slides, Meet, Drive, Chat, Vids, and Forms, enabling users to prompt for assistance with a range of work tasks directly within the product’s interface. These tasks include writing, creating images, organizing data, connecting more meaningfully with

others, and more. Users benefit from real-time support tailored to the context of their active app: e.g., generation or refinement within Docs, conditional formatting or insights in Sheets, automated summarization in Chat or Meet, and image or layout generation in Slides.

“Help Me Create” is a specialized implementation of one of these Gemini use cases. Labeled as the “Help me create a form” AI feature in Google Forms, it enables users to generate an entire draft form from a natural-language prompt. When users (with this feature activated) create a new form, they see a large card that allows them to type in a prompt. They describe the form, tap a create button, and are presented with an LLM-structured form populated with AI-generated survey questions, which the user can then refine. Users in this study either had access to this feature through a paying consumer account (Google AI Pro and Ultra) or business account (Google Business Standard or Plus, or Google Enterprise Standard or Plus). If they were on a business account, their administrator must have enabled AI features.

3 Methodology & Results

To understand the real-world impact of Gemini in Workspace on knowledge worker productivity, we conducted three distinct studies in 2024 and 2025. This multi-pronged research approach, combining controlled usability testing with randomized experiments and large-scale A/B testing, allowed us to gather a robust, mixed-methods view of AI’s effects. In the following sections, we will detail the specific methodology and key results from each of these three studies, beginning with our in-person usability analysis.

3.1 Exploratory study on Archetype Journeys in Workspace

Our first study was an exploratory in-person usability study to understand the impact of artificial intelligence (AI) on common workplace journeys. A total of 51 business professionals from the New York City area were recruited, screened for familiarity with AI, and assigned to one of two conditions: a "No AI" group and a "Gemini in WS" group that was restricted to using Gemini features within Google Workspace. Everyone in the “Gemini in WS” group had prior familiarity with AI, and the “No AI group” was composed of a mix of levels of familiarity. Although the groups are not directly comparable, they indicate the kinds of differences we can expect from people working with AI. Each participant engaged in a 90-minute session where they were tasked with completing a series of 7 tasks (e.g., “Create a meeting agenda in Docs,” “Scan a chat space and email for project updates”), designed to reflect typical knowledge worker workflows across various Google Workspace applications.

To ensure consistency, all sessions were conducted in a controlled environment using the same laptop computer, and participants received printed instructions that were also read aloud by researchers. Researchers collected both qualitative and quantitative data. Key metrics included: time on task, task success rate, AI tool discoverability, and AI tool choice. Researchers also gathered self-reported data from participants on their perceived efficiency, satisfaction with the task outcome, and confidence in their own abilities. Qualitative observations focused on usability challenges, pain points, and participants’ ideal visions for AI assistance.

The findings indicate that using Gemini in Workspace boosts user satisfaction, confidence, and efficiency, particularly for tasks involving drafting and summarization, by providing a strong starting point and reducing the cognitive load of beginning a task. This perceived efficiency was often rooted in actual time savings - for example, users in the Gemini in WS condition spent 40% less time on the note taking task than users in the non-AI condition, a finding that echoes Google survey data on Gemini in Workspace users saving a significant amount of time per week on their tasks. A handful of tasks showed mixed results in time on task, as users spent additional time refining AI-generated outputs and providing repeated context. The primary friction point identified was the siloed nature of AI, where context is lost when moving between different applications, hindering cross-product workflows. The study concludes that AI outputs are viewed as a starting point rather than a final product, and that users desire an integrated, proactive, and context-aware assistant that empowers them while maintaining user control.

Average Ratings Across All Tasks

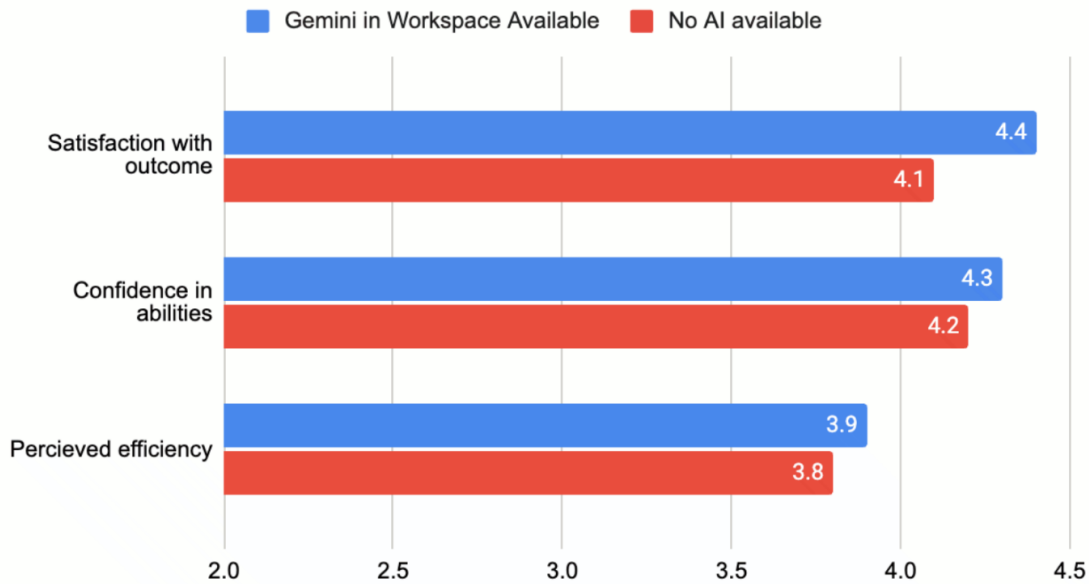


Figure 1: Satisfaction, Confidence, and Efficiency across all Tasks (Scale rating out of 5), Archetype Journeys in Workspace Study

3.2 AI Advantage for Small Businesses

The second study was a randomized experiment ($N = 320$) examining the utility of AI for small business users recruited via the *UserZoom* and *UserTesting* platforms (specialized platforms for User Experience Research). They were screened to be currently employed and using Google Workspace products for work at least once a week.

Participants completed three journeys, which were randomly selected from a full set of six. Each journey consisted of 2 – 3 tasks. Each journey took 4 – 10 minutes to complete for

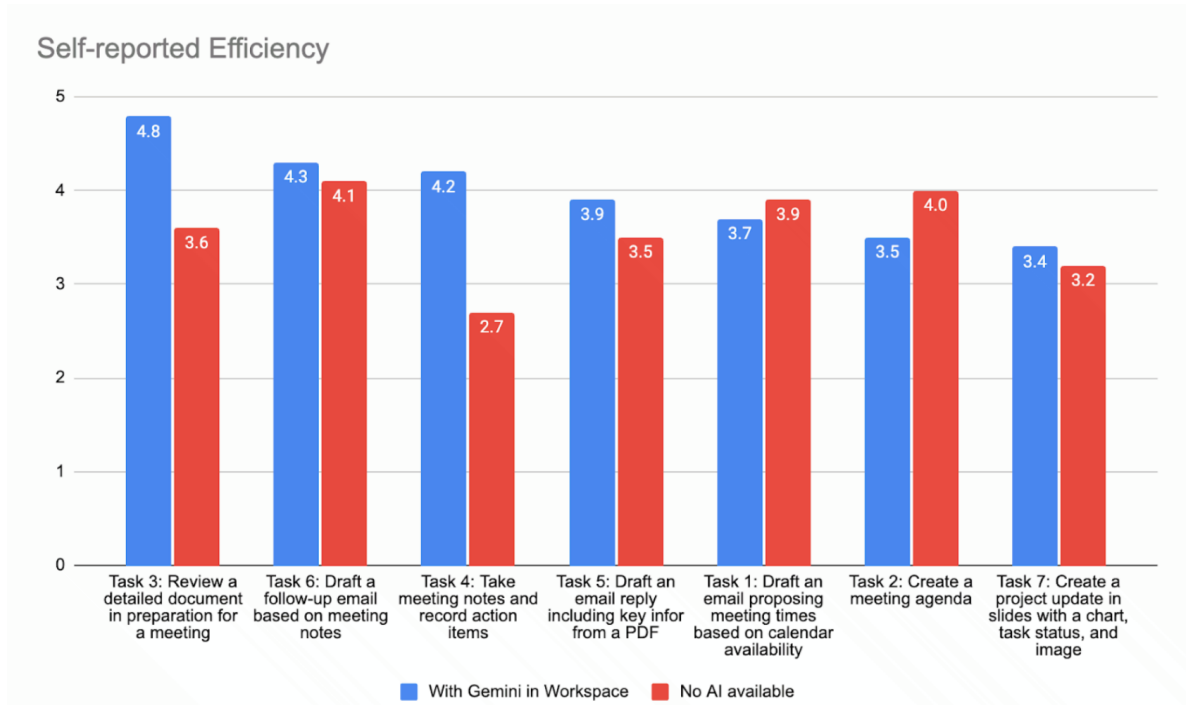


Figure 2: Self-reported Efficiency (Scale Rating out of 5), Archetype Journeys in Workspace Study

a typical participant. In total, including questionnaires and instructions, the exercise took 25–50 minutes. The tasks are centered around small business tasks, but are generally similar to those that non-specialized workers in larger firms engage in. We chose tasks identified as high-frequency, high-importance digital productivity tasks in the U.S. Department of Labor’s O*NET Database. Specifically, the tasks fell into four categories:

- **Analysis:** Tasks where participants had to decide something based on a spreadsheet.
- **Product navigation:** Tasks that required moving information between products and/or navigating a single product substantially. They did not require writing complete sentences to complete.
- **Email writing:** Tasks that required writing an email as a reply or based on other information.
- **Text processing:** Tasks that required reading and/or writing more than one paragraph of text.

Each of the three journeys participants completed was randomized to be completed with or without AI (referred to as condition). In the AI condition, users were told that they can use AI if they find it helpful and encouraged to use the workspace assistant. In the no-AI condition they were asked to abstain from using any AI tools (including the workspace assistant, or another). Some users also used AI even when explicitly instructed not to (20%). As a result, we can consider the results in the experiment to be lower-bound estimates on the effects of having access to an AI assistant.

We measured three key variables:

- **Self-reported satisfaction:** For each journey, users reported how satisfied they were with the output of their task after completing it.
- **Time on task:** Time taken to complete the entire journey was measured directly.
- **Correctness:** Raters blind to the hypotheses watched the screen recordings of the participants' attempts to complete the journeys and rated each sub-task as complete, partially complete or incomplete according to predetermined criteria for each sub-task.

The study's results show that AI assistance significantly improved productivity for text processing tasks. For example, in the "Research This Prospect" journey, which involved researching a company and writing a sales pitch, users with AI access were 25% more likely to complete the task successfully, saved an average of 132 seconds (20%), and reported 0.67 points higher satisfaction (on a 5-point scale). Similarly, in the "Managing and Facilitating Meetings" journey, which involved summarizing meeting notes and drafting a follow-up email, users with AI saved 156 seconds (32%) and reported 0.93 points higher satisfaction. For other, less text-intensive tasks, such as creating a simple invoice tracker or reformatting a table in a slide deck, there were no significant differences in completion time or satisfaction between the AI and no-AI groups.

Interestingly, in the AI condition, participants used AI most frequently in those tasks which benefited from AI, suggesting they were able to intuit when AI would be most useful. Specifically, for the text processing tasks, we saw 80% of participants using AI in the AI condition, whereas for email writing, product navigation and analysis, where there were no significant performance benefits, we saw 70%, 45%, and 33% of users using AI respectively.

Another major finding was the "leveling" effect. Less proficient users, as identified by their lower baseline completion rates, benefited more from AI assistance in terms of task completion, with a correlation of -0.60 ($p < .01$) between a user's baseline proficiency and the improvement they saw with AI. However, this effect was not observed for time on task or satisfaction, where users of all skill levels saw similar benefits.

3.3 Large Scale A/B test on Productivity Impact of “Help Me Create” on real users

In our third study, we conducted a randomized controlled trial, where 50% of a set of 23,481 business and consumer users with access to Google Workspace AI features were randomly given access to the “Help Me Create” feature in Google Forms two weeks early, in June of 2025. In this two week period, we measured the total time users with and without Help Me Create access spent editing their form. Specifically, from log sources, each time a user made a form, we measured the total number of minutes spent editing their form, from the time of the form creation to the time until they first clicked the button to publish the form, among forms that actually got sent out (whether a form was sent out was confirmed by only looking at forms which got at least one response by a user who was neither an owner nor an editor of the form). Edit actions include actions such as adding questions to a form, deleting questions, updating formatting, and so on. Note that all analyzes were performed on anonymized data and only aggregate results were examined.

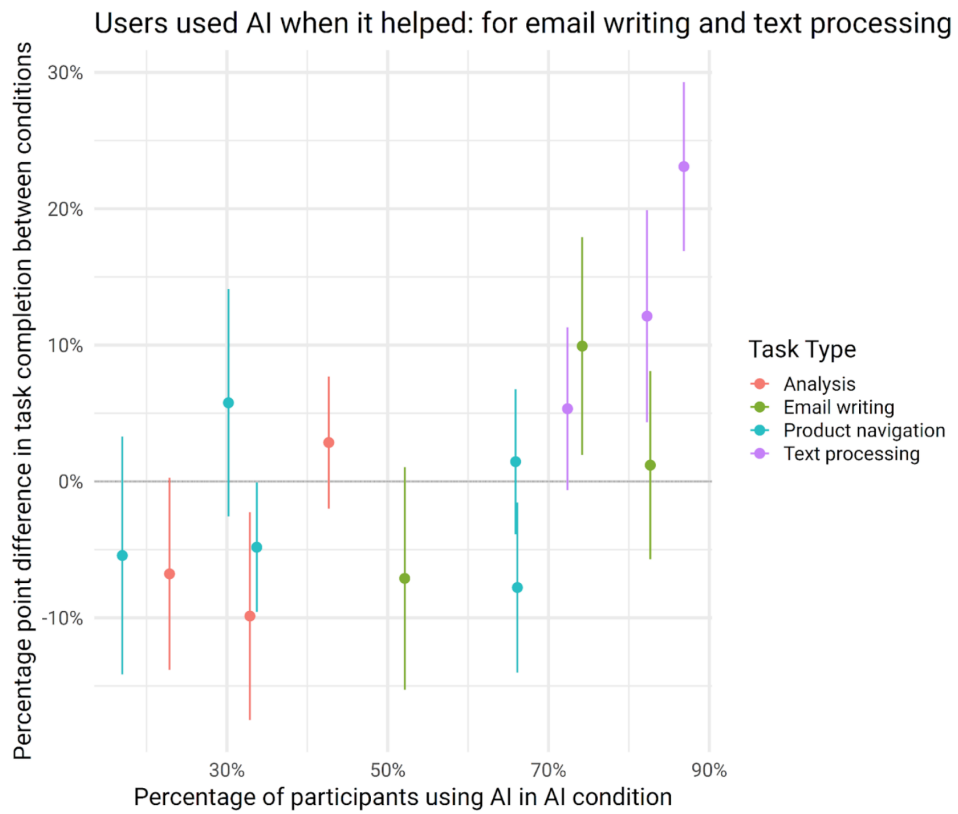


Figure 3: AI Success Rates by the Proportion of Participants Using AI, AI Advantage for Small Businesses Study

We found that providing access to Help Me Create has a causal effect on productivity, as measured by total minutes spent editing and minutes spent adding questions: The Intent-to-Treat (ITT) estimates were statistically significant, showing that users with access spent considerably less time editing and adding questions to their forms (both p 's < 0.001). Since not all users with access actually used the feature (one-sided noncompliance), we used the random assignment as an instrument for usage to estimate the Local Average Treatment Effect (LATE). For users who created a form with Help Me Create, total editing time was reduced by 27% (14.6 minutes vs. 10.8 minutes, $t = -4.4$, $p < 0.001$). This was mainly the result of a massive 68% reduction in time spent adding questions (4.6 minutes vs. 1.5 minutes, $t = -8.9$, $p < 0.001$), which was not offset by a slight increase in time spent deleting questions (0.0 seconds vs. 11.8 seconds, $t = 5.8$, $p < 0.001$).

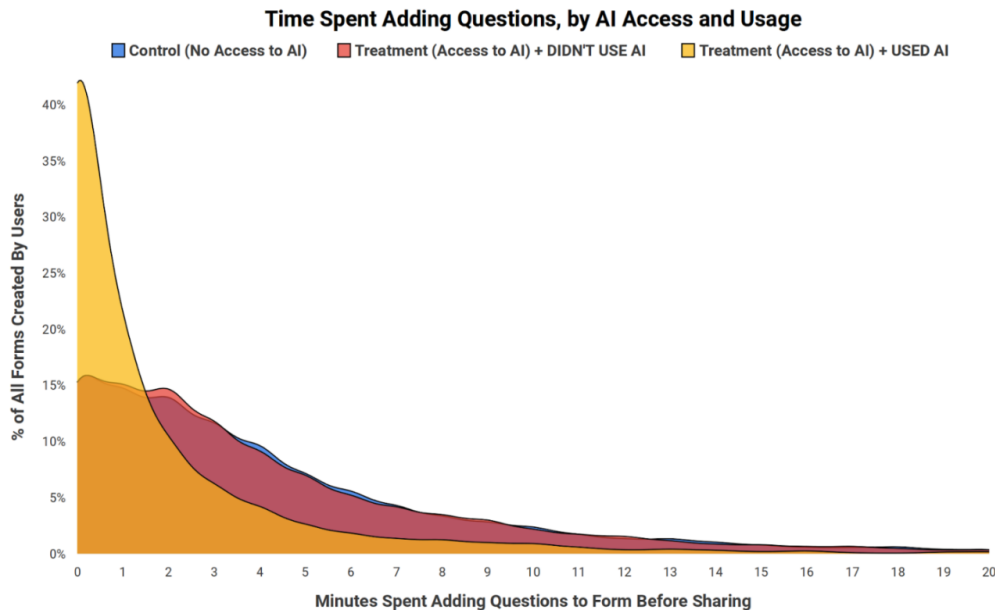


Figure 4: Time spent adding questions to Form before publishing, by condition and AI usage

4 Key Findings: Productivity Impacts of AI

The case studies discussed here depict the impact of AI across three key metrics: task success rate, completion time and user satisfaction. The first set of findings indicate that GenAI significantly improves the quality of user output and success rates, particularly for tasks involving ideation, summarization, and writing. Critically, these performance gains are most pronounced among less proficient users, “leveling” by narrowing the performance gap between users. The impact on task completion time is nuanced yet shows a clear pattern: AI delivers significant time savings on structured, repeatable tasks and complex, text-heavy work, while time on more open-ended creative tasks is static at first due to factors like iterative prompt refinement and user verification of results during the initial learning phase. After this initial learning phase, Workspace researchers observe longer-term reductions in

time on task. Finally, users with access to AI tools consistently perceive tasks as taking less time and report significantly higher rates of satisfaction and confidence.

4.1 Increase in success rates and skill gap reduction

A primary finding is that AI integration materially enhances task success and output quality. Across a range of tested activities, particularly those involving content generation (e.g., writing emails, drafting notes), users assisted by Gemini in Workspace achieved higher average success rates than those in the control group. The largest gains were observed on text-heavy tasks, with improvements in success rates of up to 33%. Moreover, a significant outcome of this performance enhancement is a “leveling” effect, where less proficient users experienced greater benefits from AI in terms of task completion, a finding that echoes recent studies in other environments.

Performance gains vary between task types: Gemini users in Workspace underperformed ‘No AI’ conditions in success rates on a handful of tasks that required combining inputs across multiple tools or generating highly tailored content, where participants iterated to get AI to reference the correct context. This reflects the concept of the “jagged technological frontier,” where AI excels at some tasks within a role, but is not yet capable of handling all of them, necessitating a continued human in the loop.

4.2 Improvements in time spent on complex and structured tasks

The impact of AI on efficiency, measured by time on task, reveals a clear pattern of significant time savings in several types of high-value work. For these tasks, AI usage led to statistically significant efficiency gains. For example, the “Help Me Create” study found that users creating forms with AI spent 27% less time editing them before publishing. Similarly, the AI Advantage for Small Businesses study demonstrated that AI delivered substantial time savings on complex, text-heavy tasks, reducing the time needed to research a sales prospect and write a pitch by 20% and cutting the time to draft a follow-up email from meeting notes by 32%. These results are in line with other studies: e.g., Paradis et al. (2024) estimate that AI drives a 21% time reduction for code developers at Google.

Largely because the majority of participants in the Archetype Journeys study were using Gemini in Workspace to execute on these tasks for the first time, AI usage did not reduce total time on task for a handful of multi-step creative tasks. This unfamiliarity led them to spend additional time on double-checking the AI’s output for accuracy, as well as greater-than-usual prompt refinement and feature exploration.

As noted above, other observations of Gemini in Workspace users indicate that workers save significant time per week after integrating AI into their workflow. This phenomenon, where new human-in-the-loop tasks emerge during the initial adoption phase, is a key consideration for measuring AI’s productivity benefit. It also suggests that as users build trust and familiarity with both the AI and the task, this verification overhead will decrease, unlocking further time savings. Interestingly, more efficient users experienced similar time-saving benefits as less efficient ones.

There was also a consistent and strong reduction in perceived time on task, regardless of the magnitude of time saved. Participants frequently reported that AI made tasks feel

faster and reduced the cognitive friction associated with starting a task (i.e., the "blank page problem") or reviewing long documents. This psychological benefit—reducing the perceived effort of a task—is a crucial, albeit distinct, value proposition of AI integration. This finding adds to the body of evidence from peer studies that found high self-reported productivity when using LLM-based tools, with perceived time saved substantially exceeding actual time saved.

4.3 User Satisfaction and Confidence

The most consistent finding is the powerful positive impact of AI on the user experience. Participants across nearly all tasks reported significantly higher satisfaction with their final output when using AI, even in cases where they spent considerable time editing the generated content. This suggests that users place a high value on the high-quality starting point provided by the AI, which they can then refine to meet their specific needs. Similar to success rate results, AI had the largest impact on text-based tasks such as writing a sales pitch or drafting emails (+36% and +19% increase in satisfaction), overcoming the cognitive load of starting from a blank page. Moreover, this increased satisfaction was coupled with a measurable boost in user confidence. By providing polished language, structured formats, and clear summaries, Gemini in Workspace empowered users, making them feel more capable in their roles.

As with time on task, the data did not reveal a "leveling" effect for satisfaction or confidence. The experiential benefits of using AI appear to be universal across the user skill spectrum, enhancing the work experience for both high and low-performing individuals alike. This broad-based improvement in job satisfaction and self-efficacy is a critical, and often overlooked, component of AI's overall productivity contribution.

5 Opportunities to enhance AI tools to drive knowledge worker productivity

These studies' findings demonstrate a promising trajectory for embedded AI assistants in digital productivity tools, alongside a clear set of opportunities to deepen their value to users. While users consistently found that AI enabled stronger writing, better summarization, and an ability to get started on complex tasks, their feedback also highlighted ways in which these tools can evolve to better match the realities of present-day knowledge work. Below, we detail a few of these opportunities for consideration by fellow tool-builders and researchers.

5.1 Continuity across contexts

Knowledge work typically spans multiple applications: an email may reference a document, which links to a spreadsheet, which lists calendar events, which need to be shared on a chat thread. While Gemini-based tools provide in-app support that feels native and accessible, experiment participants imagined even greater potential: AI that can follow their intent across apps, and surface relevant content proactively, based on their actions in other

apps. Enabling persistent context across apps for knowledge workers represents a significant opportunity to reduce friction and unlock more productive workflows.

5.2 Refinement as a “first class use case” alongside generation

The ability of Gemini-based features to generate strong first drafts was appreciated across these experiments. Many users noted that it helped them get started faster and produce more polished work. At the same time, participants often engaged in iterative refinement of AI-generated outputs, which led to mixed results in total time on tasks. This points to an opportunity to build more interactive and assistive workflows, in which AI suggests edits, adapts to tone or formatting needs, and provides multiple draft options all while highlighting areas of uncertainty as a high-quality human collaborator might. AI in apps for knowledge workers has potential to continue to grow from being a “first-pass” generator to an active partner providing in-situ editing and guidance, one that suggests quick edits rather than requiring full rewrites.

5.3 Prompt coaching

Experienced users generally knew how to prompt Gemini features effectively. Those newer to AI tools relied on trial-and-error, which again contributed to longer time on task. This reveals a valuable opportunity to embed additional prompt scaffolding directly into the user experience: suggest prompts based on task context, offer examples of high-impact phrasing, and explain how small changes to input can shape output. In short, consider thinking of AI in productivity applications as both a teacher as well as an assistant.

5.4 Transparent reasoning

Participants had high aspirations for Gemini’s capabilities and fully engaged with the tools. At the same time, they expressed a consistent desire to verify that the assistant “knows what it knows.” This suggests that trust in AI tools isn’t an all-or-nothing proposition. While users will engage with AI assistants they may not fully trust, they will likely engage more efficiently and expansively when trust is fully established. AI for knowledge workers can accelerate this process of trust calibration by clearly showing users what sources are being referenced (using capabilities like those featured in NotebookLM), highlighting uncertain content, and offering options to trace or revise AI reasoning.

6 Conclusion

Our research provides grounded, empirical evidence that embedded AI assistants improve both task success and user satisfaction for knowledge workers. A central finding is a significant leveling effect, where AI tools narrow the performance gap between less and more proficient users, particularly on text-intensive tasks. By reducing cognitive load associated with challenges like the “blank page problem,” AI makes complex work more approachable and less mentally taxing. This consistently results in increasing user confidence and higher

satisfaction with final outcomes, even when users invest time in refining the AI’s output. We also found significant time savings for highly structured work.

While these findings offer a clear snapshot of AI’s current value, it’s important to acknowledge the limitations of this work and identify avenues for future inquiry. The studies presented here are cross-sectional, capturing the productivity impact at a specific point in time. They do not fully capture the long-term effects as users move past the initial learning curve and become more adept at collaborating with AI. Furthermore, 2026 marks a milestone for Gemini in Workspace as the platform undergoes significant enhancements. These developments will warrant further research to evaluate likely additional gains for knowledge worker productivity.

These limitations point to several critical areas for future research. Longitudinal studies are needed to understand how these productivity effects evolve as both user skills and AI capabilities mature. Research is also needed to explore the organizational changes required to translate individual task improvements into tangible, firm-level productivity gains. Finally, future work could investigate the impact of AI on the evolution of skills, job design, and the qualitative aspects of work, such as creativity and team collaboration. Understanding these dynamics will be essential as AI becomes more intricately woven into the fabric of daily work.

References

- Bick, Alexander, Adam Blandin, and David J. Deming (2025). “The Rapid Adoption of Generative AI”. In: *NBER Working Paper Series*.
- Bonney, Kathryn et al. (2024). “Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey”. In: *NBER Working Paper Series*.
- DORA (2025). *State of AI-assisted Software Development 2025 Report*. <https://dora.dev/research/2025/dora-report/>. Available at: <https://dora.dev/research/2025/dora-report/>.
- Hartley, Jonathan et al. (2024). “The Labor Market Effects of Generative Artificial Intelligence”. In: *Mercatus Working Paper*.
- Paradis, Elise et al. (2024). “How much does AI impact development speed? An enterprise-based randomized controlled trial”. In: arXiv: 2410.12944 [cs.SE]. URL: <https://arxiv.org/abs/2410.12944>.

Author Contributions

This paper represents the views of Google’s AI and Economy Program. **Scott Strand** conceived this paper and was the lead author. **Nacho Cigliutti** synthesized the studies’ key findings and connected them to the broader corpus of economic research. **Zanna Iscenko** provided important inputs on framing and structure. **Fabien Curto Millet** gave critical guidance on the articulation of key arguments and oversaw the review process. **James Manyika** provided valuable support and discussion.

Data was drawn from a series of studies conducted by the Google Workspace research team. **Christyn Courtney** oversaw this group of studies, steering the methodologies and

analyses. **Sebastian Deri** led the large scale randomized controlled trial to measure the effect of AI-usage on time-savings in Form creation, designing the experiment, building the data pipelines, and analyzing and reporting the outcomes. **Aleks Sinayev** led the AI Advantage for Small Businesses study, designing the experiment and analyzing and reporting the outcomes. **Kristin Moore** designed, implemented, and reported the Archetype Journeys Study. **Robert Youmans** provided essential guidance and subject matter expertise.