

Evaluating Artificial Intelligence users needs

Learning about employees' experience and support needs while using AI

Summary

The goal of this evaluation was to understand how individuals in the U.S. General Services Administration's (GSA) Gemini artificial intelligence (AI) pilot interacted with AI tools to inform GSA's future investments and strategy. We found that frequent AI users' experience varied, they were aware of risks associated with getting inaccurate output, and they wanted more training resources and support.

Agency priority

GSA is a leader in modernizing and streamlining technology across government, including promoting responsible AI innovation in support of the Administration's [Executive Order](#) to accelerate federal AI use. Providing GSA staff AI tools to support their daily work has the potential to increase efficiency and productivity. It is a GSA priority to learn from AI pilots to inform strategies that support responsible AI innovation during agency-wide rollout, and ultimately enhance government efficiency.

What we evaluated

We partnered with GSA's Office of Information Technology (IT) to evaluate how users in GSA's Gemini AI pilot interacted with AI tools to inform GSA's future AI investments, strategies, and support. GSA IT launched the pilot in October 2024 and gave about 250 GSA employees access to Gemini in a secure test environment for 90 days. At the pilot's midpoint, we interviewed 10 of the top 13 most frequent users of Gemini to understand how they interacted with the tool, what they used it for, and types of implementation challenges.

What we learned

Even among frequent users, we found that experience, expectations, and knowledge of AI tools varied significantly. For example, while users largely expected some iteration with Gemini to get the output they wanted, their approach to that back and forth

varied in terms of tolerance for refinement and conversational style. Most interaction happened through the web app, rather than workspace integrations.

Users found that Gemini worked best at concrete, bite-sized tasks, such as helping with coding. Similarly, users found that breaking down a task, such as asking Gemini to summarize specific parts of a document rather than the whole document, was more effective than asking Gemini to take on a more complex or abstract task.

When asked about risks, users were most aware of the risk of getting inaccurate or incomplete information from Gemini, and stressed the importance of training and having a human review Gemini's output. However, the process by which users checked Gemini's outputs varied depending on their use case and experience.

Users wanted more examples of ways they can use Gemini and more opportunities to learn from their peers. A few users suggested providing clear guardrails, rules, and GSA specific examples to help others feel confident in experimenting with Gemini to develop novel and effective use cases.

Applying the findings

- Given that most interaction happened in the web app, meet demand by providing a general-use chatbot before workspace integration (e.g., AI embedded into Google Docs).
- Since users vary in how they interact with AI tools, future training and tools could be tailored based on characteristics including previous AI use, primary use case/goals in using AI, and frequency of use.
- Help users develop rules of thumb about checking the accuracy of AI output. Develop guardrails, reminders, and/or defaults to manage other risks (e.g., data security, bias, malicious output), which users may be less aware of.
- In the roll out of future AI tools, focus on: 1. validating the benefits to users (e.g., time savings), 2. understanding how variations in user behavior affect those benefits, and 3. understanding how users integrate AI output into their work.