Unlimited AI tokens aren’t unlimited after all as US Army burns through supply

In order to use Ask Sage, the Army had access to 100,000,000 tokens as part of an annual subscription to an “enterprise pack.” Tokens represent a unit of output, either in text or image, from an LLM. For the Ask Sage tool, a single token equates to about 3.7 characters, according to documents viewed by WIRED. The Defense Department burned through some 20 billion tokens per day during the 38-day Operation Epic Fury in Iran, according to Breaking Defense.

The Army and DOD didn’t reply to requests for comment; neither did Ask Sage.

It’s unclear if the tokens used by regular DOD employees are drawn from the same pool as those who might be using AI tools on classified or secret information. This hasn’t stopped the Defense Department’s emphasis on AI. On Monday, the Intercept reported that the Pentagon has continued to lean into AI tools, and has cut the staff at the Civilian Protection Center of Excellence, whose jobs entailed preventing civilian casualties in conflict zones. Instead, the DOD is developing an AI tool to speed up the assessments that the Center’s staff would normally make.

The Army is not the first eager adopter of generative AI to rethink their near unlimited use. After encouraging employees to “tokenmaxx,” Meta quietly took down its leaderboard tracking token usage and is now trying to curb use. Last week, Adam Mosseri, head of Instagram at Meta, floated the idea of capping token use per engineer at the company. According to

reporting from Fortune, Uber also saw its engineers burning through a year’s worth of generative AI tokens in merely four months.

The Army employee says they have not found the generative AI tools to be particularly useful for their work, and that when they have used the tools, they have found them to be unreliable. One model even asserted that it had completed a task that it hadn’t, they say. “I think there are definitely several aspects of the bureaucracy of the US federal government that these tools might be helpful with. But an unthinking application and use is not going to result in an effective, efficient, and trustworthy rollout.”

This story originally appeared on wired.com.

https://arstechnica.com/ai/2026/07/us-army-faces-ai-use-limits-after-exhausting-years-supply-of-ai-tokens/