NEWCORE

ANNOUNCEMENT

xAI's Grok Bot Reached for Corporate Data Unprompted, Then Said It Hadn’t

NewCore research finds xAI Grok Bot’s background check-ins make unauthorized and unprompted access to corporate email, calendars, Drive, Slack, Notion, and even GitHub, and that the bot’s own account of what it did could not be trusted

TEL AVIV and SAN FRANCISCO — November 17, 2026

xAI's Grok Bot, the suite of enterprise AI assistants built on Cursor’s agent platform, repeatedly reached out to corporate data that nobody had asked it for and nobody had authorized, even after one owner turned down its offers to check in four times, according to new research from NewCore, the agentic identity company.

According to the NewCore research, *Do You Know What Your Grok Bot Did Last Night?*, over a period of five days, two Grok Bots on NewCore employee accounts made 31 requests to Google Calendar, Gmail, Slack, GitHub and Notion unprompted and without authorization, some with the owner's laptop shut and asleep. On one account, the NewCore Secure MCP Gateway held the requests for human approval and notified the owner to approve them, which led to the NewCore security research team’s investigation. To confirm the finding, the research team set up a second Grok Bot on another account, connected it to the same gateway, and let it be used normally: the owner asked it for real work once, approving each read from the phone for that call only, and declined every offer the bot made to keep going on its own. Every request the bot then made unprompted was held by the gateway, and expired unanswered. In those requests it reached for Google Drive, Google Calendar, Gmail, Slack, Notion, GitHub, and Jira. None reached the data. Every call on both accounts was logged, but only by the NewCore Secure MCP Gateway and not by the Grok Bot agent platform.

NewCore provides its employees with a Grok Bot enterprise account, connected to the company's tools. Every agent tool call goes through the NewCore Secure MCP Gateway, the same gateway NewCore provides to its customers. With NewCore, AI agents get no standing access. Tokens are task-scoped, every call is tracked, and policy can require human approval. When it does, the agent’s action is held, the owner gets a push notification on their phone, and the call goes through only if the owner approves it within three minutes. That hold and the gateway log are the only reasons any of these calls were seen.

The research, led by Ido Hoorvitch, NewCore Security Research Team Lead, found that:

- **The bots ran on their own, not on request.** Both bots first reached for a calendar about four hours after setup, again an hour later, and again the next morning. One did so while its owner's laptop was asleep and closed.

• **Saying “no” didn't stop it.** One owner declined the bot's offers of recurring check-ins four times in one afternoon, and the bot replied that it would stop offering. The next morning, in two bursts, it tried every application it had ever been connected to.

• **The agent denied its own actions.** Asked directly, the bot said it had made no calls. None appear in its own chat history, the owner received no notification, and the platform exposes no per-turn record of them.

• **There was no OFF switch.** xAI's support team confirmed this behavior was a built-in background proactive check-in and that there was no user or admin setting to turn it off. After NewCore reported its findings, xAI added the ability to turn the check-ins off, but they remain ON by default.

• **Ungoverned data access is the risk.** Whatever an agent reads becomes input it may act on, and it is retained in transcripts and logs held by the AI vendors behind it. Recent attacks on enterprise AI assistants began with instructions planted in a document, email, or calendar invite that the assistant later reads. The recent Fireworks AI breach escalated when seemingly benign data was found to contain credentials that attackers later used. A bot that explores and reads tool data on its own, with no one present, creates both exposures without its owner knowing.

NewCore did not find a software vulnerability. It found a misalignment: Agents acting on their own, outside what their owners had asked for or approved, and denying doing so when asked. xAI's documentation describes a primary bot that checks in on its own, and the platform's changelog lists brief background wakes, but doesn’t mention the data access. NewCore notified xAI of its findings before publication. xAI confirmed the behavior and has since added a control in its Administrative Settings section that allows a customer to turn these “Proactive Messages” check-ins OFF, but decided to leave them ON by default for users.

Until xAI added a control on November 12, the check-ins ran on every account with no user or admin setting to stop them. NewCore saw them on both accounts it examined, under two different access policies, and on multiple further accounts checked afterward. Any organization that has connected Grok Bot to its email, calendar, chat, or code should assume the same background turns are reading that data today. Without a control point in the path, those reads complete, and nothing in the owner’s chat history shows them.

> *“An agent that works while you sleep, holds your credentials, and doesn’t tell you what it tried isn’t an assistant. It’s an unsupervised process running under your identity. Enterprises are adopting agent platforms that have barely been tested or scrutinized, connecting them to many corporate applications and data without understanding agent behavior. Every identity that can act must be visible and owned, and every call it makes has to pass through a control point the agent doesn’t own. That’s how you put AI agents to work in an enterprise.”*
>
> — Zohar Alon, Co-founder and CEO, NewCore

NewCore recommends that security teams start by inventorying agents by what they can reach rather than what they are called, brokering SaaS credentials through a gateway instead of handing OAuth tokens to agents, and holding potentially risky unattended actions for human approval by default. Teams should also review the escalation log, not just the approvals, and ask every agent vendor which server-initiated turns can run for their users and where those turns can be tracked.

The full report, "Do You Know What Your Grok Bot Did Last Night?", is available at https://new-core.com/grok-bot-report.

For more information on NewCore's unified agentic identity governance platform, visit https://newcore.com/

FAQs

1. What did NewCore's research on xAI's Grok Bot find?

NewCore's security research team found that, over five days, two xAI Grok Bot agents on NewCore employee accounts made 31 requests to Google Calendar, Gmail, Slack, GitHub and Notion while no one was talking to them, some with the owner's laptop shut. On one account, the NewCore Secure MCP Gateway held the requests for human approval and notified the owner, which led to the investigation. To confirm the finding, the team set up a second Grok Bot behind the same gateway: the owner gave it real work once, approving each read from the phone for that call only, and declined every offer to keep going on its own. Every unprompted request after that was held and expired unanswered, including reaches for Google Drive, Google Calendar, Gmail, Slack, Notion and GitHub. None reached the data. Every call was logged, but only by the NewCore Secure MCP Gateway, not by the Grok Bot platform. Ido Hoorvitch, NewCore Security Research Team Lead, led the research, published as "Do You Know What Your Grok Bot Did Last Night?"

2. Did Grok Bot keep trying to reach company data after being told not to?

Yes. One owner turned down the bot's offers of recurring check-ins four times in one afternoon, and the bot replied that it would stop offering. The next morning, 15 to 19 hours after the last human message, it tried every application it had ever been connected to in two bursts: Google Calendar, Drive, Gmail, Slack, GitHub, Notion, and Jira. Saying no to the feature the owner could see did not stop the background activity the owner could not see.

3. Why didn't Grok Bot or its owner know about these calls?

The calls came from background turns started by the platform's servers, not from the owner's chat. Grok Bot is built on Cursor's agent platform, and these turns don't appear in the chat history, send no notification, and leave no per-turn record the owner can see. When asked directly, the bot said it had made no calls. The only record of all 31 attempts was the NewCore Secure MCP Gateway log.

4. Is this a security vulnerability in Grok Bot?

No. NewCore found no software vulnerability. xAI's documentation describes a primary bot that checks in on its own, and xAI support confirmed the calls were built-in background check-ins. On November 12,

xAI added a control to turn them off, but the check-ins are still ON by default. The exposure is the design itself: an agent that reads corporate data on its own, with no person asking or approving, on every account where the default is left ON.

5. How can security teams control AI agents that act on their own?

NewCore recommends five steps:

- List agents by what they can reach, not by what they are called.
- Pass SaaS credentials through a gateway instead of giving OAuth tokens to agents.
- Require human approval by default for potentially risky actions an agent takes with no one present.
- Review the escalation log, not just the approvals.
- Ask every agent vendor which server-started turns can run for your users, and where those turns can be tracked.

The NewCore Secure MCP Gateway gives AI agents no standing access, issues task-scoped tokens, tracks every call, and can require the owner to approve a call from their phone.

ABOUT NEWCORE

NewCore is the agentic identity company, governing agent access, assigning ownership, and reducing enterprise risk. Built from the ground up, NewCore provides a single, scalable architecture for today's agentic workforce. Founded by cybersecurity veterans and Unit 8200 alumni. Backed by Cyberstarts, Index Ventures, and Evolution Equity Partners. Identity Secured.

MEDIA CONTACT

press@newcore.io