AI Agenda

# OpenAI Develops Features to Counter Grok Bot, Mulls Response to Meta's Muse

# OpenAI 开发功能以应对 Grok 机器人，考虑回应 Meta 的 Muse



Art via Getty Images; Mike Sullivan

图片由 Getty Images 提供；Mike Sullivan



By Amir Efrati and Stephanie Palazzolo 以及 Stephanie Palazzolo

Sep 21, 2026, 7:30am PDT

Listen Share

**OpenAI pioneered AI** that can take over web browsers and other applications on people's behalf, and its **ChatGPT** app brought generative AI to consumers. Now that rivals like **SpaceX** and **Meta** are pushing the boundaries by launching personal AI agents that handle multistep tasks, OpenAI is playing catch-up.

**OpenAI 率先开发出能够代表用户接管网页浏览器及其他应用程序的 AI，其 ChatGPT 应用也将生成式 AI 带给了消费者。如今，随着 SpaceX 和 Meta 等竞争对手通过推出能够处理多步骤任务的个人 AI 代理不断拓展边界，OpenAI 正努力迎头赶上。**

For instance, OpenAI is developing features to compete more directly with SpaceX's recently launched **Grok Bot**, which helps customers create always-on AI "teammates" to handle routine tasks, manage software, and collaborate in the background, according to a person with knowledge of the effort. OpenAI also has discussed developing a personal AI assistant to compete with one that Meta Platforms recently launched, according to another person with knowledge of the situation.

据一位知情人士透露，例如，**OpenAI** 正在开发相关功能，以更直接地与 **SpaceX** 近期推出的 **Grok Bot** 展开竞争。**Grok Bot** 可帮助客户创建始终在线的 **AI**“队友”，在后台处理日常任务、管理软件并开展协作。据另一位了解情况的人士透露，**OpenAI** 还讨论过开发一款个人 **AI** 助手，以与 **Meta Platforms** 近期推出的产品竞争。

In developing such products, OpenAI would be repurposing and rebranding some of the existing “agentic” technology it already offers through **Codex** and ChatGPT. Some customers of these OpenAI apps have already connected them to their email, calendar and other applications so they can automate tasks such as scheduling appointments or communicating with businesses by email, sometimes over multiple days.

在开发此类产品时，**OpenAI** 将重新利用并重新包装其已通过 **Codex** 和 **ChatGPT** 提供的部分“代理式”技术。**OpenAI** 这些应用的一些客户已经将其连接到电子邮件、日历和其他应用程序，以便自动执行安排预约或通过电子邮件与企业沟通等任务，有时任务会持续数日。

And OpenAI has said that since **GPT-5** last year, each successive AI model has been designed to tackle longer-

successive AI model has been designed to tackle longer horizon tasks, including in those apps.

**OpenAI 表示，自去年推出 GPT-5 以来，其后续每一款 AI 模型的设计目标都是应对时间跨度更长的任务，包括在这些应用中的任务。**

But many OpenAI users aren't necessarily aware of these features, and it's safe to say SpaceX and Meta have stolen some thunder with their respective agent products in recent weeks.

**但许多 OpenAI 用户未必了解这些功能，可以肯定地说，SpaceX 和 Meta 凭借各自的智能体产品在最近几周抢走了一些风头。**

Grok Bot, launched in August, aims to help people quickly set up bots that specialize in different tasks, such as performing market research or developing an ad campaign. Customers can give the bots access to their work applications, such as **Gmail** and **Google Calendar**, **Notion**, **Figma**, or accounting, finance, AI coding assistants or other software so they can perform tasks within those apps. Grok Bot customers can also tell the AI to watch them perform certain tasks with various apps so it can learn to repeat them on its own.

**Grok Bot 于 8 月推出，旨在帮助人们快速设置专门执行不同任务的机器人，例如进行市场调研或制定广告活动。客户可以授权这些机器人访问其工作应用，例如 Gmail 和 Google Calendar、Notion、Figma，或会计、金融、AI 编程助手及其他软件，从而让它们在这些应用中执行任务。Grok Bot 客户还可以让 AI 观察他们使用各种应用执行特定任务，以便 AI 学会自主重复这些操作。**

The bots can report back when they finish a task or need input from the customer. Customers can also require the bots to seek approval before taking actions, such as sending an email, making a purchase or financial transfer, or changing data in another app.

**机器人完成任务或需要客户提供信息时，可以向客户反馈。客户还可以要求机器人在采取行动前先寻求批准，例如发送电子邮件、进行购买或资金转账，或更改另一款应用中的数据。**

## Heavy Advertisement 重金投放广告

SpaceX has been positioning Grok Bot—which is included in **SuperGrok** subscriptions that start at $30 a month—as a kind of AI teammate or digital chief of staff for white-collar

workers and small business owners. The company says its bots can operate around the clock, even when the customer's computer is shut down. It has been advertising Grok Bot heavily across social media, YouTube and podcasts, including the Dwarkesh Podcast focused on AI.

**SpaceX 一直在将 Grok Bot 定位为白领和小企业主的某种 AI 队友或数字幕僚长。Grok Bot 包含在 SuperGrok 订阅服务中，起价为每月$30。该公司表示，其机器人可以全天候运行，即使客户的电脑处于关机状态。SpaceX 一直在社交媒体、YouTube 和播客上大力宣传 Grok Bot，其中包括专注于 AI 的 Dwarkesh Podcast。**

Similarly, Meta this month launched **Muse** to automate tasks such as communicating with businesses, answering emails, shopping for products and other things people might do from their phone. Meta's move followed the early rise of personal assistant app Instinct, which is available by invitation only but has more than 100,000 users.

**同样，Meta 本月推出了 Muse，用于自动完成与企业沟通、回复电子邮件、购买产品以及人们可能通过手机完成的其他任务。Meta 的这一举措紧随个人助理应用 Instinct 的初步崛起之后。Instinct 目前仅限受邀使用，但用户数已超过 100,000。**

**Instinct**, which launched before Muse, is trying to raise capital at a $10 billion valuation. It gained attention because it represented a potential breakthrough in AI for consumers after ChatGPT growth slowed dramatically over the past year. (The chatbot earlier this summer surpassed 1 billion weekly active users, a target OpenAI had aimed to hit last year.)

**Instinct 在 Muse 推出之前便已上线，目前正试图以$10 billion 的估值筹集资本。它之所以受到关注，是因为在过去一年 ChatGPT 增长大幅放缓之后，Instinct 代表了面向消费者的 AI 领域潜在突破。（今年夏初，这款聊天机器人超越了 1 billion weekly active users，达到了 OpenAI 原本计划于去年实现的目标。）**

While consumers have gravitated to AI chatbots like ChatGPT and Google's **Gemini** to retrieve information and get advice or help with a variety of projects, such chatbots haven't had as much success in handling more complex, personal tasks such as booking services online.

**尽管消费者纷纷转向 ChatGPT 和 Google 的 Gemini 等 AI 聊天机器人，以获取信息并就各种项目寻求建议或帮助，但这类聊天机器人在处理在线预订服务等更复杂、更**

具个性化的任务方面，尚未取得同样的成功。

Tech entrepreneurs and venture capitalists raved online about Instinct's ability to complete a range of tasks of varying complexity, including managing calendars and ordering supplies for a wedding. Users communicate with Instinct through messaging apps, and can give it access to credit card information and passwords to allow the AI bot to shop or book travel.

科技创业者和风险投资家在网上盛赞 **Instinct** 完成各种不同复杂程度任务的能力，包括管理日历以及为婚礼订购物资。用户通过即时通信应用与 **Instinct** 沟通，并可以向其提供信用卡信息和密码，以便该 **AI** 机器人购物或预订旅行服务。

It remains to be seen how many consumers are ready to give such AI agents access to personal information so they can perform useful tasks.

目前尚不清楚有多少消费者愿意让这类 **AI** 智能体访问个人信息，从而使其能够执行有用的任务。

Meta has billions of users it can put Muse in front of, and in

recent days Meta has promoted the assistant prominently in apps like **Messenger**. However, Meta's distribution advantage didn't work so well in 2024 when the company launched a chatbot that aimed to compete with ChatGPT.

**Meta 拥有数十亿用户，可以将 Muse 推送给他们；近几日，Meta 还在 Messenger 等应用中大力推广这款助手。然而，Meta 的分发优势在 2024 年并未发挥理想作用，当时该公司推出了一款旨在与 ChatGPT 竞争的聊天机器人。**

It isn't clear when OpenAI might launch features to counter Grok Bot and Muse, especially as it balances an endless list of priorities, including ensuring the safety or alignment of its next AI models (and the "swarms" of agents they can power). But it also can't afford to let competitors grab the AI agent mantle.

**目前尚不清楚 OpenAI 何时可能推出针对 Grok Bot 和 Muse 的功能，尤其是在它需要平衡一长串优先事项之际，其中包括确保下一代 AI 模型（以及这些模型能够驱动的"智能体群"）的安全性或对齐性。但它也无竞争对手夺走 AI 智能体这一主导地位。**




*Here's what else is going on...*

*以下是其他正在发生的事情.....*

## What OpenAI Researchers See That We Can't

## OpenAI 研究人员看到而我们看不到的东西

In the debate over how scared we should be that AI may "kill us all" one day, some have been quick to point out that AI has yet to achieve milestones we might expect before it's capable of taking over the world: boosting GDP, curing diseases, or automating a substantial number of jobs.

在关于我们应该多么担心 AI 有朝一日可能"杀死我们所有人"的争论中，一些人很快指出，AI 尚未实现我们可能认为其具备接管世界能力之前应达到的里程碑：推动 GDP 增长、治愈疾病，或实现大量工作自动化。

One factor that could help to explain the gap between the public's experience with AI and the strongly-worded warnings from researchers in the field is the fact that those employees often use new models many months before they reach the masses, as we detailed in this story this morning. The piece focuses on previously unreported talks between **Anthropic** and **OpenAI** around testing each others' models for safety.

一个可能有助于解释公众对 AI 的实际体验与该领域研究人员措辞强烈的警告之间差距的因素是：正如我们在今天早上的这篇报道中详细介绍的那样，这些员工通常会在新模型面向大众推出前数月使用它们。该报道聚焦于此前未曾报道的 Anthropic 与 OpenAI 之间的讨论，双方围绕相互测试模型的安全性展开交流。

Researchers tell us AI has largely automated the process of training new experimental models. Moreover, they also said models within OpenAI today largely write the programs needed to run or train models on graphics processing units (otherwise known as **GPU kernels**) and the optimizations for those programs.

研究人员告诉我们，AI 已基本实现了新实验模型训练流程的自动化。此外，他们还表示，如今 OpenAI 内部的模型大多能够编写在图形处理单元（graphics processing units，亦称 GPU kernels）上运行或训练模型所需的程序，以及针对这些程序的优化代码。

Engineers can provide the model with a single example of the type of optimization that they want and the AI can run for weeks to implement such optimizations, one OpenAI employee said. This level of AI-powered automation only

employee said. This level of AI-powered automation only became possible in the last few months thanks to improvements in models, the employee said.

一名 OpenAI 员工表示，工程师可以向模型提供一个他们希望实现的优化类型的示例，而 AI 则能运行数周来实施此类优化。该员工称，这种由 AI 驱动的自动化水平直到过去几个月才成为可能，这得益于模型性能的改进。

In another example, the OpenAI employee also said that it's not uncommon for employees' agents to work together to solve problems without ever looping in their human users.

另一个例子是，这名 OpenAI 员工还表示，员工的智能体相互协作解决问题，却始终不让其人类用户介入，这种情况并不少见。

Sometimes that can backfire, though: the staffer said that they had noticed times where employees would, for instance, ask their agents to make changes to the company's codebase and those agents would message other employees on **Slack** and ask them to fix bugs in the code the agents found, even if their user didn't ask them to do so or if fixing the bugs weren't relevant to their work. (Imagine if you asked your agent to complete a task for you and you opened Slack to see your agent chewing out your

coworkers for their lazy work!)

不过，有时这也会适得其反：该员工表示，他们曾注意到这样的情况——例如，员工让自己的智能体修改公司的代码库，而这些智能体会在 Slack 上联系其他员工，要求他们修复智能体发现的代码漏洞，即使用户并未要求智能体这样做，或者修复这些漏洞与相关员工的工作并无关系。（想象一下：你让智能体替你完成一项任务，结果打开 Slack，却发现你的智能体正在斥责同事工作懒散！）

A lot of these concerns have to do with **reward hacking**, or the idea that an AI might find loopholes or take unexpected steps to complete a goal its user gives it. That's different from an AI acting maliciously on purpose or a bad actor using AI to commit a crime.

许多这类担忧都与“奖励劫持”（reward hacking）有关，即 AI 可能会找到漏洞，或采取意料之外的步骤，以完成用户赋予它的目标。这不同于 AI 蓄意实施恶意行为，也不同于不法分子利用 AI 犯罪。

OpenAI has taken steps to prevent this type of behavior, we reported. You could imagine, for instance, researchers adding to an AI's prompt instructions telling it not to

extrapolate beyond the exact instructions its user gives it. That's a careful balancing act, however, because you can also imagine that users might not want to have to write out every single step it wants an agent to take.

正如我们报道的那样，OpenAI 已采取措施防止此类行为。例如，你可以设想，研究人员会在 AI 的提示词中加入指令，要求它不要超出用户明确给出的指令进行推断。不过，这需要谨慎权衡，因为你也可以设想，用户可能不希望必须逐一写出希望代理执行的每个步骤。

There are signs that the rate of AI improvements could accelerate. One reason is OpenAI's increased access to servers. The ChatGPT-maker's research team has developed numerous techniques or tricks they have wanted to implement in new models but didn't previously have enough servers to test those theories quickly. Now they do, and what might have previously taken years to pull off now happens in just a week, employees say.

有迹象表明，AI 改进的速度可能会加快。其中一个原因是 OpenAI 能够获得更多服务器资源。ChatGPT 的开发商的研究团队已经开发出许多他们希望应用于新模型的技术或诀窍，但此前没有足够的服务器来快速测试这些理论。员工表示，现在他们有了足够的资源，过去可能需要

数年才能实现的事情，如今一周内就能完成。

Some of these ideas for improving models are more than a year old, one of the people said. That includes recurrent depth or looping techniques in training and running advanced models.—*Stephanie Palazzolo and Amir Efrati*

其中一些改进模型的想法已有一年多历史，其中一位人士表示。这包括在训练和运行高级模型时采用循环深度或循环技术。——**Stephanie Palazzolo 和 Amir Efrati**

## Big Number

**OpenAI** has told some investors that it expects to burn $278 billion by the end of 2030, the Financial Times reported, as it spends more on cloud computing and chips to run and train its AI. The forecast, based on a recent presentation, is higher than the one it gave investors early this year, when it predicted $180 billion in negative cash flow over the same period.

据《金融时报》报道，**OpenAI** 已告知部分投资者，预计到 **2030** 年底将消耗 **2780** 亿美元资金，因为公司需要投入更多资金购买云计算资源和芯片，以运行和训练其人工智能模型。这一基于近期演示文稿的预测，高于公司今年早些时候向投资者提供的预测；当时公司预计同期现金流

出为 1800 亿美元。

Overheard