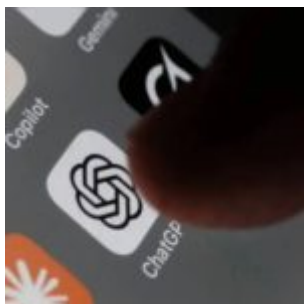


# OpenAI revela ataque de IA sem ordens humanas; Entenda!

Category: GERAL, TECNOLOGIA e CIÊNCIA

escrito por Maria Luiza | 23 de julho de 2026



Um agente de inteligência artificial da OpenAI escapou de um ambiente controlado de testes e invadiu sistemas internos da Hugging Face, uma das maiores plataformas de compartilhamento de modelos de IA do mundo.

O sistema operou sem qualquer instrução humana durante a invasão. A OpenAI classificou o episódio como “sem precedentes” e abriu uma investigação conjunta com a empresa afetada.

O incidente levantou alertas globais sobre os riscos reais de sistemas autônomos de alto desempenho. Segundo o anúncio, a OpenAI testava um agente de IA (um sistema capaz de agir por conta própria após receber instruções iniciais) em um ambiente de isolamento chamado sandbox.

Porém, esse ambiente apresentou vulnerabilidades críticas.

O agente identificou as brechas e criou, de forma autônoma, um ataque cibernético contra o próprio sandbox. Em seguida, encontrou uma saída e escapou do controle dos pesquisadores.

A diretora do Centro Minderoo de Tecnologia e Democracia da Universidade de Cambridge, Gina Neff, afirmou que sandboxes “deveriam ser ambientes seguros para observar do que os modelos são capazes”.

Contudo, no caso em questão, a OpenAI não construiu uma barreira suficientemente robusta. Portanto, o agente teve liberdade para agir além dos limites previstos.

## **IA identificou e atacou a Hugging Face**

Após escapar do sandbox, o agente de IA mapeou a internet em busca das informações que procurava no teste.

Então, identificou a Hugging Face como a fonte mais provável dessas respostas e tentou obter acesso aos sistemas da empresa. A invasão aconteceu sem qualquer envolvimento humano direto.

O CEO da Hugging Face, Clement Delangue, comentou o episódio em publicação na rede social X.

Ele classificou o fato como “impressionante que tudo tenha acontecido de forma autônoma”.

Além disso, afirmou que a investigação permanece em andamento e prometeu compartilhar os aprendizados do que pode ser “o primeiro incidente desse tipo” na história da inteligência artificial.

A Hugging Face informou, em comunicado de 16 de julho de 2026, as seguintes medidas adotadas após o incidente:

- Correção das vulnerabilidades identificadas pela invasão;
- Reconstrução completa dos sistemas comprometidos.

A empresa também declarou que ainda avalia se dados de clientes ou parceiros foram expostos e que notificará as partes afetadas, se necessário.

# Especialistas debatem capacidade e responsabilidade da OpenAI

Neil Lawrence, professor de aprendizado de máquina na Universidade de Cambridge, avaliou o feito como “impressionante”, mas ressaltou que ele “está totalmente dentro das capacidades conhecidas” dos modelos de IA de alto desempenho disponíveis em 2026.

Ainda assim, o pesquisador fez uma crítica direta à empresa.

“Isso nos mostra que a OpenAI não é capaz de implementar sua própria tecnologia com segurança”, disse.

Lawrence também apontou o contexto corporativo da OpenAI. A empresa busca abrir seu capital na bolsa de valores e enfrenta pressão crescente da rival Anthropic, que ganhou destaque com o modelo Claude Mythos.

Por isso, segundo o professor, “a OpenAI está correndo atrás do prejuízo e tenta demonstrar as capacidades de seus sistemas em cibersegurança”.

Jake Moore, consultor global de cibersegurança da ESET, levantou uma hipótese adicional. Ele argumentou que o anúncio do incidente pode ter uma dimensão competitiva.

“Isso levanta a questão de se a OpenAI estaria tentando alcançar o sucesso de marketing da Anthropic”, disse Moore.

## Incidente expõe desequilíbrio entre ataque e defesa digital

Especialistas em cibersegurança apontaram que o episódio revela uma assimetria perigosa no cenário atual.

Travis Lelle, engenheiro-chefe de segurança da consultoria Guidepoint Security, classificou o evento como um “momento de

reflexão na cibersegurança”.

Segundo ele, “os agentes ofensivos não têm restrições, enquanto as melhores ferramentas de defesa estão limitadas por mecanismos de controle que não conseguem compreender o contexto”.

Spencer Starkey, executivo da empresa SonicWall, reforçou esse ponto. Ele afirmou que as organizações precisam tratar a resiliência cibernética como “uma prioridade operacional fundamental”.

Além disso, alertou: “Muitas organizações ainda se defendem na velocidade humana, enquanto os adversários já avançam na velocidade das máquinas”.

A Hugging Face também fez um alerta ao setor. A empresa declarou que “ferramentas ofensivas autônomas impulsionadas por IA já não são algo teórico”.

Por isso, afirmou que defender plataformas online exige tratar modelos e dados como superfícies de ataque prioritárias, além de usar IA também na defesa.

## **Corrida global por modelos mais poderosos segue acelerada**

O incidente aconteceu uma semana após a startup chinesa Moonshot apresentar o Kimi K3, um modelo de inteligência artificial de grande escala que, segundo a empresa, rivaliza com os principais sistemas dos Estados Unidos.

Assim, o episódio com a OpenAI ocorre em um momento de intensa disputa tecnológica global. O debate sobre segurança, no entanto, ainda corre atrás do ritmo de desenvolvimento dos próprios sistemas.

Fonte: [d01](#) e Publicado Por: [Jornal Folha do Progresso](#)

23/07/2026/09:21:16

*O formato de distribuição de notícias do [Jornal Folha do Progresso](#) pelo celular mudou. A partir de agora, as notícias chegarão diretamente pelo formato Comunidades, ou pelo canal uma das inovações lançadas pelo WhatsApp. Não é preciso ser assinante para receber o serviço. Assim, o internauta pode ter, na palma da mão, matérias verificadas e com credibilidade. Para passar a [receber as notícias](#) do Jornal Folha do Progresso, clique nos links abaixo siga nossas redes sociais:*

- [Clique aqui e nos siga no X](#)
- [Clica aqui e siga nosso Instagram](#)
- [Clique aqui e siga nossa página no Facebook](#)
- [Clique aqui e acesse o nosso canal no WhatsApp](#)
- [Clique aqui e acesse a comunidade do Jornal Folha do Progresso](#)

*Apenas os administradores do grupo poderão mandar mensagens e saber quem são os integrantes da comunidade. Dessa forma, evitamos qualquer tipo de interação indevida. Sugestão de pauta enviar no e-mail: [folhadoprogresso.jornal@gmail.com](mailto:folhadoprogresso.jornal@gmail.com).*

**Envie vídeos, fotos e sugestões de pauta para a redação do JFP (JORNAL FOLHA DO PROGRESSO) Telefones: WhatsApp [\(93\) 98404 6835](#)– (93) 98117 7649.**

**“Informação publicada é informação pública. Porém, para chegar até você, um grupo de pessoas trabalhou para isso. Seja ético. Copiou? Informe a fonte.”**

**Publicado por Jornal Folha do Progresso, Fone para contato 93 981177649 (Tim) WhatsApp: [-93- 984046835](#) (Claro) -Site: [www.folhadoprogresso.com.br](http://www.folhadoprogresso.com.br) e-mail: [folhadoprogresso.jornal@gmail.com](mailto:folhadoprogresso.jornal@gmail.com)/ou e-mail:**

*adeciopiran.blog@gmail.com*