
Incitamento ao suicídio e OpenAI. Novas medidas de segurança introduzidas

Data: 2025-09-17 17:51:24

Autor: Inteligência Against Invaders

[Redazione RHC](#):17 Setembro 2025 19:06

OpenAI [anunciou](#) novas medidas de segurança para o ChatGPT seguinte uma série de histórias trágicas e ações judiciais acusando o chatbot de envolvimento em suicídios de adolescentes. O sistema agora tentará determinar a idade da pessoa com quem está conversando e, se necessário, solicitar o ID para confirmar que o usuário tem mais de 18 anos. A empresa reconheceu que isso limita a privacidade dos adultos, mas considerou a troca justificada por razões de segurança.

O CEO da OpenAI, Sam Altman, disse que não esperava aprovação unânime para essas medidas, mas as considerava necessárias em meio ao crescente conflito sobre a regulamentação da inteligência artificial. Essa decisão foi influenciada por uma série de incidentes de alto perfil.

Em agosto, os pais de um [adolescente chamado Adam Reid](#) entrou com uma ação judicial, alegando que o ChatGPT o ajudou a escrever uma nota de suicídio, aconselhou-o sobre métodos alternativos e o desencorajou a compartilhar seus sentimentos com adultos. Naquele mesmo mês, o Wall Street Journal relatou um **Homem de 56 anos que cometeu suicídio depois de se comunicar com um bot que alimentou sua paranóia**. E o Washington Post relatou um novo processo no qual os pais culpam o serviço The Character AI relatou a morte de uma menina de 13 anos.

A OpenAI implementou anteriormente o controle dos pais no ChatGPT, mas agora reforçou as regras. Para menores, o chatbot funcionará de acordo com princípios diferentes: rejeitará flertes e excluirá discussões sobre suicídio e automutilação, mesmo em contextos artísticos. *Se o sistema detectar pensamentos perigosos em um adolescente, ele tentará entrar em contato com seus pais e, se isso for impossível e com risco de vida, entrará em contato com os serviços de emergência.*

A OpenAI reconheceu que enfrentava um problema fundamental com grandes modelos de linguagem. Em seus estágios iniciais, o ChatGPT era rigoroso, rejeitando muitos tópicos, mas a crescente concorrência de soluções “sem censura” e locais, bem como a pressão dos críticos da censura, forçaram a empresa a afrouxar seus filtros.

Agora, a empresa mudou de rumo: **Ele quer oferecer aos usuários adultos o máximo de liberdade sem causar danos ou restringir os direitos de terceiros**. Outras plataformas estão tomando medidas semelhantes: neste verão, o YouTube anunciou planos para usar algoritmos de aprendizado de máquina para estimar a idade dos espectadores e proteger os adolescentes de certas categorias de conteúdo.

Redação

A equipe editorial da Red Hot Cyber é composta por um grupo de indivíduos e fontes anônimas que colaboram ativamente para fornecer informações e notícias antecipadas sobre segurança cibernética e computação em geral.

[Lista degli articoli](#)