

Especialistas discutem se é possível processar ChatGPT por difamação

17/04/2023

Professores de Direito de várias universidades dos EUA estão empenhados no debate do momento: a quem responsabilizar civilmente, quando um *chatbot* de inteligência artificial, como o ChatGPT, produz um texto sobre uma pessoa com informações convincentes, atribuídas a fontes confiáveis, mas inteiramente falsas — e difamatórias.



Freepik Chatbots podem distorcer fatos com muita competência e até criar fontes de informações falsas

O professor de Direito Eugene Volokh, da Universidade da Califórnia, mostrou como isso acontece. Ele pediu ao ChatGPT informações sobre professores de Direito acusados de assédio sexual. Um de seus conhecidos estava na lista.

O ChatGPT informou que o professor Jonathan Turley, do Georgetown University Law Center, foi acusado por uma aluna de assédio sexual, durante uma viagem da classe ao Alasca, patrocinada pela universidade. O texto cita uma notícia de março de 2018 no Washington Post, como fonte da informação.

Problemas: o professor Turley é de outra faculdade de Direito, a da George Washington University; ele nunca levou seus alunos a uma viagem ao Alasca ou a qualquer outro lugar; nunca foi acusado de assédio sexual; a suposta notícia no Washington Post não existe, como confirmado pelo próprio jornal.

Os professores reagem a esses problemas com preocupação, porque esses chatbots (como o ChatGPT, o Bing da Microsoft e o Bard da Google) podem distorcer fatos relevantes com muita competência e até mesmo fabricar fontes confiáveis para sustentar suas informações.

Às vezes, reagem com descaso. A professora de Direito Kate Crawford, da University of Southern California e pesquisadora da Microsoft Research, disse ao Washington Post que chama essas citações de fontes de "alucitações", um trocadilho com a palavra "alucinações". O



professor da Princeton University, Arvind Narayanan, chama o ChatGPT de "gerador de lorotas".

Em um artigo para o Wall Street Journal, o cartunista Ted Rall escreveu que considerava a possibilidade de processar o ChatGPT, que produziu um texto em que afirma que ele foi acusado de plágio de outro cartunista, com o qual ele mantinha um relacionamento "complicado" e "contencioso".

Na verdade, os dois cartunistas são amigos, o relacionamento dos dois nunca foi contencioso ou complicado e ninguém nunca o acusou de plágio. Ele perguntou a especialistas se poderia mover uma ação contra o ChatGPT por difamação.

O professor emérito da Faculdade de Direito de Harvard, Laurence Tribe, disse a ele que não importa, para propósitos de responsabilização, se as mentiras são geradas por um ser humano ou por um chatbot.

O professor RonNell Jones, da Faculdade de Direito da University of Utah, disse ao cartunista que, nesse caso, mover uma ação por difamação pode ser difícil para uma figura pública, que tem de provar que houve "[malícia real](#)" (ou intencional) da publicação, para se obter uma indenização.

Ele lembrou que alguns juristas têm sugerido que o melhor remédio é mover uma ação baseada no "modelo de responsabilização do produto", em vez de no "modelo de difamação".

Na Austrália, o prefeito de Hepburn Shire, Brian Hood, disse a Reuters que irá mover uma ação contra a [OpenAI, criadora do ChatGPT](#), a menos que a empresa corrija a falsa informação de que ele foi condenado e preso por corrupção, quando trabalhava em uma subsidiária do banco nacional do país.

Na verdade, Hood foi delator de um caso de corrupção e nunca foi acusado de qualquer crime. Seus advogados enviaram uma notificação à OpenAI, o que é o primeiro passo formal para se mover uma ação por difamação na Austrália. Essa pode ser a primeira vez que a OpenAI é processada por difamação, por causa de conteúdo criado pelo ChatGPT.

O jornal Washington Post também considera a possibilidade de processar a OpenAI/ChatGPT, por ter sido citado como fonte de informações falsas. Mas ainda há uma questão não resolvida: se a OpenAI pode evitar responsabilização com base na [Seção 230 da Lei da Decências nas Comunicações](#), que protege as publicações online contra ações por conteúdo postado por terceiros.

Outra questão é se o autor da ação pode provar que houve dano a sua reputação por causa de informações falsas. Na verdade, é muito fácil fazer *chatbots* produzirem desinformação ou discurso de ódio, se é isso que uma pessoa está buscando, diz o Jornal da ABA (American Bar Association).

Um estudo do Center for Countering Digital Hate mostrou que pesquisadores induziram o Bard (da Google) a produzir informações contaminadas por ódio em 78 das 100 vezes tentadas, sobre tópicos como o holocausto e mudança climática.



Um porta-voz da OpenAI enviou ao Washington Post a seguinte declaração: "Quando usuários se registram no ChatGPT, somos tão transparentes quanto possível ao advertir que o software nem sempre produz informações corretas. Nosso foco é melhorar a precisão dos fatos e estamos fazendo progressos."

A OpenAI descreve o ChatGPT como uma fonte confiável de asserções de fatos, não apenas uma fonte de entretenimento. No entanto, "o atual e futuro modelo de negócios da empresa depende, inteiramente, da credibilidade do ChatGPT e de sua capacidade de produzir sumários razoavelmente precisos dos fatos", escreveu o professor Eugene Volokh, que conduz um estudo sobre os chatbots.

Fonte: <https://conjur.jumps.com.br/2023-abr-17/especialistas-discutem-possivel-processar-chatgpt-difamacao/>