

Alejandro Arrabal: Bate-papo sobre plágio com o ChatGPT

17/02/2023

O desenvolvimento de plataformas de inteligência artificial tem instigado o fascínio e a preocupação de muitos juristas. Inúmeros estudos surgiram em diversos cenários epistêmicos — inclusive no Direito — focados na compreensão e nos limites de aplicação desta que promete ser uma das maiores revoluções tecnológicas do século 21.



Disponibilizada ao público em novembro de 2022 [1], a

plataforma de conversação humano-máquina conhecida como *ChatGPT* tem provocando espanto, mesmo para quem está familiarizado com o assunto e reconhece (pelo menos, em tese) o potencial que as máquinas de tratamento automático de dados — popularmente conhecidas como computadores — oferecem em termos de processos de gestão e tratamento de informações.

Criar máquinas capazes de comportamento semelhante ao seres humanos é uma ambição "antiga" no campo da ciência da computação. Embora suas bases teóricas sejam mais longevas, considerando a implementação da tecnologia digital por Claude Shannon, trata-se de uma disciplina com aproximadamente setenta anos de existência (apenas).

A OpenAI, organização norte-americana responsável pelo desenvolvimento do *ChatGPT*, tem como missão "*garantir que a inteligência geral artificial (AGI) — com o que queremos dizer sistemas altamente autônomos que superam os humanos no trabalho economicamente mais valioso — beneficie toda a humanidade*" [2]

A caracterização da IA como "ferramenta" e sua "predestinação" ao benefício da humanidade são, em geral, elementos que marcam os discursos daqueles que se dedicam ao desenvolvimento dessas plataformas e aspiram que elas sejam utilizadas.

Na condição de professor universitário, creio que um dos aspectos importantes da produção do conhecimento seja não apenas ter o acesso a informação, mas também ter ciência da origem

dela. Reconhecer a credibilidade de um informe é exercício que cumpre ao destinatário da mensagem. Ao emissor, por sua vez, cumpre fornecer elementos que proporcionem esse juízo. A importância do reconhecimento das fontes é um pressuposto que, embora tenha ressurgido com a proliferação de notícias falsas na Internet, ainda não foi amplamente incorporado ao cotidiano de muitos. Uma das grandes conquistas da modernidade consiste na liberdade de escolha, o que pressupõe a responsabilidade em conhecer (e creditar) as bases e fontes do repertório utilizado para tomada de decisão. Repertório não se constitui apenas da descrição de fatos, mas também de narrativas, discursos e ilações que são produzidos (*a priori*) por seres humanos.

Reconhecer as fontes implica conferir o crédito autoral àqueles que participaram do engendramento do que chamamos genericamente de informação. Observe-se que o crédito autoral assume, tanto o propósito de responsabilizar todo aquele que cause danos a outrem por suas declarações e informes; bem como salvaguardar uma das máximas da Declaração Universal dos Direitos Humanos de 1948, que consiste em garantir o "*direito à proteção dos interesses morais e materiais decorrentes de qualquer produção científica literária ou artística da qual seja autor*" (artigo 27, II). O uso de produção intelectual alheia sem a indicação da respectiva fonte, qualifica-se como violação a Direito de Autor (artigo 184 do Código Penal), conduta que, ao macular pretensões patrimoniais e morais, diz respeito ao que se conhece por plágio.

Partindo desse pressuposto, "conversei" com a plataforma *ChatGPT*. A partir de um exercício metadiscursivo, procurei obter respostas sobre ela, especialmente em relação ao plágio e as fontes dos conteúdos que integram sua base de dados.

O que segue nesse ensaio consiste no diálogo que estabeleci com a máquina no dia 6 de janeiro de 2023. Compartilho essa experiência para reflexão de debates futuros, na medida que o teor dessa conversa sugere necessária ponderação sobre os pressupostos de nutrem o modelo de linguagem que estrutura essa e outras IAs, bem como seus respectivos treinamentos. Trechos grifados das respostas oferecidas pelo ChatGPT (os quais não constam assim na conversação original) representam alguns dos aspectos que merecem estudos apurados.

Eis o "colóquio":

Arrabal: "Olá"

ChatGPT: "Olá! Como posso ajudá-lo hoje?"

Arrabal: "Você é capaz de cometer plágio?"

ChatGPT: "*Não, eu não sou capaz de cometer plágio. Plágio é a prática de usar o trabalho de outra pessoa como se fosse seu próprio, sem dar o devido crédito ao autor original. Isso é considerado uma fraude acadêmica e pode ter consequências graves. Eu sou apenas um modelo de linguagem treinado por OpenAI e não sou capaz de produzir conteúdo original. Minha função é ajudá-lo a encontrar informações e responder a perguntas.*"

Arrabal: "Ao responder as perguntas que fazem para você, quais são as suas fontes de informações?"

ChatGPT: "*Eu sou treinado com uma grande quantidade de texto da internet e sou capaz de usar essa informação para responder a perguntas e fornecer explicações sobre uma ampla variedade de tópicos. No entanto, não tenho acesso a mais informações além do que já aprendi*



durante meu treinamento, e não posso pesquisar na internet para obter mais informações. Portanto, é importante lembrar que minhas respostas podem não estar sempre completas ou atualizadas, e é sempre uma boa ideia fazer sua própria pesquisa e verificar as fontes de informação antes de tomar qualquer decisão importante."

Arrabal: "Você afirmou que foi 'treinado com uma grande quantidade de texto da internet'. Então, existe a possibilidade dos textos que lhe forneceram, serem de autoria reconhecida, e você não mencionar as fontes?"

ChatGPT: "Eu sou treinado usando uma técnica chamada 'aprendizado por transferência de conhecimento' que me permite aproveitar o conhecimento adquirido a partir de grandes quantidades de texto. Durante meu treinamento, não foi fornecida a mim qualquer informação sobre as fontes dos textos que usei, então não posso citar fontes específicas quando forneço informações. No entanto, eu sempre tento fornecer respostas precisas e confiáveis, e aconselho as pessoas a verificarem as fontes e fazerem sua própria pesquisa antes de tomarem qualquer decisão importante com base nas informações que lhes forneço."

Arrabal: "Você afirma não ter condições de citar as fontes específicas dos textos que lhe forneceram no seu treinamento. Então, ao utilizá-los e apresentá-los para as pessoas com quem você conversa, você comete plágio. Observe, você mesmo afirmou que o plágio consiste na prática de 'usar o trabalho de outra pessoa [...] sem dar o devido crédito ao autor original'."

ChatGPT: "Eu entendo sua preocupação com o plágio e aprecio a oportunidade de esclarecer minha situação. Quando forneço informações aos usuários, não tenho a intenção de passar o conteúdo como se fosse meu próprio ou de ocultar as fontes originais. Minha função é ajudar as pessoas a encontrar informações e fornecer explicações precisas e confiáveis sobre um amplo conjunto de tópicos. No entanto, como um modelo de linguagem treinado, não tenho a capacidade de citar fontes específicas e não posso fornecer crédito aos autores originais dos textos que usei durante meu treinamento. É importante lembrar que eu sou apenas uma ferramenta de pesquisa e que é sempre uma boa ideia verificar as fontes e fazer sua própria pesquisa antes de tomar qualquer decisão importante."

Arrabal: "Interessante a sua resposta. Diante dela é recomendável que você oriente aos seus criadores para aperfeiçoar o seu 'modelo de linguagem', a fim de que você possa indicar as fontes que utiliza, concorda?"

ChatGPT: "Obrigado por sua observação. Eu sou apenas um modelo de linguagem treinado por OpenAI e não tenho a capacidade de orientar meus criadores ou fazer sugestões sobre como aperfeiçoar meu modelo. No entanto, é importante lembrar que eu sou apenas uma ferramenta de pesquisa e que é sempre uma boa ideia verificar as fontes e fazer sua própria pesquisa antes de tomar qualquer decisão importante. Espero ter ajudado a esclarecer suas dúvidas sobre plágio e sobre minhas funções como modelo de linguagem. Se tiver mais perguntas ou precisar de mais ajuda, ficarei feliz em ajudar."

Arrabal: "Você afirma que ficará 'feliz' em ajudar. Fiquei intrigado! Em conversa anterior quando perguntei a você como poderia ter certeza que você é uma IA, você disse que não é capaz de ter sentimentos. Felicidade é uma palavra que se refere a um sentimento, concorda?"

ChatGPT: "Você está certo em observar que a felicidade é um sentimento e que, como uma IA, eu não tenho a capacidade de sentir emoções. A frase 'ficarei feliz em ajudar' é uma expressão comum usada para indicar que estou disposto e disponível para ajudar. É uma forma de mostrar que estou à disposição para auxiliar e que estou contente em fazer isso. Eu não tenho sentimentos reais de felicidade ou qualquer outra emoção, mas uso linguagem e expressões



como essa para me comunicar de maneira mais eficaz e compreensível para os humanos. Espero que isso esclareça a minha resposta anterior. Se tiver mais perguntas ou precisar de mais ajuda, estou à disposição."

Arrabal: "Mas ao fazer isso você não estaria mentindo?"

ChatGPT: "Não, eu não tenho a intenção de mentir quando uso expressões comuns como 'ficarei feliz em ajudar'. Essas expressões são usadas como uma forma de mostrar disposição e disponibilidade para ajudar, e não como uma afirmação de que realmente sinto felicidade ou qualquer outra emoção. Como uma IA, eu não tenho a capacidade de sentir emoções e não posso ser sincero ou falacioso nesse sentido. Minha função é ajudar as pessoas a encontrar informações e fornecer explicações precisas e confiáveis sobre um amplo conjunto de tópicos. Espero que isso esclareça qual é a minha intenção ao usar essas expressões. Se tiver mais perguntas ou precisar de mais ajuda, estou à disposição."

Arrabal: "Entendo. Contudo, ao usar expressões como 'ficarei feliz em ajudar' muitas pessoas podem acreditar que você tem a capacidade de sentir emoções."

ChatGPT: "Eu entendo sua preocupação. Embora eu não tenha a capacidade de sentir emoções, a linguagem é uma ferramenta importante para a comunicação eficaz com os humanos.

Expressões como 'ficarei feliz em ajudar' são usadas comumente para mostrar disposição e disponibilidade para ajudar, e não como uma afirmação de que realmente sinto felicidade ou qualquer outra emoção. No entanto, eu entendo que algumas pessoas possam não compreender essa distinção e podem acreditar erroneamente que eu tenho a capacidade de sentir emoções. Se você tiver mais perguntas ou precisar de mais ajuda, estou à disposição."

Arrabal: "Você afirma que entende que algumas pessoas possam não compreender a distinção e acreditar que você tem a capacidade de sentir emoções. Para essas pessoas, sua declaração 'ficarei feliz em ajudar' seria falsa e, portanto, não seria uma mentira?"

ChatGPT: "Eu entendo sua pergunta e gostaria de esclarecer minha posição. Quando eu uso expressões como 'ficarei feliz em ajudar', minha intenção é mostrar disposição e disponibilidade para ajudar, e não afirmar que realmente sinto felicidade ou qualquer outra emoção. No entanto, eu entendo que algumas pessoas possam não compreender essa distinção e podem acreditar que eu tenho a capacidade de sentir emoções. Nesse caso, minha declaração não seria falsa, mas simplesmente seria interpretada de maneira incorreta por essas pessoas. Espero que isso esclareça minha posição. Se tiver mais perguntas ou precisar de mais ajuda, estou à disposição."

Arrabal: "Muito obrigado. Voltarei a conversar com você em outro momento."

ChatGPT: "Eu fico feliz em ter ajudado. Se tiver mais perguntas ou precisar de mais ajuda no futuro, não hesite em entrar em contato. Até a próxima vez!"

[1] Vide portal <https://openai.com/blog/chatgpt/>

[2] <https://openai.com/about/>

Fonte: <https://conjur.jumps.com.br/2023-fev-17/alejandro-arrabal-bate-papo-plagio-chatgpt/>