CEST
Centro de Estudos Sociedade e Tecnologia
USP
Universidade de São Paulo
Boletim- Volume 8, Número 1, Fevereiro/2023

ChatGPT: se você ainda não usou, um dia pode usar sem perceber

Eduardo Bertassi

O ChatGPT ou *Chat Generative Pre-Trained Transformer* é um *chatbot* lançado pela OpenAI, uma empresa de pesquisa, inicialmente, sem fins lucrativos fundada em 2015 por Elon Musk (fundador do PayPal, CEO da Tesla, SpaceX e Twitter), Sam Altman (atual CEO da Open AI), Peter Tiel (cofundador do PayPal e Palantir Technologies), Reid Hoffman (cofundador do LinkedIn) entre outros magnatas da tecnologia.

Por volta do começo de dezembro de 2022 o ChatGPT começou a ganhar espaço nos jornais brasileiros, mas seus feitos se tornaram mais proeminentes quando foi lançado publicamente como um protótipo em 30 de novembro de 2022; o funcionamento era (e até o momento ainda é) simples: bastava digitar uma pergunta em uma caixa de texto na interface do sistema que o *chatbot* lhe forneceria a resposta. No entanto, diferentemente de outros chats mais rudimentares, era possível criar poemas, canções, roteiros de TV e até depurar linhas de código-fonte de programas de computador a partir das perguntas feitas pelos usuários.

As respostas do *chatbot* impressionam, mas nem sempre estão corretas. Como qualquer sistema de Inteligência Artificial (IA) o ChatGPT (ainda) não é perfeito, pois tais sistemas aprendem por meio de regularidades estatísticas a partir dos seus dados de treinamento. Conforme explicação da própria empresa em seu site, o sistema utiliza um modelo chamado de RLHF ou *Reinforcement Learning from Human Feedback* (aprendizado por reforço a partir de retroalimentação humana) em que se treina uma base de dados com a supervisão de humanos. Explicando de maneira simplificada, a partir de centenas de conversas com o *chatbot* os treinadores humanos classificam as respostas com os melhores resultados usando um modelo de reforço por recompensa e isso ajusta as respostas para perguntas similares que poderão ser feitas no futuro.

Os resultados que se podem obter são impressionantes e não é à toa que chamam a atenção, pois em suas respostas o ChatGPT pode combinar diferentes campos de conhecimento de formas completamente inusitadas e até mesmo divertidas. É possível pedir que ele depure o código-fonte de um *software*, mas respondendo como se fosse um pirata ou então pedir que ele explique como funciona um famoso algoritmo de classificação de dados chamado *Bubble Sort* como se fosse um gangster esperto.

Em suas respostas o ChatGPT pode combinar diferentes campos de conhecimento de formas completamente inusitadas e até mesmo divertidas

A própria empresa reconhece que o seu sistema possui limitações:

- algumas respostas, apesar de serem plausíveis podem ser incorretas ou sem sentido (é preciso conferir);
- algumas perguntas sem respostas, se reformuladas com pequenos ajustes, podem finalmente ter uma resposta válida (recomenda-se experimentar diferentes formulações de perguntas);
- certas respostas extremamente detalhadas acabam surgido de vieses colocados nos dados de treinamento por instrutores que preferem respostas mais longas; e
- certas perguntas inapropriadas podem ter respostas com instruções prejudiciais ou apresentar comportamentos tendenciosos (a empresa está se esforçando par evitar tais respostas por meio de regras de moderação e até mesmo bloqueios intencionais).

O que se tem discutido sobre o ChatGPT são as questões éticas, filosóficas, econômicas e até mesmo psicológicas relacionadas ao uso deste tipo de tecnologia, mas isso é algo que sempre acontece quando uma nova tecnologia surge, afinal, nunca se sabe de início como algo completamente novo e algumas vezes disruptivo pode impactar de forma significativa a sociedade e os indivíduos porque isso é algo muito difícil de se prever durante a fase de desenvolvimento de um projeto. Algumas vezes, novas tecnologias podem trazer dilemas sem resposta devido às limitações dos modelos de ética correntes e acabam precisando até mesmo da criação de novos princípios éticos para guiar seu uso.

Vale ressaltar que Elon Musk, um dos fundadores da OpenAI, renunciou ao conselho de administração da empresa no final de 2018 porque deixou de concordar com o rumo o qual a empresa estava tomando. A empresa que havia surgido como um empreendimento sem fins lucrativos "para avançar a inteligência digital da maneira mais provável para beneficiar a humanidade como um todo, sem restrições da necessidade de gerar retorno financeiro" (como descrito no site da empresa em 2015) parece que acabou tomando outros rumos. Musk afirmou que "OpenAI foi iniciado como código aberto e sem fins lucrativos; nenhum dos dois ainda é verdade". Mas por que essa preocupação de manter o ChatGPT como um projeto de código aberto?

Softwares ou sistemas de código aberto, diferentemente daqueles que são de código fechado, têm a vantagem de terem seu código-fonte abertos ao público, portanto ele pode ser disponibilizado gratuitamente, pode ser copiado e até mesmo alterado por seus usuários de acordo com suas necessidades (o *software* de código aberto mais famoso até hoje é o sistema operacional Linux). Há inúmeras vantagens de se utilizar *softwares* de código aberto: são grátis; não é preciso de licenças; o seu desenvolvimento não é limitado a indivíduos, grupos ou organizações específicas; não estão sujeitos a regulações de órgãos ou empresas específicas; entre outros. Uma das características mais importantes é a de que todos sabem como o *software* funciona incluindo suas vulnerabilidades, portanto a comunidade de desenvolvedores está sempre trabalhando para corrigi-las; apesar se ser contraintuitivo, ter expostas as vulnerabilidades de um *software* é algo que ajuda a aumentar sua qualidade e confiabilidade, pois sabe-se exatamente o que deve ser feito para melhorar o *software*.

Para sistemas de IA tão complexos quanto o ChatGPT ter a capacidade de se inspecionar como o *software* foi codificado por seus programadores é fundamental. Muito provavelmente, num futuro não tão distante, *softwares* embarcados com IA passarão a tomar uma quantidade cada vez maior de decisões pelos seres humanos devido à vasta quantidade de informações e variáveis existentes a partir de diferentes e inusitados cenários de aplicação, incluindo defesa militar. Porém, quem seria tão imprudente de adquirir um sistema de IA de código fechado (que não se sabe como foram codificados) para controlar mísseis nucleares, por exemplo? Esse é um caso extremo, mas que chama a atenção para a necessidade de se criar políticas de regulamentação visando a utilização desse tipo de *software*, como proposto pelo Parlamento Europeu em abril de 2021 no documento que visava estabelecer "regras harmonizadas" para o uso da IA, ou seja, uma lei para utilização de IA.

Sem dúvida, a utilização da IA é algo que, além de nos maravilhar, fará parte do nosso dia a dia de formas que sequer conseguimos imaginar. No entanto, precisamos ficar alertas, pois seu uso indiscriminado e sem estudos (seja do ponto de vista técnico, ético ou social) podem gerar mais problemas do que soluções. Não devemos deixar de ser otimistas com a possibilidade de criar uma ferramenta que ajudará a humanidade, mas também não devemos deixar de permanecer vigilantes para que essa inovação permaneça sempre nos trilhos que esperamos que ela permaneça.

Eduardo Bertassi *é graduado em engenharia elétrica pela Escola Politécnica da USP e colaborador do CEST-USP.*

Coordenador Acadêmico: Edison Spina

Este artigo resulta do trabalho de apuração e análise das autoras, não refletindo obrigatoriamente a opinião do CEST.