

Moreira Rêis: Primeiras impressões da febre do ChatGPT-3

Nos últimos dias, virou febre na internet conversar com o robô *ChatGPT-3* [1]. Trata-se de um poderoso modelo de inteligência artificial para o processamento de linguagem natural (PLN), pertencente à empresa OpenAI, que consegue manter uma conversa por escrito com um ser humano sobre



A coisa ainda está num estágio experimental, mas já aberta

para uso por qualquer um que se interesse. Basta acessar [este site](#) e seguir as instruções.

Já há notícias de que o ChatGPT conseguiu ser aprovado nos testes de proficiência para ser médico, advogado e concluir um MBA nos Estados Unidos [2]. Uma mulher britânica de 37 anos, residente em Londres, decidiu deixar o marido e assumir o amante que conheceu online, depois de ser aconselhada pelo *ChatGPT* [3]. No Brasil, o robô escreveu, em 50 segundos, uma redação sobre o tema do Enem 2022, com um nível bem razoável de expressividade [4]. Um homem publicou um livro infantil escrito e ilustrado inteiramente pelo GPT-3 [5], em poucos dias, suscitando a revolta de alguns artistas com o uso da criatividade humana por máquinas.

É possível pedir ao *ChatGPT-3* para traduzir e resumir um livro em outro idioma, escrever poesias ou músicas, gerar diálogos entre personagens de um livro ou filme, com base nos perfis oferecidos pelo usuário etc. Enfim, o *ChatGPT-3* pode funcionar como um assistente pessoal superinteligente; um oráculo capaz de responder virtualmente a qualquer pergunta que se faça sobre qualquer tema, em questão de segundos e com satisfatória redação. Lembra um pouco aqueles assistentes pessoais de super-heróis dos quadrinhos e do cinema, como o J.A.R.V.I.S., do Homem de Ferro. Com o detalhe de que, por enquanto, o robô GPT-3 se comunica apenas por escrito com o usuário.

A grande novidade desse *chatbot* é que ele foi treinado numa colossal base de dados e dispõe de nada menos que 175 bilhões de parâmetros para decidir sobre o texto que vai escrever como resposta a alguma demanda que lhe é apresentada. Para que se tenha uma ideia do tamanho disso, o ChatGPT2 tinha "apenas" 1,5 bilhão de parâmetros.



Como resultado dessa vasta "experiência" com dados, o ChatGPT-3 pode responder por escrito, sem nenhum ajuste especial, desde questões filosóficas, até receitas de bolo, passando por matemática, ciências, biologia, astronomia etc. O GPT-3, por isso mesmo, pode em tese ser usado como tradutor, programador, professor, escritor de ficção, jurista, conselheiro amoroso, poeta, e por aí vai.

Apesar dessa grande versatilidade, não dá para dizer ainda que o *ChatGPT-3* é uma inteligência artificial geral (AGI, na sigla em inglês para *artificial general intelligence*) — até porque a inteligência humana, que seria o parâmetro absoluto para medir o que é a inteligência em si, não atua apenas por meio de textos, mas também por meio de imagens, sons, sensações táteis e até pelo silêncio e pela inação. Mas, sem dúvida, é o mais próximo que se chegou de uma AGI, até este momento.

Os algoritmos Processamento de Linguagem Natural (PLN) de última geração eram treinados em tarefas particulares — como, por exemplo classificação de sentimentos e classificação textual —, usando aprendizado supervisionado. O *ChatGPT-3* foi pensado para lidar com qualquer assunto a respeito dos quais ele tenha sido treinado, desde que a pergunta seja feita por escrito em certos idiomas, e isso tudo sem qualquer supervisão humana específica dos seus desenvolvedores.

Os modelos GPT têm sido concebidos para aprender um modelo de linguagem usando dados não rotulados, isto é, sem significado explicitamente indicado pelo programador. Apenas para ajustar o modelo são fornecidos exemplos de tarefas específicas. Nos modelos de aprendizado supervisionado, o programador apresenta o objeto e o associa ao respectivo nome (rótulo); no GPT, os objetos e os nomes são apresentados como aparecem no dia a dia e o programa deve inferir as conexões entre eles pela massa dos exemplos que são oferecidos. Mal comparando: no aprendizado supervisionado, a máquina aprende uma língua com um professor; no não supervisionado, tem que aprender sozinha, por superexposição aos dados e inferência estatística de padrões, com correções e ajustes sendo feitos a partir dos *feedbacks* dos "nativos" do idioma.

Em todo caso, mesmo no aprendizado não supervisionado, o desenvolvimento do modelo não se faz por exposição a dados brutos. Os dados de treinamento são tratados antes de serem apresentados ao modelo de linguagem, num procedimento conhecido como pré-processamento de dados. Isso é para dizer que, mesmo num programa tão avançado e complexo como o GPT, o recorte inicial do que a máquina deve conhecer é escolhido pelo desenvolvedor. Há um efeito de enquadramento aqui e isso tem consequências posteriores sobre o *design* final do modelo.

É impossível dizer agora quais serão as consequências sociais da incorporação de "motores de conversação", como o *ChatGPT-3*, à vida cotidiana das pessoas, das empresas e dos governos. Mas certamente elas serão muitas.

Em primeiro lugar, o uso da internet deve mudar bastante. Quando olhamos para a maneira como a *web* é usada hoje pela maioria das pessoas — com exceção das redes sociais, que têm uma lógica diversa — percebemos que ela ainda é muito tributária da visão original de Tim Berners-Lee, o pai da *web*, a saber: há um grande livro (*hipertexto*), formado por muitas páginas (*sites*), que podemos "navegar" (não seria melhor "folhear"?) conforme o nosso interesse, com a ajuda de índices móveis *ad hoc*, dos quais o Google é o mais importante.



Com os motores de conversação, deverá ocorrer uma mudança importante: as fontes das informações passarão a segundo plano e o usuário focará em um resultado interpretado, ou seja, um resumo bem estruturado do tema, em vez de vários textos para ler (muitos deles irrelevantes).

Vamos a um exemplo, típico da área jurídica, para esclarecer melhor o que dizemos: imaginemos que um aluno de graduação (ou mesmo um bacharel), deseje saber o que é "hipoteca". Ele então digita a palavra "hipoteca" no Google. O motor de busca produz mais de 34 milhões de resultados em 0,45 segundos. Na prática, dificilmente alguém passa da primeira página dos milhões de resultados, o que indica haver certa desproporção dos resultados em relação à expectativa média dos usuários. Enfim, os resultados são vários *links* sobre o tema, com artigos, notícias e toda ordem de postagens que tangenciam o tópico e sobre os quais o usuário deverá ainda se debruçar para extrair a informação que deseja.

Agora, se o mesmo usuário fizer a pergunta "O que é hipoteca?" para o ChatGPT-3, a resposta será essa (ou algo aproximado, já que o modelo evolui): *"Hipoteca é um empréstimo garantido por uma propriedade imobiliária. Ela é concedida pelo banco ou outra instituição financeira e é usada para financiar a compra de uma casa ou outro imóvel. O mutuário, ou tomador do empréstimo, coloca a propriedade como garantia para o empréstimo e se compromete a pagar as parcelas do empréstimo, juntamente com juros e outros encargos, de acordo com as condições acordadas no contrato de hipoteca. Se o mutuário não pagar as parcelas do empréstimo, a instituição financeira pode tomar posse da propriedade e vendê-la para recuperar o valor do empréstimo"*.

Certo, essa não é uma resposta perfeita. Ela contém algumas imprecisões, mas é mais do que suficiente para que se tenha uma boa ideia do que é hipoteca. E o mais interessante é que vem num formato imediatamente legível e compreensível (e "colável", infelizmente), sem a necessidade de novos cliques em inúmeros outros *links* para aumentar a legibilidade da pesquisa.

Comparando esses dois formatos de resultados de pesquisa, é justo supor que o design dos motores de busca — e até o modelo de negócios de empresas responsáveis por eles — deve mudar em breve, dada a concorrência dos motores de conversação. Hoje o Google e seus homólogos têm importante fatia do seu faturamento atrelada à colocação de páginas patrocinadas no topo de seus resultados de busca. Como não haverá "resultados de busca" nos motores de conversação, senão a própria resposta já formulada em formato imediatamente legível pelo usuário, então provavelmente esses serviços serão oferecidos por meio de *streaming*, mediante assinatura mensal, visto que não será viável, em princípio, a promoção explícita de certas páginas patrocinadas.

Em contrapartida, uma preocupação importante trazida pelo uso do *ChatGPT-3* é que ficará ainda mais difícil para os produtores de conteúdo controlar o uso de suas criações. O programa não indica especificamente de onde extrai as suas respostas diretas e sucintas, embora seja claro que ele tira de conteúdos hospedados na web. Indagado sobre quais são as bases de dados que usa para produzir as suas incríveis respostas sobre tudo que há, o *ChatGPT-3* respondeu: *"Eu sou treinado com uma variedade de fontes de dados, incluindo artigos de notícias, livros eletrônicos, sites e fóruns na internet. Isso me permite responder a uma variedade de perguntas e tópicos"*. Resposta bastante elusiva, como se vê.



A questão da fraude em provas e trabalhos acadêmicos é outro ponto sensível. Os maus alunos de cursos de graduação ou pós-graduação poderão, em tese, assumir para si a autoria de textos produzidos mediante perguntas ao *ChatGPT-3*. Isso evidentemente é uma manipulação da IA para um propósito malicioso — que, no entanto, por infelicidade pode ocorrer na prática.

Para além dos eventuais controles pedagógicos das universidades, colégios e instituições de ensino em geral, estima-se que a solução para isso talvez esteja na criação de outros algoritmos, capazes de sondar o texto para saber se ele foi produzido por máquinas; ou então na incorporação no próprio *design* do *ChatGPT-3* de uma espécie de "marca d'água" digital que aponte a origem do texto [6]. Essa é uma questão que promete muitas vicissitudes ainda.

Os vieses também merecem cuidadosa atenção. Já se sabe que um dos grandes problemas dos algoritmos é que eles tendem a reproduzir visões distorcidas da realidade, o que se torna mais grave por causa da escala que isso alcança na internet. Se, em vez de dar uma lista de fontes, como nos motores de busca, o algoritmo de conversação tem de dar uma resposta resumida sobre o que foi perguntado pelo usuário, aumentam proporcionalmente as chances de enviesamento, por conta da condensação que precisa ser feita para produzir a síntese de tudo que foi encontrado, deixando de lado detalhes contextuais que eventualmente podem ser relevantes para uma boa compreensão de certos temas.

A superação dos vieses depende fundamentalmente de um conjunto de dados de treinamento apropriado e de ajustes posteriores no *design* do algoritmo, conforme sejam apresentados elementos objetivos que indiquem a efetiva ocorrência de enviesamento. Esse não é um problema exclusivo ou particularmente mais grave no *ChatGPT-3*, mas antes é típico de todos os algoritmos que lidam com grandes conjuntos de dados. É provável, inclusive, que medidas de profilaxia anti-vieses se tornem rotineiras em algoritmos que operem com o público em geral, dada a inerente tendência para o enviesamento de modelos estatísticos.

Olhando pela perspectiva do trabalho, o *ChatGPT-3* tem potencial para abalar uma série de profissões. Se, no passado, as tecnologias eliminaram postos de trabalho entre os menos letrados, agora parece que estamos diante de uma ferramenta que tem capacidade para abalar profissões intelectuais — na verdade, principalmente elas.

Com efeito, a existência de uma coisa que pode responder bem, por escrito, a dúvidas sobre tudo, é uma clara ameaça, em tese, a todas as profissões que se baseiam em conhecimentos técnicos e que se exercem predominantemente por escrito (assim, por exemplo, as profissões jurídicas em geral). É claro que temos de considerar também a contracorrente formada pelas pressões de ordem política, econômica e mesmo cultural que se farão para manter os postos de trabalho nas mãos das pessoas, mas não podemos desprezar a força dos fatos. Há empregos e profissões em perigo, sim.



Para ilustrar isso, convém lembrar que recentemente Joshua Browder, dono da empresa DoNot Pay, anunciou pelo Twitter que um robô de sua empresa daria assistência, como se advogado fosse, a uma pessoa num tribunal americano. Convenientemente, ele deixou de mencionar onde e quando se daria o julgamento e qual seria o processo, para evitar impugnações. Mas, poucos dias depois do anúncio, o conselho dos advogados locais conseguiu impedir a experiência [7], ameaçando de prisão Joshua Browder. Outro exemplo que, graças a uma reportagem da revista *Wired* de 2019 [8], sempre é lembrado quando se fala de perda de postos de trabalho para computadores, é o de robôs-juízes que estariam atuando na Estônia. Tal fato, no entanto, foi categoricamente negado pelo Ministério da Justiça daquele país [9]. Isso não quer dizer, claro, que a ideia não possa ser cogitada em algum momento no futuro, aqui ou no exterior.

Ao lado desses senões e de muitos outros que ainda se apresentarão, não podemos deixar de ver, já nesses primeiros contatos, muitas coisas boas que o *ChatGPT-3* e seus similares poderão trazer. De fato, ter um assistente virtual superinteligente que responda a todas as nossas dúvidas, a qualquer hora do dia ou da noite, pode nos poupar tempo e energia, simultaneamente aumentando a nossa produtividade. Basta pensar no próprio Google, que tem desempenhado, ao longo dos últimos anos, função semelhante à de assistente pessoal em algumas circunstâncias. Quantas tarefas do dia a dia e do trabalho foram facilitadas pelo Google? Inúmeras! O mesmo deve ocorrer quando o *ChatGPT-3* entrar em operação para valer e se incorporar aos hábitos cotidianos.

Yann LeCun, uma das maiores autoridades do mundo em inteligência artificial, fez um prognóstico auspicioso, numa recente publicação em uma rede social. Ele disse que, num mundo em que todos tiverem acesso a um assistente pessoal com inteligência artificial, o conhecimento e a inteligência se tornarão menos importantes do que a motivação, o senso moral e a habilidade para ouvir, assim como se passa com grandes líderes, que são cercados por gente que sabe mais do que eles mesmos, e nem por isso perdem a sua proeminência.

Esse é um destino possível e francamente desejável, mas receio que, para chegar lá, tenhamos ainda um longo e acidentado caminho a percorrer.

[1] O "chat" é de conversa, em inglês, e o "GPT-3" é para se referir à terceira geração de um algoritmo de *Generative Pretrained Transformer*.

[2] [ChatGPT: inteligência artificial é aprovada em provas para médico, advogado e MBA nos EUA | Exame](#)

[3] [Mulher decide se divorciar de seu marido e ficar com seu amante porque o bot de IA ChatGPT lhe disse para | Daily Mail Online](#)

[4]



[Robô 'ChatGPT' escreve redação do Enem em 50 segundos; saiba quanto ele tiraria na prova | Educação | G1 \(globo.com\)](#)

[\[5\] AI-generated kids' book using ChatGPT, Midjourney caught in art debate – *The Washington Post*](#)

[\[6\] Há possibilidade técnica de se fazer isso, conforme se vê em: \[A Watermark for Large Language Models \\(arxiv.org\\)\]\(#\)](#)

[\[7\] AI-powered "robot" lawyer won't argue in court after jail threats – *CBS News*](#)

[\[8\] Can AI Be a Fair Judge in Court? Estonia Thinks So | *WIRED*](#)

[\[9\] Estonia does not develop AI Judge | *Justiitsministeerium*](#)

Meta Fields