O trem da IA descarrilou?

Carlos A. Afonso

Data da publicação:

Julho | 2023

> "Os fatos são subversivos. Subversivos das reivindicações feitas por líderes democraticamente eleitos, bem como ditadores, por biógrafos e autobiógrafos, espiões e heróis, torturadores e pós-modernistas. Subversivos de mentiras, meias-verdades, mitos; de todos aqueles "discursos fáceis que confortam os homens cruéis".
> -- Timothy Garton Ash, *Facts are Subversive*

Recentemente, centenas de cientistas e pessoas envolvidas com a "indústria dos algoritmos" assinaram um manifesto de uma frase: "Mitigar o risco de extinção causado pela inteligência artificial (IA) deve ser uma prioridade global, juntamente com outros riscos em escala social, como pandemias e guerra nuclear." Um alerta direto ao ponto, assinado inclusive pelos criadores do estopim que desatou a crise da "inteligência artificial generativa" (IAG): o chatGPT.[1] Variantes desse alerta têm sido publicados por instituições e entidades especialistas, tendo em comum o mantra da "IA ética".

Em um movimento oposto, o Washington Post reporta que o Vale do Silício vivia um ambiente sombrio, com demissões em massa, até que foi bafejado pelo tsunami da IAG. Só no mês de maio os investimentos de risco em "start-ups" de IA somaram US$11 bilhões, um salto de 86% em relação ao mesmo mês do ano passado.[2] Essa febre, combinada com a desenfreada prospecção de criptomoedas, empurrou a principal fabricante de processadores gráficos de alta performance, a Nvidia, para o pedestal das empresas multibilionárias. Os processadores gráficos (as GPUs) têm sido utilizados para processamento rápido de volumes imensos de dados, por ter um desempenho muito superior aos processadores de uso geral – uma capacidade exigida pelos atuais sistemas de IAG.

É uma curiosa contradição entre o pavor de uma extinção causada pela IA e o desejo incontido de fazer fama e dinheiro com os avanços espetaculares e assustadores da mesma.

A IAG é uma variante de um campo da programação de sistemas conhecido como Processamento de Linguagem Natural (PLN), que inclui sistemas como geradores de textos, chatbots de atendimento, aplicativos de conversão e manipulação de mídia, emulação de sistemas biológicos etc.

É relevante entender que as origens da IA (cujo ponto de partida como objeto formal de pesquisa data de 1956)[3] estão na própria programação de computadores, em particular dos programas que interagem com um usuário humano ou com outro programa. A cada momento usuários de dispositivos conectados à Internet (ou mesmo offline) interagem com uma máquina de estado finito e jogadores online (ou offline) interagem com máquinas de estado difuso. Você visita um sítio Web e busca algo de interesse em um menu – que representa uma máquina de estado finito, com algumas opções predeterminadas, e você pode escolher apenas uma. Uma versão mais sofisticada ("fuzzy state" ou estado difuso) é encontrada por exemplo na interação em jogos ou com veículos autônomos, em que as opções são dinâmicas.

Essas máquinas de estado são as precursoras do que se convencionou chamar de inteligência artificial. Eram e são nada mais que algoritmos em software criados por humanos. Esta explicação simplista é apresentada apenas

para lembrar que os fundamentos da IA estão na própria gênese da programação de computadores.

A evolução de capacidade/velocidade de processamento e de memória, bem como o avanço dos sistemas em rede, permitiram grandes saltos na possibilidade de programas interativos cada vez mais sofisticados poderem consultar rapidamente grandes bases de dados distribuidas em um ou mais datacentros. Essa evolução também permitiu que grandes capacidades de processamento e memória fossem embarcadas em um computador portátil, um “tablet” ou um celular, ou mesmo em computadores dedicados em pequenos dispositivos como câmeras e sensores.

Will Douglas Heaven fez um rápido histórico da evolução da IAG, mostrando que os fundamentos surgiram de várias equipes de desenvolvedores.[4] Um desses fundamentos é o avanço, a partir da década de 80, na emulação por software da forma em que os neurônios dos animais interagem, formando uma rede neural, com a capacidade de reter e combinar informação para gerar informação a partir de bases textuais – são os modelos de linguagem. Esse avanço beneficiou-se de uma invenção de pesquisadores do Google que permitiu combinar significados na geração de frases com sentido – os "transformers", que viabilizaram redes neurais recorrentes.

Um dos produtos desses avanços foi o processador de linguagem natural (PLN) "Generative Pre-trained Transformer" (GPT), criado em 2018 pela empresa OpenAI, e que evoluiu para as versões GPT-2, GPT-3 (2020) e GPT-4 ou ChatGPT (2022). Sua fonte de dados é a Internet, trazendo para seus resultados todos os riscos da qualidade de informação (ou desinformação) na rede.

A iniciativa da OpenAI não foi a única. Outros grupos de software, além do Google com o LaMDA e o Bard, a Microsoft com um novo Bing (utilizando uma variante do ChatGPT), bem como um derivado do GPT-3 desenvolvido por um consórcio de voluntários conhecido como BLOOM, continuam a avançar no campo da IAG. A Meta também produziu uma variante do GPT-3 com o nome de OPT.

As questões e desafios trazidos por esses sistemas provocam um interessante efeito colateral: o surgimento de várias iniciativas que produzem legitimadores ou detectores dos conteúdos gerados por esses sistemas. Chomsky alerta que os textos resultantes dos PLNs podem ser úteis para nichos específicos, mas diferem profundamente de como humanos raciocinam e usam linguagem.[5]

Baseados nessas diferenças, estão sendo desenvolvidos detectores, ironicamente utilizando os mesmos algoritmos e fontes de informações, e uma resenha de seis deles já existentes foi apresentada por Funmi Looi Somoye.[6] Alguns deles, como o GPTZero, são ainda de uso livre, e anunciam uma precisão de 96% ou mais na detecção de conteúdo gerado por IAG. Confirmada essa capacidade, os detectores passam a ser elementos a considerar nas estratégias de combate à desinformação ou ao uso indevido de conteúdos derivado do uso da IAG – um desafio especialmente para o ambiente acadêmico que busca combater o plágio.

Quem sabe essas possibilidades de resistência poderão mitigar o “fim-do-mundo” preconizado pelos especialistas que assinaram o sombrio manifesto de uma frase mencionado no início deste texto? Karen Hao sintetiza a natureza dos desafios, e notemos que seu texto é de maio de 2021, antes do “tsunami” do ChatGPT e similares, destacando que desvios ou anormalidades da humanidade refletem-se em seus algoritmos:

> “Estudos já mostraram como ideias racistas, sexistas e abusivas estão embutidas nesses modelos. Eles associam categorias como médicos a homens e enfermeiras a mulheres; palavras boas com os brancos e más com os negros. Sonde-os com as instruções certas e eles também começarão a encorajar coisas como genocídio, automutilação e abuso sexual infantil. Por causa de seu tamanho, eles têm uma pegada de carbono incrivelmente alta. Por causa de sua fluência, eles confundem facilmente as pessoas fazendo-as pensar que um humano escreveu suas saídas, o que os especialistas alertam que pode permitir a produção em massa de desinformação.”[7]

Hao lembra também que esses sistemas em grande escala devoram energia em valores comparáveis aos grandes sistemas de mineração de criptomoedas.[8] Mais pessimista é o professor Eugenio Bucci, já sob o impacto do burburinho do ChatGPT:

> “As ferramentas de IA [generativa] vão aos poucos tomando posse dos protocolos discursivos que, desde

> sempre, orientam as condutas humanas. O jargão jurídico é um desses protocolos. O método científico é outro. A atividade dos médicos é um terceiro tipo. As religiões também têm os seus, que não se confundem com os anteriores. Todos esses protocolos têm um traço comum: eles são construídos na linguagem. Quando a IA aprende a falar, como se fosse gente, ela se apropria dos protocolos que formatam comportamentos individuais e sociais e, a partir daí, tudo muda de figura. Como resultado, o ser humano perderá relevância, enquanto os protocolos desumanizados se expandirão. Da nossa irrelevância brotará o ciclo vicioso que vai nos escantear e, depois, nos extinguir. A menos que a democracia tome providências. Segundo o grupo seleto que assinou o manifesto de uma única frase, ainda há tempo."[9]

As ditas "plataformas sociais" representadas nas propostas regulatórias atuais por serviços também tradicionais de busca de informação e de troca de mensagens, priorizando os serviços de maior escala como os oferecidos por empresas como Alphabet, Amazon, Meta, Apple, Microsoft, são uma parte do desafio maior -- o alcance de novos serviços como os oferecidos por variantes da IAG, a profusão de aplicativos envolvendo grandes volumes de recursos financeiros dos cassinos online (a maioria deles sediados em paraísos fiscais), os desafios para a segurança e privacidade nas inúmeras variantes de serviços de nuvem, etc.

Não há alcance nas propostas regulatórias atuais para abranger esses novos desafios. Há ainda outro espaço que essas propostas estão longe de alcançar: o universo cada vez mais diversificado na Internet das Coisas (IoT, na sigla em inglês). Neste espaço há uma infinidade e variedade de dispositivos cuja origem não é clara, em que a responsabilidade pelo software embarcado ("firmware") é difícil de determinar, e em que riscos de segurança não são em consequência mitigados pelos fabricantes.

Atualizações de "firmware" nos bilhões de dispositivos de IoT são na quase totalidade inexistentes. Tampouco há clareza sobre a funcionalidade desses "firmwares" -- para onde uma câmera wi-fi envia de fato as imagens obtidas, que tipo de interação não perceptível um assistente digital caseiro mantém com seu fabricante, etc.

Em suma, há um grande risco das propostas regulatórias, se sacramentadas em lei, já nascerem datadas, ou alcançarem uma parte menor do espaço interativo da Internet. Em particular, um desafio grave aparece agora, com a IAG. Se havia dúvidas sobre o impacto preocupante nos direitos autorais e trabalhistas dessa nova modalidade de interação envolvendo bases de dados gigantescas capturadas (legal ou ilegalmente) da Internet e software sofisticado, o exemplo atual do movimento grevista em Hollywood as elimina.

Atores e atrizes têm suas interpretações usurpadas por empresas que reproduzem suas atuações originais digitalmente em outras performances, muitas vezes sem autorização desses artistas. Roteiristas têm seus textos originais usurpados pelo uso de IAG para gerar novas redações por parte dos estúdios, sem remunerar os autores humanos.

São riscos cujas consequências ainda serão mais precisamente avaliadas, quando a poeira da atual explosão dessas novas modalidades de interação com IAG baixar.

--

1. Ver https://www.safe.ai/statement-on-ai-risk

2. Ver https://www.washingtonpost.com/technology/2023/06/04/ai-bubble-tech-indu...

3. Ver https://en.wikipedia.org/wiki/Dartmouth_workshop

4. Heaven, W.D., "ChatGPT is everywhere. Here's where it came from", MIT Technology Review, fevereiro de 2023.

5. Chomsky, N. et al., "The False Promise of ChatGPT", New York Times, 08-03-2023.

6. Somoye, F.L., "ChatGPT detectors in 2023", PCGuide, abril de 2023.

7. Hao, K., "The race to understand the exhilarating, dangerous world of language AI", Technology Review, 20-05-2021.

8. Ver, por exemplo, Strubell, E., Ganesh, A., McCallum, A., "Energy and Policy Considerations for Deep Learning in NLP", College of Information and Computer Sciences, University of Massachusetts Amherst, 05-06-2019.

9. Bucci, E., “O Inteligentíssimo Fim do Mundo”, O Estado de São Paulo, 06-01-2023.

Categoria:

- poliTICs 36