poliTICs 36 - editorial

Esta edição da poliTICs é dedicada aos novos desafios resultantes do avanço da inteligência artificial generativa (IAG). Um texto introdutório de Carlos A. Afonso faz uma rápida revisão histórica sobre o próprio conceito de inteligência artificial e uma análise dos possíveis impactos culturais, sociais e políticos das inúmeras variantes da IAG.

Publicamos também uma análise de Naomi Klein sobre as implicações amplas do advento da IAG, destacando quatro ilusões (as "alucinações" mencionadas pelos próprios criadores desses sistemas): a IAG resolveria a crise climática; a IAG traria sabedoria à política; os gigantes da tecnologia não quebrarão o mundo; a IAG nos libertará do trabalho penoso.

Trazemos ainda o provocativo (e por que não dizer, irreverente) texto de Cory Doctorow, que analisa o caso específico do envolvimento arriscado do Google com a IAG na esteira das iniciativas da Microsoft, Apple e Meta.

O texto de Claire Wardle descreve os resultados de pesquisa de sua equipe com uma amostra de 20 milhões de mensagens do Facebook, Instagram e Twitter, sintetizando em quatro componentes as ações derivadas de uma visão mais ampla e integrada de como e por que as informações circulam. A autora procura esclarecer as supostas diferenças entre informação errada, desinformação e informação maliciosa, do ponto de vista das pessoas que buscam informações e reagem a elas de várias formas em suas expressões nas redes sociais.

Reproduzimos também o texto do professor César Rodríguez-Garavito, que expõe os resultados de testes com o ChatGPT e resume: "a inteligência artificial generativa pode aumentar a desinformação e as mentiras online, mas também ser uma formidável ferramenta para o exercício legal da liberdade de expressão; pode proteger ou minar os direitos de migrantes e refugiados, dependendo se é usado para monitorá-los ou para detectar padrões de abuso contra eles; e pode ser útil para grupos tradicionalmente marginalizados, mas também pode aumentar os riscos de discriminação contra a comunidade LBGBTI+, cujas identidades fluidas muitas vezes não se encaixam nas caixas algorítmicas de AIs."

A poliTICs também publica o estudo de caso crítico de Tomás Balmaceda, Karina Pedace e Tobias Schleider sobre a iniciativa do governador de Salta, Argentina, de solicitar à Microsoft uma plataforma de IA para, segundo, ele, "prevenir a gravidez na adolescência usando inteligência artificial com a ajuda de uma empresa de software de renome mundial". Os autores descrevem mais um exemplo confirmando que "não existe 'IA objetiva' ou IA que não seja contaminada por valores humanos", exemplificando com um revelador isomorfismo entre *O Mágico de Oz* e a condição inevitavelmente humana da IA.

Boa leitura!

O trem da IA descarrilou?

Carlos A. Afonso

> 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".
>
> -- Thimothy 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

[1] Ver *https://www.safe.ai/statement-on-ai-risk*
[2] Ver *https://www.washingtonpost.com/technology/2023/06/04/ai-bubble-tech-industry-outlook/*
[3] Ver *https://en.wikipedia.org/wiki/Dartmouth_workshop*

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

[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.

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

---

[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.

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.

Máquinas de IA não estão “alucinando”. Mas seus criadores estão[1]

Naomi Klein* – 05-08-2023

Nos muitos debates que giram em torno do rápido lançamento de novos sistemas da chamada inteligência artificial (IA), há uma escaramuça relativamente obscura focada na escolha da palavra “alucinar”.

Este é o termo que os arquitetos e impulsionadores da IA generativa[2] escolheram para caracterizar as respostas fornecidas por chatbots que são totalmente fabricadas ou totalmente erradas. Como, por exemplo, quando você pede a um bot uma definição de algo que não existe e ele, de forma bastante convincente, lhe dá uma definição completa, com notas de rodapé inventadas.[3] “Ninguém neste campo resolveu ainda os problemas de alucinação”, disse Sundar Pichai, CEO do Google e da Alphabet, a um entrevistador recentemente.[4]

Isso é verdade – mas por que chamar os erros de “alucinações”? Por que não lixo algorítmico? Ou falhas? Alucinação refere-se à misteriosa capacidade do cérebro humano de perceber fenômenos que não estão presentes, pelo menos não em termos convencionais e materialistas. Ao se apropriar de uma palavra comumente usada em psicologia, no campo dos psicodélicos e em várias formas de misticismo, os impulsionadores da IA, embora reconheçam a falibilidade de suas máquinas, ao mesmo tempo alimentam a mitologia mais querida do setor: ao construir esses grandes modelos de linguagem e treiná-los em tudo o que nós, humanos, já escrevemos, dissemos e representamos visualmente, eles pretendem gerar uma inteligência animada prestes a desencadear um salto evolutivo para nossa espécie. De que outra forma bots como Bing e Bard poderiam estar viajando à solta no éter?

Alucinações distorcidas estão de fato acontecendo no mundo da IA, no entanto – mas não são os bots que alucinam; foram os CEOs de tecnologia que as desencadearam, junto com suas falanges fãs, que caíram nas garras de alucinações selvagens, individuais ou coletivas. Aqui estou definindo alucinação não no sentido místico ou psicodélico - estados alterados da mente que podem de fato ajudar no acesso a verdades profundas e anteriormente não percebidas. Não. Essas pessoas estão simplesmente viajando: vendo, ou pelo menos alegando ver, evidências que não existem, e até mesmo criando ilusões sobre mundos inteiros que colocarão seus produtos em uso para nossa elevação e educação universais.

A IA generativa acabará com a pobreza, dizem eles. Ela vai curar todas as doenças. Vai mitigar as mudanças climáticas. Tornará nosso trabalho mais significativo e emocionante. Irá desencadear vidas de lazer e contemplação, ajudando-nos a recuperar a humanidade que perdemos para a mecanização do capitalismo tardio. Vai

---

[1] Publicado originalmente no jornal inglês *The Guardian*:
https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein
Reproduzido em português com autorização da autora.

[2] *https://pt.wikipedia.org/wiki/Intelig%C3%AAncia_artificial_generativa*

[3] *https://www.wsj.com/articles/hallucination-when-chatbots-and-people-see-what-isnt-there-91c6c88b*

[4] *https://www.cbs.com/shows/video/SR6ZcCYjoD3O0sn_ZmVUw87daawsZ5V3/*

acabar com a solidão. Isso tornará nossos governos racionais e responsivos. Essas, eu temo, são as verdadeiras alucinações da IA - e as ouvimos de forma recorrente desde que o ChatGPT foi lançado no final de 2022.

Existe um mundo em que a IA generativa, como uma poderosa ferramenta de pesquisa preditiva e executora de tarefas tediosas, poderia de fato ser organizada para beneficiar a humanidade, outras espécies e nosso lar compartilhado.[5] Mas para que isso aconteça, essas tecnologias precisariam ser implantadas dentro de uma ordem econômica e social muito diferente da nossa, que tivesse como propósito atender às necessidades humanas e proteger os sistemas que sustentam toda a vida no planeta.

E, como aqueles de nós que não estão iludidos sabem bem, nosso sistema atual não é nada disso. Em vez disso, é construído para maximizar a extração de riqueza e lucro – tanto dos humanos quanto do mundo natural – uma realidade que nos trouxe ao que poderíamos pensar como o estágio "necrotécnico" do capitalismo. Nessa realidade de poder e riqueza hiperconcentradas, a IA – longe de corresponder a todas essas alucinações utópicas – tem muito mais probabilidade de se tornar uma ferramenta temível de mais desapropriação e espoliação.

Vou me aprofundar sobre por que isso acontece. Mas primeiro é útil pensar sobre a que *propósito* servem as alucinações utópicas sobre IA. Qual o papel dessas histórias benevolentes em nossas culturas, no momento em que nos deparamos com essas estranhas novas ferramentas? Aqui está uma hipótese: elas são as manchetes poderoras e atraentes daquilo que pode vir a ser o maior e mais importante roubo da história da humanidade. Isso porque o que estamos testemunhando são as empresas mais ricas da história (Microsoft, Apple, Google, Meta, Amazon …) apoderando-se unilateralmente da soma total do conhecimento humano que existe em formato digital capturável, e emparedando-o dentro de produtos proprietários - muitos dos quais irão mirar diretamente nos humanos cujas vidas inteiras de trabalho serviu para treinar as máquinas, sem que as pessoas dessem permissão ou consentimento para isso.

Isso não deveria ser juridicamente legal. No caso de material protegido por direitos autorais - que agora sabemos que treinou os modelos (incluindo este jornal)[6]-, vários processos foram movidos argumentando que isso era evidentemente ilegal.[7] Por que, por exemplo, deveria ser permitido a uma empresa com fins lucrativos copiar as pinturas, desenhos e fotografias de artistas vivos para dentro de um aplicativo como o Stable Diffusion ou o Dall-E 2, e usá-las para gerar versões *doppelganger*[8] do trabalho desses artistas, beneficiando muita gente, menos os próprios artistas?

A pintora e ilustradora Molly Crabapple está ajudando a liderar um movimento de artistas que apontam para esse roubo. "Os geradores de arte de IA são treinados em enormes conjuntos de dados, contendo milhões e milhões de imagens protegidas por direitos autorais, colhidas sem o conhecimento de seus criadores, muito menos compensação ou consentimento. Este é efetivamente o maior roubo de arte da história. Perpetrado por entidades corporativas aparentemente respeitáveis apoiadas pelo

[5] *https://www.nature.com/articles/d41586-020-03348-4*
[6] *https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/*
[7] *https://news.artnet.com/art-world/class-action-lawsuit-ai-generators-deviantart-midjourney-stable-diffusion-2246770*
[8] *https://pt.wikipedia.org/wiki/Doppelg%C3%A4nger*

capital de risco do Vale do Silício. É roubo à luz do dia", afirma uma nova carta aberta da qual ela é co-autora.[9]

O truque, é claro, é que o Vale do Silício costuma chamar esse roubo de "disrupção" – em geral impunemente. Conhecemos este movimento: atacar em território sem lei; afirmar que as regras antigas não se aplicam à sua nova tecnologia; gritar que a regulamentação só ajudará a China – tudo isso enquanto você estabelece solidamente uma boa posição para seus interesses. Até chegar o momento em que todos superamos a novidade desses novos brinquedos e começarmos a fazer um balanço dos destroços sociais, políticos e econômicos, a tecnologia já se tornou tão onipresente, que tribunais e formuladores de políticas acabam desistindo da batalha.[10]

Vimos isso na digitalização de arte e livros do Google; com a colonização espacial de Musk; com o ataque do Uber à indústria de táxis; com o ataque do Airbnb ao mercado de aluguel; com a promiscuidade do Facebook manipulando nossos dados. Não peça permissão, gostam de dizer os disruptores, peça perdão (e lubrifique os pedidos com generosas contribuições de campanha).

No livro *The Age of Surveillance Capitalism*, Shoshana Zuboff detalha meticulosamente como os mapas do Street View do Google ultrapassaram as normas de privacidade ao enviar seus carros com câmeras para fotografar nossas vias públicas e o exterior de nossas casas.[11] No momento em que os processos defendendo os direitos de privacidade começaram, o Street View já era tão onipresente em nossos dispositivos (e tão legal e tão conveniente …) que poucos tribunais fora da Alemanha12 estavam dispostos a intervir.

Agora, a mesma coisa que aconteceu com o exterior de nossas casas está acontecendo com nossas palavras, nossas imagens, nossas músicas, toda a nossa vida digital. Tudo isso está sendo capturado e usado para treinar as máquinas que simulam o pensamento e a criatividade. Essas empresas devem saber que estão envolvidas em roubo ou, pelo menos, que pode haver fortes evidências disso[13]. Elas estão apenas esperando que a velha tática funcione mais uma vez – que a escala do roubo seja tão grande e se desenrole com tanta rapidez que os tribunais e os formuladores de políticas irão mais uma vez dar o braço a torcer diante da suposta inevitabilidade de tudo isso.[14]

É também por isso que suas alucinações sobre todas as coisas maravilhosas que a IA fará pela humanidade são tão importantes. Porque essas previsões arrogantes disfarçam esse roubo massivo e o apresentam como uma dádiva – ao mesmo tempo em que ajudam a racionalizar os perigos inegáveis da IA.

A essa altura, a maioria de nós já ouviu falar sobre a pesquisa que pediu a cientistas e desenvolvedores de IA para estimar a probabilidade de que sistemas avançados de IA

---

[9] *https://artisticinquiry.org/AI-Open-Letter*
[10] *https://www.reuters.com/article/us-google-books-idUSKCN0SA1S020151016*
[11] *https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook*
[12] *https://archive.nytimes.com/bits.blogs.nytimes.com/2013/04/23/germanys-complicated-relationship-with-google-street-view/*
[13] *https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem*
[14] *https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/*

causem "extinção humana ou desempoderamento igualmente permanente e severo da espécie humana". Assustadoramente, a resposta média foi que havia 10% de chance.[15]

Como alguém racionaliza trabalhar no desenvolvimento de ferramentas que carregam tais riscos existenciais? Muitas vezes, a razão dada é que esses sistemas também carregam enormes vantagens potenciais – não levando em conta que essas vantagens são, em sua maior parte, alucinatórias. Vamos aprofundar-nos em alguns dos riscos mais graves.

Alucinação nº 1: a IA resolverá a crise climática

Quase invariavelmente, no topo das listas de vantagens da IA está a afirmação de que esses sistemas resolverão de alguma forma a crise climática. Ouvimos isso de todos, do Fórum Econômico Mundial[16] ao Conselho de Relações Exteriores[17] e ao Boston Consulting Group[18], que explica que a IA "pode ser usada para apoiar todas as partes interessadas a adotar uma abordagem mais informada e orientada por dados para combater as emissões de carbono e construir uma sociedade mais verde. Também pode ser empregada para redirecionar os esforços climáticos globais para as regiões de maior risco".

O ex-CEO do Google, Eric Schmidt, resumiu o caso quando disse à revista *Atlantic* que vale a pena correr os riscos da IA, porque "se você pensar nos maiores problemas do mundo, eles são todos realmente difíceis – mudança climática, organizações humanas e assim por diante. Portanto, eu sempre quero que as pessoas sejam mais inteligentes".[19]

De acordo com essa lógica, o fracasso em "resolver" grandes problemas como a mudança climática se deve a um déficit de inteligência. Não importa que pessoas inteligentes, cheias de PhDs e prêmios Nobel, digam aos nossos governos há décadas o que precisa acontecer para sairmos dessa bagunça: reduzir nossas emissões, deixar o carbono no solo, combater o consumo excessivo dos ricos e o subconsumo dos pobres, porque nenhuma fonte de energia é isenta de custos ecológicos.

A razão pela qual esse conselho muito inteligente foi ignorado não é devido a um problema de compreensão de leitura ou porque, de alguma forma, precisamos de máquinas para pensar por nós. É porque fazer o que a crise climática exige de nós encalharia trilhões de dólares em ativos de combustíveis fósseis, ao mesmo tempo em que desafiaria o modelo de crescimento baseado no consumo, que é o coração de nossas economias interconectadas.[20] A crise climática não é, de fato, um mistério ou um enigma que ainda não resolvemos devido a conjuntos de dados insuficientemente robustos. Sabemos o que seria necessário, mas não é uma solução rápida – é uma mudança de paradigma. Esperar que as máquinas cuspam uma resposta mais palatável e/ou rentável não é uma cura para esta crise, é mais um sintoma dela.

Deixando de lado as alucinações, é muito mais provável que a IA seja trazida ao mercado de maneira a aprofundar gravemente a crise climática. Primeiro, os servidores gigantes que possibilitam ensaios instantâneos e obras de arte de chatbots

---

[15] *https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/#Extinction_from_AI*
[16] *https://www.weforum.org/agenda/2021/08/how-ai-can-fight-climate-change/*
[17] *https://world101.cfr.org/global-era-issues/climate-change/how-can-artificial-intelligence-combat-climate-change*
[18] *https://www.bcg.com/publications/2022/how-ai-can-help-climate-change*
[19] *https://www.theatlantic.com/technology/archive/2023/03/open-ai-gpt4-chatbot-technology-power/673421/*
[20] *https://www.wsj.com/articles/trillions-in-assets-may-be-left-stranded-as-companies-address-climate-change-11637416980*

são uma fonte enorme e crescente de emissões de carbono.[21] Em segundo lugar, à medida que empresas como a Coca-Cola começam a fazer grandes investimentos em IA generativa para vender mais produtos[22], fica muito evidente que essa nova tecnologia será usada da mesma forma que a mais recente geração de ferramentas digitais: aquilo que começa com promessas de espalhar a liberdade e a democracia acaba em microanúncios direcionados a nós, para que compremos mais coisas inúteis que expelem carbono.

E há um terceiro fator, este um pouco mais difícil de definir. Quanto mais nossos canais de mídia são inundados com *deep fakes* e clones de vários tipos, mais temos a sensação de estar afundando em areia movediça informacional. Geoffrey Hinton, muitas vezes tido como "o padrinho da IA" - porque a rede neural que ele desenvolveu há mais de uma década forma os blocos de construção dos grandes modelos de linguagem de hoje - entende isso muito bem. Ele acabou de deixar um cargo sênior no Google para poder falar livremente sobre os riscos da tecnologia que ajudou a criar, incluindo, como disse ao *New York Times*, o risco de que as pessoas "não sejam mais capazes de saber o que é verdade"[23].

Isso é altamente relevante para a crença que a IA ajudará a combater a crise climática. Porque quando desconfiamos de tudo o que vemos e lemos em nosso ambiente de mídia cada vez mais misterioso, ficamos ainda menos preparados para resolver problemas coletivos urgentes. A crise de confiança é anterior ao ChatGPT, é claro, mas não há dúvida que uma proliferação de *deep fakes* será acompanhada por um aumento exponencial de culturas de conspiração já muito presentes. Então, que diferença fará se a IA apresentar avanços tecnológicos e científicos? Se o tecido da realidade compartilhada estiver se desfazendo em nossas mãos, seremos incapazes de responder com qualquer coerência.

Alucinação nº 2: a IA proporcionará uma governança sábia

Essa alucinação evoca um futuro próximo em que políticos e burocratas, aproveitando a vasta inteligência agregada de sistemas de IA, são capazes de "ver padrões de necessidade e desenvolver programas baseados em evidências" que trazem maiores benefícios para seus eleitorados. Essa afirmação vem de um artigo publicado pela fundação do Boston Consulting Group[24], mas está reverberando em muitos *thinktanks* e consultorias de gestão. E é revelador que essas empresas em particular – as firmas contratadas por governos e outras corporações para identificar economias de custos, muitas vezes demitindo um grande número de trabalhadores – foram as mais rápidas a aderir ao movimento da IA. A PwC (anteriormente PricewaterhouseCoopers) acaba de anunciar um investimento de US$ 1 bilhão[25] ,e a Bain & Company, assim como a Deloitte, estão entusiasmadas com o uso dessas ferramentas para tornar seus clientes mais "eficientes".

Tal como acontece com as reivindicações climáticas, é necessário perguntar: a razão pela qual os políticos impõem políticas públicas cruéis e ineficazes é a falta de

---

[21] *https://penntoday.upenn.edu/news/hidden-costs-ai-impending-energy-and-resource-strain*
[22] *https://www.coca-colacompany.com/news/coca-cola-invites-digital-artists-to-create-real-magic-using-new-ai-platform*
[23] *https://www.nytimes.com/2023/05/01/technology/ai-google-chatbot-engineer-quits-hinton.html*
[24] *https://www.centreforpublicimpact.org/*
[25] *https://venturebeat.com/ai/the-power-of-infrastructure-purpose-built-for-ai/*

evidências? Uma incapacidade de "ver padrões", como sugere o artigo do BCG? Eles não entendem os custos humanos de reduzir as despesas em saúde pública em meio a pandemias[26], ou de abandonar o investimento público em moradias quando as barracas lotam nossos parques urbanos, ou de aprovar novas infraestruturas de exploração de combustíveis fósseis enquanto as temperaturas sobem? Os políticos precisam de IA para torná-los "mais inteligentes", para usar o termo de Schmidt – ou são inteligentes o suficiente para saber quem vai apoiar sua próxima campanha ou financiar seus rivais, caso mudem seu discurso?

Seria muito bom se a IA realmente pudesse cortar o vínculo entre o dinheiro corporativo e a formulação de políticas imprudentes – mas esse vínculo tem tudo a ver com o motivo pelo qual empresas como Google e Microsoft foram autorizadas a liberar seus chatbots ao público, apesar da avalanche de avisos e riscos conhecidos. Schmidt e outros estão em uma campanha de lobby de anos, dizendo aos políticos em Washington que, se não forem livres para avançar com IA generativa, livres do peso de uma regulamentação séria, as potências ocidentais serão deixadas para trás pela China[27]. No ano passado, as principais empresas de tecnologia gastaram um recorde de US$ 70 milhões para fazer lobby em Washington – mais do que o setor de petróleo e gás – e essa quantia, observa a *Bloomberg News*, está bem além dos milhões gastos "em sua ampla gama de grupos comerciais, instituições sem fins lucrativos e *thinktanks*"[28].

E, no entanto, apesar do conhecimento íntimo destas empresas sobre como o dinheiro molda a política em nossas capitais nacionais, quando você ouve Sam Altman, o CEO da OpenAI – criador do ChatGPT – falar sobre os melhores cenários para seus produtos, tudo isso parece ser esquecido. Em vez disso, ele parece estar criando a alucinação de um mundo totalmente diferente do nosso, no qual os políticos e a indústria tomam decisões com base nos melhores dados e nunca colocariam inúmeras vidas em risco por lucro e vantagem geopolítica. O que nos leva a outra alucinação.

Alucinação nº 3: pode-se confiar nos gigantes da tecnologia para não quebrar o mundo

Questionado se está preocupado com a frenética corrida do ouro que o ChatGPT já desencadeou, Altman disse que sim, mas acrescentou com otimismo: "espero que tudo acabe dando certo". Sobre seus colegas CEOs de tecnologia – os que competem para lançar seus chatbots rivais – ele disse: "acho que os bons anjos vencerão".[29]

Bons anjos? No Google? Tenho certeza que a empresa demitiu a maioria deles porque estavam publicando artigos críticos sobre IA ou denunciando a empresa por racismo e assédio sexual no local de trabalho[30]. Outros "anjos bons" desistiram alarmados, sendo o mais recente deles, Hinton[31]. Isso porque, ao contrário das alucinações das pessoas que mais lucram com a IA, o Google não toma decisões com base no que é melhor para o mundo – ela toma decisões com base no que é melhor para os acionistas da Alphabet,

[26] *https://www.theguardian.com/society/2022/aug/03/how-the-tory-party-has-systematically-run-down-the-nhs*
[27] *https://epic.org/wp-content/uploads/foia/epic-v-ai-commission/EPIC-19-09-11-NSCAI-FOIA-20200331-3rd-Production-pt9.pdf*
[28] *https://www.bnnbloomberg.ca/tech-giants-broke-their-spending-records-on-lobbying-last-year-1.1877988*
[29] *https://steno.ai/lex-fridman-podcast-10/367-sam-altman-openai-ceo-on-gpt-4-chatgpt-and*
[30] *https://www.engadget.com/google-fires-ai-researcher-over-paper-challenge-132640478.html*
[31] *https://www.engadget.com/google-engineers-leave-over-timnit-gebru-exit-093645678.html*

que não querem perder a última bolha, especialmente quando a Microsoft, a Meta e a Apple já entraram nela.

Alucinação nº 4: a IA nos libertará do trabalho penoso

Se as alucinações aparentemente inofensivas do Vale do Silício parecem plausíveis para muitos, há uma razão simples para isso. A IA generativa está atualmente no que poderíamos chamar de estágio de falso socialismo. Isso faz parte de um já conhecido manual do Vale do Silício. Primeiro, crie um produto atrativo (um motor de busca, uma ferramenta de mapeamento, uma rede social, uma plataforma de vídeo, um aplicativo de viagens…); distribua-o de graça ou quase de graça por alguns anos, sem nenhum modelo de negócios viável discernível ("brinque com os bots", eles nos dizem, "veja que coisas divertidas você pode criar!"); faça muitas afirmações grandiosas sobre como você só está fazendo isso porque deseja criar uma "praça pública virtual" ou uma "comunidade de informação" ou "conectar as pessoas", enquanto espalha liberdade e democracia (e sem ser "maligno"). Então observe enquanto as pessoas aderem ao uso dessas ferramentas gratuitas, e seus concorrentes declaram falência. Uma vez que o terreno esteja limpo, introduza os anúncios direcionados, a vigilância constante, os contratos com instituições policiais e militares, as vendas de dados às escondidas e as crescentes taxas de assinatura.

Muitas vidas e setores foram dizimados por iterações anteriores deste manual, de motoristas de táxi a mercados de aluguel e jornais locais. Com a revolução da IA, esses tipos de perdas podem parecer erros irrelevantes, com professores, programadores, artistas visuais, jornalistas, tradutores, músicos, profissionais de saúde e tantos outros enfrentando a perspectiva de ter suas rendas substituídas por códigos problemáticos.

Não se preocupe, alucinam os entusiastas da IA – será maravilhoso. Quem gosta de trabalhar afinal? Dizem-nos que a IA generativa não será o fim do emprego, apenas o fim do "trabalho chato"[32] – com os chatbots prestativamente desempenhando todas as tarefas repetitivas e destruidoras de almas, e os humanos meramente supervisionando-os. Altman, por sua vez, vê um futuro onde o trabalho "pode ser um conceito mais amplo, não algo que você tenha que fazer para poder comer, mas algo que você faça como uma expressão criativa e uma forma de encontrar realização e felicidade"[33].

Essa é uma visão empolgante de uma vida mais bonita e tranquila, compartilhada por muitos esquerdistas (incluindo o genro de Karl Marx, Paul Lafargue, que escreveu um manifesto intitulado *O Direito à Preguiça*) [34]. Mas nós, esquerdistas, também sabemos que, se ganhar dinheiro não for mais o imperativo da vida, deverá haver outras maneiras de atender às nossas necessidades de abrigo e sustento. Um mundo sem empregos de baixa qualidade significa que o aluguel deve ser gratuito, a assistência médica gratuita e todas as pessoas devem ter direitos econômicos inalienáveis. E então, de repente, não estamos mais falando sobre IA – estamos falando sobre socialismo.

É exatamente devido ao fato de não vivermos num mundo racional e humanista inspirado em *Star Trek,* que Altman parece estar alucinando. Vivemos sob o capitalismo e, sob esse sistema, inundar o mercado com tecnologias que podem

[32] *https://www.nytimes.com/2023/04/22/opinion/jobs-ai-chatgpt.html*
[33] *https://steno.ai/lex-fridman-podcast-10/367-sam-altman-openai-ceo-on-gpt-4-chatgpt-and*
[34] *https://www.marxists.org/archive/lafargue/1883/lazy/*

executar de forma plausível as tarefas econômicas de incontáveis trabalhadores não faz com que as pessoas estejam repentinamente livres para se tornarem filósofas e artistas. Isso significa que essas pessoas se descobrirão à beira do abismo – e os artistas de verdade estarão entre os primeiros a cair.

Essa é a mensagem da carta aberta de Crabapple, que convida "artistas, editores, publishers, e líderes de sindicatos de jornalistas a comprometerem-se com os valores humanos contra o uso de imagens geradas por IA" e "a comprometerem-se a apoiar a arte editorial feita por pessoas, não por fazendas de servidores". A carta, agora assinada por centenas de artistas, jornalistas e outros profissionais, afirma que todos, exceto os artistas mais elitistas, encontram seu trabalho "em risco de extinção"[35]. E de acordo com Hinton, o "padrinho da IA", não há razão para acreditar que a ameaça não se espalhe. Os chatbots "tiram o trabalho árduo", mas "podem tirar mais do que isso".

Crabapple e seus co-autores escrevem: "A arte feita por IA generativa é vampírica, banqueteando-se com gerações passadas de obras de arte, mesmo quando suga a força vital de artistas vivos". Mas há formas de resistir: podemos nos recusar a usar esses produtos e nos organizar para exigir que nossos empregadores e governos também os rejeitem.

Uma carta de proeminentes estudiosos da ética da IA, incluindo Timnit Gebru, que foi demitida pelo Google em 2020 por desafiar a discriminação no local de trabalho, apresenta algumas das ferramentas regulatórias que os governos podem introduzir imediatamente – incluindo total transparência sobre quais conjuntos de dados estão sendo usados para treinar o modelos[36]. Os autores escrevem: "Não só deve estar sempre explícito quando estamos diante de mídia sintética, mas as organizações que constroem esses sistemas também devem ser obrigadas a documentar e divulgar os dados de treinamento e os modelos de arquiteturas... Devemos construir máquinas que funcionem para nós, em vez de 'adaptar' a sociedade para ser legível e gravável por máquinas."

Embora as empresas de tecnologia queiram que acreditemos que já é tarde demais para reverter esse produto de imitação em massa e substituto humano, existem precedentes legais e regulatórios altamente relevantes que podem ser aplicados. Por exemplo, a Comissão Federal de Comércio dos EUA (FTC) forçou a Cambridge Analytica, bem como a Everalbum, proprietária de um aplicativo de fotos, a destruir algoritmos inteiros que foram treinados com fotos capturadas e dados obtidos de forma ilegítima[37]. Em seus primeiros dias, o governo Biden fez muitas afirmações ousadas sobre a regulamentação das *big techs*, incluindo reprimir o roubo de dados pessoais para construir algoritmos proprietários. Com uma eleição presidencial aproximando-se rapidamente, agora seria um bom momento para cumprir essas promessas – e evitar o próximo conjunto de demissões em massa antes que elas aconteçam.

Um mundo de *deep fake*, círculos infinitos de imitação e piora da desigualdade não é uma realidade inevitável. É um conjunto de escolhas políticas. Podemos eliminar o

[35] *https://artisticinquiry.org/AI-Open-Letter*
[36] *https://www.dair-institute.org/blog/letter-statement-March2023*
[37] *https://digiday.com/media/why-the-ftc-is-forcing-tech-firms-to-kill-their-algorithms-along-with-ill-gotten-data/*

atual modelo de chatbots vampíricos – e começar a construir o mundo no qual as promessas mais promissoras da IA seriam mais do que alucinações do Vale do Silício.

Isso porque fomos nós que treinamos as máquinas. Todos e todas nós. Mas nunca demos o nosso consentimento. Eles se alimentaram da engenhosidade, da inspiração e das descobertas coletivas da humanidade (juntamente com algumas de nossas características mais desprezíveis). Esses modelos são máquinas de aprisionamento e apropriação, devorando e privatizando nossas vidas individuais, bem como nossas heranças intelectuais e artísticas coletivas. E o objetivo deles nunca foi resolver a mudança climática ou tornar nossos governos mais responsáveis, ou nossa vida diária mais tranquila. Sempre foi lucrar com a miséria em massa que, sob o capitalismo, é a consequência flagrante e lógica da substituição de funções humanas por bots.

Tudo isso é excessivamente dramático? Uma resistência enfadonha e reflexiva às inovações empolgantes? Por que esperar o pior? Altman tenta tranquilizar-nos : "Ninguém quer destruir o mundo"[38]. Talvez não. Mas, como a crise climática e a crise de extinção cada vez mais graves nos mostram todos os dias, muitas pessoas e instituições poderosas parecem saber muito bem que estão ajudando a destruir a estabilidade dos sistemas planetários de suporte à vida, desde que possam continuar alcançando recordes de lucros que eles acreditam que irão protegê-los e às suas famílias dos piores efeitos[39].

Altman, como muitas criaturas do Vale do Silício, é um prevenido – em 2016, ele se gabou: "Tenho armas, ouro, iodeto de potássio, antibióticos, baterias, água, máscaras de gás da Força de Defesa de Israel e um grande pedaço de terra em Big Sur para onde eu posso voar."[40]

Tenho certeza de que esses fatos dizem muito mais sobre o que Altman realmente acredita em relação ao futuro que ele está ajudando a desencadear do que quaisquer alucinações floridas que ele está escolhendo compartilhar em entrevistas à imprensa.

(*) Naomi Klein é jornalista canadense premiada e autora reconhecida como best-seller no *New York Times*. Colunista do jornal inglês *The Guardian*, em 2018 foi nomeada para a cátedra inaugural Gloria Steinem na Rutgers University. É professora honorária de Mídia e Clima na Rutgers. Em setembro de 2021, ingressou na Universidade da Colúmbia Britânica como professora titular de Justiça Climática e codiretora do Centro de Justiça Climática da mesma universidade.

[38] *https://steno.ai/lex-fridman-podcast-10/367-sam-altman-openai-ceo-on-gpt-4-chatgpt-and*
[39] *https://www.theguardian.com/business/2023/apr/28/exxonmobil-chevron-record-profits*
[40] *https://www.newyorker.com/magazine/2016/10/10/sam-altmans-manifest-destiny*

https://pluralistic.net/2023/02/16/tweedledumber/#easily-spooked

O pânico do chatbot do Google : sobre as infinitas inseguranças de um autodenominado gênio criativo que realmente apenas compra as ideias de outras pessoas.

Cory Doctorow – 16-02-2023

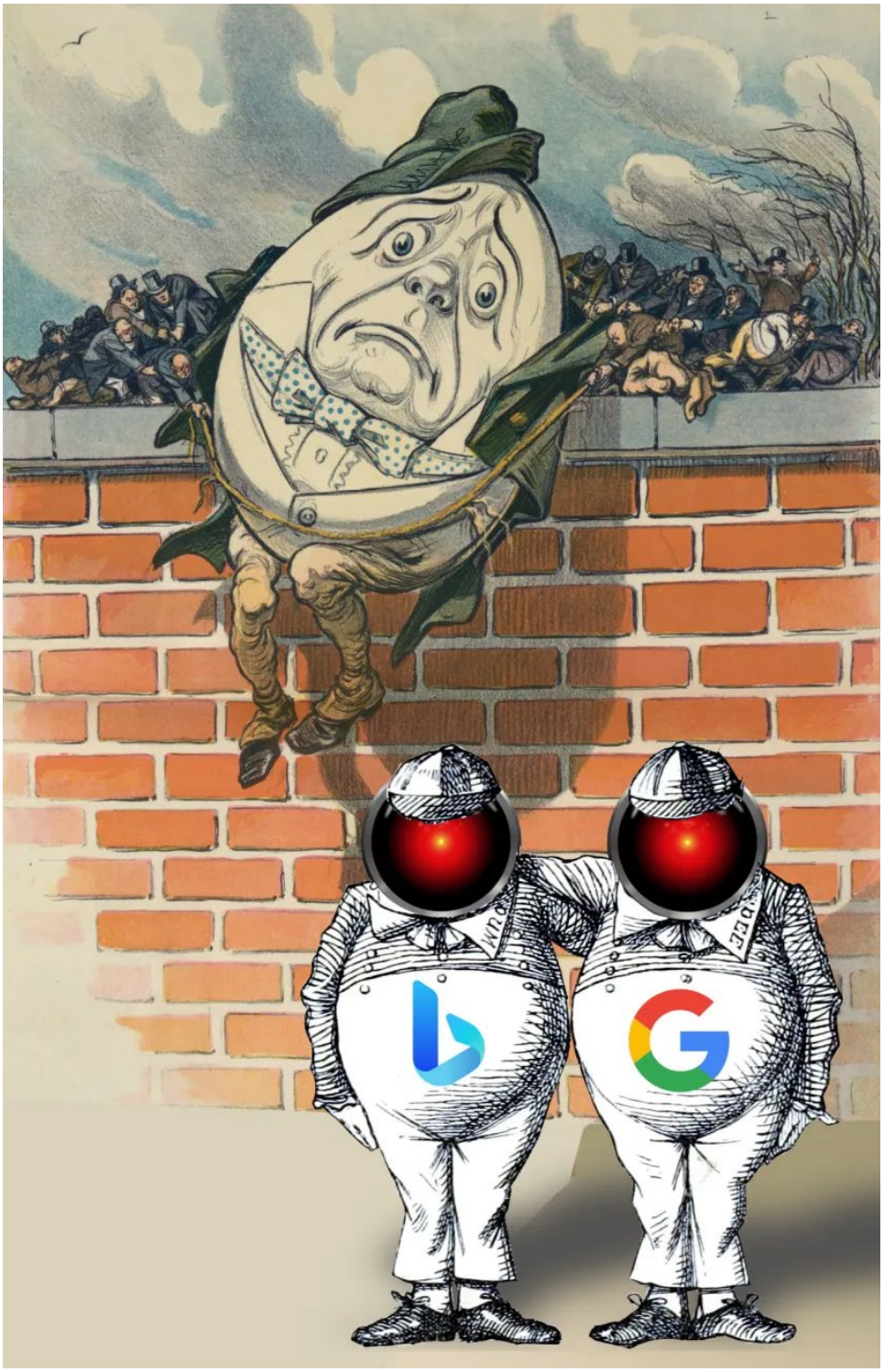

Imagem: Cryteria , CC BY 3.0 , modificado

O realmente notável não é apenas que a Microsoft decidiu que o futuro da pesquisa não é links para materiais relevantes, mas, em vez disso, parágrafos longos e floreados

escritos por um chatbot que por acaso é um mentiroso habitual – ainda mais notável é que o Google concorda .

A Microsoft não tem nada a perder. Ela gastou bilhões no Bing, um mecanismo de busca que ninguém usa voluntariamente. Seria melhor tentar algo tão estúpido que talvez funcione. Mas por que o Google, um monopolista com mais de 90% de participação nas pesquisas em todo o mundo, está pulando da mesma ponte que a Microsoft?

Há um delicioso tópico do Mastodon sobre isso, escrito por Dan Hon, onde ele compara os sistemas de busca emporcalhados[1] pelo chatbot ao Bing e ao Google nesta conversa entre Alice, Tweedledee e Tweedledum:[2]

> Na frente da casa, Alice encontrou dois curiosos personagens, ambos buscadores.
>
> "'Eu sou o Google-E', disse aquele recheado com anúncios.
>
> "'E eu sou Bingle-Dum', disse o outro, que era o menor dos dois, e fazia beicinho, por ter menos visitas e oportunidades de conversa do que o outro.
>
> "'Eu conheço você', disse Alice. 'Você vai me apresentar um quebra-cabeça? Talvez um de vocês diga a verdade e o outro minta?'
>
> "'Ah, não', disse Bingle-Dum.
>
> "'Nós dois mentimos', acrescentou o Google-E."
>
> A conversa vai ficando melhor:
>
> "'Esta é realmente uma situação intolerável. Se vocês dois mentem...'
>
> " '... e mentimos de forma convincente', acrescentou Bingle-Dum.
>
> "'Sim, obrigado. Se for assim, então como vou confiar em qualquer um de vocês?'
>
> "Googl-E e Bingle-Dum se viraram e deram de ombros."

A busca por chatbot é uma péssima ideia, especialmente em uma era em que a Web estará provavelment recheada com besteirol de IA, a tagarelice constante de papagaios estocásticos que alimenta esses bots.[3]

A estratégia de uso de chatbots do Google não deveria adicionar mais "madlibs"[4] à internet – ao contrário, deveria descobrir como excluir a desinformação dos *spammers* e os conteúdos criados para otimizar presença nos buscadores (ou, no mínimo, verificar os fatos).

E, no entanto, o Google está apostando tudo nos chatbots, com o CEO da empresa ordenando uma ação agressiva para inserir chatbots onde for possível em todo o

[1] A expressão original é "enshittified".
[2] *https://mamot.fr/@danhon@dan.mastohon.com/109832788458972865*
[3] *https://dl.acm.org/doi/10.1145/3442188.3445922*
[4] *https://en.wikipedia.org/wiki/Mad_Libs*

“googleverso”. Por que diabos a empresa está competindo com a Microsoft para ver quem pode ser o primeiro a saltar do pico das expectativas infladas?[5]

Acabei de publicar uma teoria no *The Atlantic* , sob o título "Como o Google ficou sem ideias", em que recorro à teoria da concorrência para explicar a suada insegurança do Google, um complexo de ansiedade que atormenta a empresa quase desde a sua criação.[6]

A teoria central: há um quarto de século, os fundadores do Google tiveram uma ideia incrível – uma maneira melhor de fazer buscas. O mercado de capitais encheu a empresa de dinheiro, e ela contratou as melhores, mais brilhantes e criativas pessoas que pôde encontrar, mas depois criou uma cultura corporativa incapaz de capitalizar suas ideias.

Todos os produtos produzidos internamente pelo Google – exceto o clone do Hotmail – morreram. Alguns desses produtos eram bons, alguns eram terríveis, mas não importava. O Google – uma empresa que cultivava o capricho de uma fábrica de Willy Wonka – não conseguia “inovar” de jeito nenhum.

Todos os produtos bem-sucedidos do Google, exceto pesquisa e Gmail, são uma aquisição: dispositivos móveis, tecnologia de anúncios,, vídeos, gerenciamento de servidor, documentos, calendário, mapas etc. A empresa deseja desesperadamente ser uma empresa de "fazer coisas", mas na verdade é uma empresa de "comprar coisas". Claro, é bom para operacionalizar e dimensionar produtos, mas isso é uma aposta para qualquer monopolista.[7]

A dissonância cognitiva de um autodenominado "gênio criativo", cujo verdadeiro gênio é gastar o dinheiro de outras pessoas para comprar produtos de outras pessoas e obter crédito por eles, leva as pessoas a fazerem coisas realmente malucas (como qualquer usuário do Twitter pode atestar).

O Google há muito exibe essa patologia. Em meados dos anos 2000 – depois que o Google perseguiu o Yahoo na China e começou a censurar seus resultados de pesquisa e a colaborar com a vigilância estatal – costumávamos dizer que a maneira de levar o Google fazer algo estúpido e autodestrutivo era que o Yahoo fizesse isso primeiro.

Este foi um bom tempo. O Yahoo estava desesperado e falido, um cemitério de aquisições promissoras que foram destruídas e deixadas para sangrar ali mesmo na Internet pública, enquanto os príncipes em duelo da alta administração do Yahoo jogavam um LARP[8] traiçoeiro que os fazia competir para ver quem poderia sabotar os outros. Ir para a China foi um ato de desespero depois que a empresa foi humilhada pela busca imensamente superior do Google. Ver o Google copiar as jogadas idiotas do Yahoo foi desconcertante.

Desconcertante na época, diga-se. Com o passar do tempo e o Google copiando servilmente outros rivais, sua patologia de insegurança se revelou. O Google repetidamente falhou em criar um produto "social" popular e, como o Facebook detinha uma fatia cada vez maior do mercado de anúncios, o Google fez uma pressão

[5] *https://en.wikipedia.org/wiki/Gartner_hype_cycle*
[6] *https://www.theatlantic.com/ideas/archive/2023/02/google-ai-chatbots-microsoft-bing-chatgpt/673052/*
[7] *https://www.eff.org/deeplinks/2020/06/technical-excellence-and-scale*
[8] *https://pt.wikipedia.org/wiki/Live_action_(RPG)*

judicial para competir com ele. A empresa fez da integração do Google Plus um "indicador-chave de desempenho" para todas as divisões, e o resultado foi um pântano bizarro de recursos "sociais" malfadados em todos os produtos do Google -- produtos nos quais bilhões de usuários confiavam para operações de alto risco, que foram repentinamente enfeitados com botões "sociais" que não faziam sentido.

O desastre do G+ foi realmente incrível: alguns recursos e integrações do G+ foram ótimos e desenvolveram seguidores leais, mas foram ofuscados pela insistência incoerente e de cima para baixo de tornar o Google uma empresa "social em primeiro lugar". Quando o G+ entrou em colapso, ele implodiu totalmente, e as partes úteis do G+ nas quais as pessoas passaram a confiar desapareceram junto com as partes estúpidas.

Para quem viveu a tragicomédia do G+, o pivô do Google para Bard – um chatbot para resultados de busca – é terrivelmente familiar. É um verdadeiro momento "morra como herói ou viva o suficiente para tornar-se um vilão". A Microsoft – a monopolista que só foi impedida de estrangular o Google em seu berço pelo trauma de seu arrastamento antitruste – transformou-se de uma empresa de criação de produtos em uma empresa de aquisições e operações, e o Google está logo atrás nisso.

No ano passado, o Google demitiu 12.000 funcionários para agradar um "investidor ativista" -- no mesmo ano, declarou uma recompra de ações de US$70 bilhões, extraindo capital suficiente para pagar os salários desses 12.000 googlers pelos próximos 27 anos. O Google é uma empresa financeira com uma linha lateral em tecnologia de propaganda. Tem que ser assim: quando seu único caminho de sucesso para o crescimento requer acesso aos mercados de capitais para financiar aquisições anticompetitivas, você não pode se dar ao luxo de irritar os deuses do dinheiro, mesmo que tenha uma estrutura de "ação dupla" que permita aos fundadores voto superior a todos os outros acionistas.[9]

O ChatGPT e seus imitadores têm todas as características de uma moda passageira de tecnologia e são realmente os sucessores do Web versão 3 da última temporada e dos “pump-and-dumps”[10] de criptomoedas. Uma das críticas mais claras e inspiradoras dos chatbots vem do escritor de ficção científica Ted Chiang, cuja crítica clássica instantânea foi chamada de "ChatGPT é um JPEG borrado da Web".[11]

Chiang aponta uma diferença fundamental entre as respostas do ChatGPT e dos autores humanos: o primeiro rascunho de um autor humano costuma ser uma ideia original, mal expressa, enquanto o melhor que se pode esperar do ChatGPT é uma ideia não original, expressa com competência. O ChatGPT está perfeitamente posicionado para melhorar o copiar-e-colar de páginas Web que legiões de trabalhadores mal pagos lançam em uma tentativa de subir nos resultados de pesquisa do Google.

Em relação ao ensaio de Chiang no podcast *This Machine Kills*,[12] Jathan Sadowski perfura habilmente a bolha do *hype* do ChatGPT4, que afirma que a próxima versão do chatbot será tão incrível que qualquer crítica à tecnologia atual se tornará obsoleta.

[9] *https://abc.xyz/investor/founders-letters/2004-ipo-letter/*
[10] *https://pt.wikipedia.org/wiki/Pump_and_dump*
[11] *https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web*
[12] *https://soundcloud.com/thismachinekillspod/232-400-hundred-years-of-capitalism-led-directly-to-microsoft-viva-sales*

Sadowski observa que os engenheiros da OpenAI estão fazendo de tudo para garantir que a próxima versão não seja treinada em nenhum dos resultados do ChatGPT3. Isso é revelador: se um grande modelo de linguagem pode produzir materiais tão bons quanto o texto produzido por humanos, então por que os resultados do ChatGPT3 não podem ser usados para criar o ChatGPT4?

Sadowski tem um ótimo termo para descrever esse problema: "IA de Habsburg". Assim como a endogamia real produziu uma geração de supostos super-homens que eram incapazes de se reproduzir, também alimentar um novo modelo com o fluxo de exaustão do último produzirá um giro cada vez pior de bobagens em espiral que eventualmente desaparecem em seu próprio fiofó.

https://issues.org/incompreensão-misinformation-wardle/

Incompreensão da desinformação

Claire Wardle* -- publicado originalmente em 08 de maio de 2023

> A obsessão em medir a precisão de postagens individuais é equivocada. Para fortalecer os ecossistemas de informação, concentre-se em narrativas e em por que as pessoas compartilham o que fazem.

No outono de 2017, o Collins Dictionary nomeou "fake news" como a palavra do ano.[1] Foi difícil contestar a decisão. Os jornalistas estavam usando o termo para aumentar a conscientização sobre informações falsas e enganosas online. Os acadêmicos começaram a publicar copiosamente sobre o assunto e até mesmo a batizar conferências com seu nome.[2] E, claro, o presidente dos EUA, Donald Trump, usava regularmente o epíteto no pódio para desacreditar quase tudo de que não gostava.[3]

Na primavera daquele ano, eu já estava exasperada com a forma como esse termo estava sendo usado para atacar a mídia. O que é pior, o termo nunca refletiu de fato o real problema: a maior parte do conteúdo taxado de "fake news" não era notícia falsa, mas sim conteúdo verdadeiro usado fora de contexto – e muito raramente aparentava de fato ser uma notícia. Eu fiz um apelo para que se parasse de usar a expressão "fake news" e, em vez disso, que se passasse a usar as expressões "informação errada", "desinformação" e "desinformação maliciosa" sob o conceito guarda-chuva "transtorno da informação".[4] [5] Esses termos, especialmente os dois primeiros, passaram a ser bastante adotados - mas representam uma estrutura excessivamente simplificada e arrumadinha que eu já não considero útil.

Tanto "desinformação" quanto "informação errada" descrevem afirmações mentirosas ou enganosas, mas a desinformação é distribuída com a intenção de causar danos, enquanto a informação errada é o compartilhamento equivocado da informação enganosa. As análises de ambos os tipos de conteúdo geralmente se concentram em saber se uma postagem é precisa e se tem a intenção de enganar. O resultado? Nós, pesquisadores, ficamos tão obcecados em rotular os pontos que não conseguimos ver o padrão maior que eles revelam.

Ao se concentrar estritamente no conteúdo problemático, os pesquisadores falham em compreender o número cada vez maior de pessoas que criam e compartilham esse tipo de conteúdo e também negligenciam a análise do contexto mais amplo, de quais informações as pessoas realmente precisam. Os acadêmicos não vão fortalecer efetivamente o ecossistema informacional até que mudemos nossa perspectiva - passando da classificação de cada postagem à compreensão dos contextos sociais dessas informações, de como elas se encaixam em narrativas e identidades e quais seus impactos de curto prazo e danos de longo prazo.

[1]Ver *https://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist*
[2]Ver *https://firstmonday.org/ojs/index.php/fm/article/view/11645/10152*
[3]Ver *https://www.nytimes.com/video/us/politics/100000004865825/trump-calls-cnn-fake-news.html*
[4]Ver *https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77*
[5]NT: Utilizamos os termos em português com base na tese de doutorado de Tatiana Dourado. Ver *https://www.researchgate.net/publication/342082587_Fake_news_na_eleicao_presidencial_de_2018_no_Brasil*

O que está ficando de fora

Para entender o que esses termos deixam de fora, consideremos “Lynda”, uma pessoa fictícia baseada nas muitas que acompanho online. Lynda acredita veementemente que as vacinas são perigosas. Ela vasculha bancos de dados em busca de pesquisas científicas recém-publicadas, acompanha audiências regulatórias para aprovações de vacinas, lê bulas de vacinas para analisar ingredientes e advertências. Em seguida, ela compartilha o que aprendeu com sua comunidade online.

Ela é uma divulgadora de informação falsa? Não. Ela não está compartilhando por engano informações que não se preocupou em verificar. Ela dedica tempo para buscar informações.

Ela também não é uma agente de desinformação, como comumente definido. Ela não está tentando causar danos ou ficar rica. Minha sensação é que Lynda é motivada a postar porque sente uma necessidade irresistível de alertar as pessoas sobre um sistema de saúde que ela acredita sinceramente que a prejudicou ou prejudicou a um ente querido. Ela está escolhendo informações estrategicamente para se conectar com as pessoas e promover uma visão de mundo. Seus critérios para escolher o que postar dependem menos de fazer sentido racionalmente e mais de suas identidades sociais e afinidades.

Desconsiderar Lynda por sua interpretação seletiva e falta de credenciais de pesquisa gera o risco de não percebermos o que ela está conseguindo fazer: coletar trechos de conteúdo ou clipes que apoiam seus sistemas de crenças de informações, publicadas por instituições autorizadas (o que pode ser apenas uma fala de um cientista admitindo que mais pesquisas [sobre vacinas] são necessárias, ou uma advertência sobre efeitos colaterais conhecidos) e compartilhar isso sem oferecer qualquer contexto ou explicação mais ampla. Essa informação “precisa” que ela descobriu por meio de sua própria pesquisa é usada para apoiar narrativas imprecisas – como por exemplo que os governos estejam lançando vacinas para controle populacional, ou que os médicos estejam sendo enganados ou vendidos às empresas farmacêuticas.

Para entender o ecossistema de informação contemporâneo, os pesquisadores precisam se afastar da fixação na precisão do detalhe e ampliar a percepção para entender as características de alguns desses espaços online que são alimentados pela necessidade das pessoas por conexão, comunidade e afirmação. Como escreveu a estudiosa da comunicação Alice Marwick , “Nos ambientes sociais, as pessoas não estão necessariamente procurando informar os outros: elas compartilham histórias (e fotos e vídeos) para se expressar e divulgar sua identidade, afiliações, valores e normas”.[6] Essa motivação pode aplicar-se aos fãs dos Beatles, bem como aos amantes de gatos, ativistas pela justiça social ou promotores de várias teorias da conspiração.

Pesquisa isolada

O mundo online de Lynda aponta para outra coisa que os rótulos “informação errada” e “desinformação" não conseguem capturar: conexões. Embora Lynda possa postar principalmente em grupos anti-vacinas no Facebook, se eu seguir suas atividades, é muito provável que também a encontre postando em *#stopthesteal*[7] ou grupos

[6]Ver *https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Marwick-pp-474-512.pdf*

[7]NT: Movimento de desinformação criado pelo operador político republicano Roger Stone em 2016, em antecipação a possíveis perdas eleitorais futuras que poderiam ser retratadas como roubadas por suposta fraude. Ver em

semelhantes e compartilhando no Instagram memes de negação das mudanças climáticas ou teorias da conspiração sobre o último tiroteio em massa[8]. Mas isso é um grande "se"; ninguém espera que eu, como pesquisadora, faça perguntas tão amplas.

Um dos desafios de estudar essa arena é o fato de que manter um foco estreito faz com que o papel das Lyndas do mundo seja mal compreendido. Um crescente conjunto de pesquisas aponta para o volume de conteúdo on-line problemático que pode ser rastreado até um número surpreendentemente pequeno[9] dos chamados superespalhadores[10] - mas até agora, mesmo esse trabalho estuda apenas aqueles que amplificam o conteúdo[11] dentro de um determinado tópico, e não aqueles que criam os conteúdos – o que faz com que os impactos de verdadeiros crentes devotados como Lynda ainda sejam pouco estudados.

Isso reflete um problema maior. Aqueles de nós que são financiados para rastrear informações danosas on-line muitas vezes trabalham em silos. Trabalho em uma escola de saúde pública, então as pessoas acham que eu deveria apenas estudar a desinformação sobre saúde. Meus colegas nos departamentos de ciência política são financiados para investigar o discurso que pode corroer a democracia. Eu suspeito que pessoas como Lynda navegam sobre uma quantidade enorme de conteúdo problemático de amplo alcance, mas elas não operam da maneira que nós, acadêmicos, somos configurados para pensar sobre nossos sistemas de informação falhos.

Todos os meses, há conferências acadêmicas e políticas focadas em desinformação sobre saúde, desinformação política, comunicação climática ou desinformação vinda da Rússia ou da Ucrânia. Freqüentemente, cada um destes eventos tem especialistas muito diferentes falando sobre problemas idênticos com pouca consciência da produção de conhecimento de outras disciplinas. Agências de financiamento e formuladores de políticas criam inadvertidamente ainda mais silos ao se concentrarem em nações ou regiões distintas, como a União Europeia.

Eventos e incidentes também se tornam silos. Os financiadores se concentram em eventos datados, de alto nível, como uma eleição, o lançamento de uma nova vacina ou a próxima conferência das Nações Unidas sobre mudanças climáticas. Mas aqueles que tentam manipular, monetizar, recrutar ou inspirar pessoas se destacam em explorar momentos de tensão ou indignação, quer seja o último documentário sobre a realeza britânica, um julgamento de divórcio de uma celebridade ou a Copa do Mundo. Ninguém financia investigações sobre a atividade online que esses momentos geram, embora isso possa trazer informações cruciais.

As respostas das autoridades também são isoladas. Em novembro de 2020, minha equipe publicou um relatório sobre 20 milhões de postagens que coletamos no Instagram, Twitter e Facebook, incluindo conversas sobre as vacinas COVID-19.[12] (Observe que não pretendíamos coletar postagens contendo desinformação; simplesmente queríamos saber como as pessoas estavam falando sobre as vacinas.) A

https://en.wikipedia.org/wiki/Attempts_to_overturn_the_2020_United_States_presidential_election

[8]Ver *https://www.nytimes.com/2021/03/26/us/far-right-extremism-anti-vaccine.html*

[9]Ver *https://counterhate.com/research/the-disinformation-dozen/*

[10]Ver *https://arxiv.org/pdf/2207.09524.pdf*

[11]Ver *https://dl.acm.org/doi/10.1145/3577213*

[12]Ver *https://firstdraftnews.org/long-form-article/under-the-surface-covid-19-vaccine-narratives-misinformation-and-data-deficits-on-social-media/*

partir desse grande conjunto de dados, a equipe identificou várias narrativas importantes, incluindo segurança, eficácia e necessidade de se vacinar e os motivos políticos e econômicos para produzir a vacina. Mas a conversa mais frequente sobre vacinas nas três plataformas foi uma narrativa que rotulamos de "liberdade e autonomia". As pessoas eram pouco propensas a discutir a segurança das vacinas e muito mais propensas a discutir sobre a obrigação de vacinar-se ou apresentar comprovante de vacinação. No entanto, agências como os Centros de Controle e Prevenção de Doenças dos EUA estão preparadas apenas para tratar da narrativa única sobre segurança, eficácia e necessidade.

Não "átomos", mas narrativas e redes

Infelizmente, a maioria dos estudiosos que estudam e respondem a informações poluídas ainda pensa em termos do que chamo de átomos de conteúdo, e não em termos de narrativas. As plataformas de mídia social têm equipes que tomam decisões sobre se uma postagem individual deve ser verificada, rotulada, rebaixada ou removida. As plataformas tornaram-se cada vez mais hábeis em praticar arbitrariedades com postagens que podem nem mesmo violar suas diretrizes. Mas, ao se concentrar em postagens individuais, os pesquisadores não conseguem ver o quadro geral: as pessoas não são influenciadas por uma postagem tanto quanto pelas narrativas nas quais essa postagem se encaixa.

Nesse sentido, postagens individuais não são átomos, mas algo como gotas de água. É improvável que uma gota d'água persuada ou prejudique, mas, com o tempo, a repetição começa a se consolidar em narrativas abrangentes – muitas vezes, narrativas que já estão alinhadas com o pensamento das pessoas. O que acontece com a confiança do público quando as pessoas veem repetidamente, ao longo de meses e meses, postagens que "estão apenas fazendo perguntas" sobre instituições governamentais ou organizações de saúde pública? Como gotas de água em pedra, uma gota não fará mal, mas com o tempo os sulcos podem ser profundos.

O que fazer?

Nos últimos anos, tem sido muito mais fácil culpar os *trolls* russos no Facebook ou os adolescentes no *4chan* do que reconhecer como aqueles atores encarregados de fornecer informações claras e úteis para atender às necessidades das comunidades falharam sistematicamente em fazê-lo. Os maus atores que estão tentando manipular, dividir e semear o caos se aproveitaram desses vácuos. Nesse espaço confuso, instituições confiáveis não têm atuado à altura do desafio.

Para realmente avançar, os defensores de ecossistemas saudáveis de informação precisam de uma visão mais ampla e integrada de como e por que as informações circulam:

Organizar e financiar pesquisas transversais: aqueles que desejam promover ecossistemas de informação saudáveis devem aprender a avaliar fluxos de conteúdo multilíngues e em rede que ultrapassem os limites convencionais de disciplinas e regiões. Eu presidi uma força-tarefa que propunha uma instituição global permanente para monitorar e estudar informações, que seria financiada centralmente e, portanto, independente tanto de países quanto de empresas de tecnologia.[13] No momento, os

[13]Ver *https://edmo.eu/2022/06/29/10-recommendations-by-the-taskforce-on-disinformation-and-the-war-in-*

esforços para monitorar a desinformação geralmente fazem um trabalho sobreposto, mas falham em compartilhar dados e mecanismos de classificação e têm capacidade limitada de responder em uma situação de crise.

Aprender a participar: o ecossistema de informações poluídas é participativo -- um local de experimentação constante, uma vez que os participantes impulsionam engajamento e se conectam melhor com as preocupações de seu público. Embora os meios de comunicação e as agências governamentais pareçam adotar as mídias sociais, eles raramente usam os recursos interativos bidirecionais que caracterizam a Web 2.0. A comunicação científica tradicional ainda é feita de cima para baixo, com base no modelo paternalista de déficit, que presume que os especialistas sabem quais informações fornecer, e que o público consumirá passivamente as informações e responderá conforme o esperado. Esses sistemas têm muito a aprender com pessoas como Lynda sobre como se conectar com o público, em vez de fazer apresentações para ele. Um primeiro passo essencial é treinar as equipes de comunicação do governo, de organizações comunitárias, bibliotecários e jornalistas para pesquisar e ouvir as perguntas e preocupações do público.

Apoie a resiliência liderada pelas comunidades: Hoje, os financiadores globais e nacionais têm um foco exagerado em plataformas, filtros e regulamentação – ou seja, em como eliminar as "coisas ruins" em vez de como expandir as "coisas boas". Em vez de prosseguir com esses esforços ineficazes, os principais financiadores deveriam encontrar uma maneira de apoiar respostas específicas e baseadas em contextos locais que dialoguem com o que as comunidades precisam. Por exemplo, o pesquisador de saúde Stephen Thomas criou a campanha *Health Advocates In-Reach and Research* (HAIR)[14] que treina proprietários de barbearias e salões de beleza locais para ouvir seus clientes sobre problemas de saúde e, em seguida, fornecer conselhos e direcionar as pessoas aos recursos apropriados para acompanhamento e cuidados. Outro exemplo: depois de avaliar as necessidades de informação da comunidade local de língua espanhola em Oakland, Califórnia, e descobri-la terrivelmente mal atendida, a jornalista Madeleine Bair fundou o site participativo de notícias on-line *El Tímpano* em 2018.[15]

Campanhas educacionais direcionadas com análise do ciclo de vida [NdT: da informação]: Estas também podem ajudar as pessoas a aprender a navegar em sistemas de informação poluídos. Ensinar às pessoas técnicas como o método SIFT (que descreve as etapas para avaliar fontes e rastrear reivindicações em seu contexto original)[16] e leitura lateral (que ensina como verificar informações enquanto as consome)[17] tem se mostrado eficaz, assim como programas para equipar as pessoas com habilidades para entender como suas emoções são direcionadas e outras técnicas usadas por manipuladores.[18]

Para cada uma dessas tarefas, as pessoas e entidades que desejam promover ecossistemas de informação saudáveis devem se comprometer com o longo prazo. A

ukraine/#pst-1

[14]Ver *https://sph.umd.edu/hair*

[15]Ver *https://internews.org/resource/mas-informacion/*

[16]Ver *SIFT (The Four Moves)*, *https://hapgood.us/2019/06/19/sift-the-four-moves/*

[17]Ver *https://www.cip.uw.edu/2021/12/07/lateral-reading-canada-civix-study/*

[18]Ver *https://theconversation.com/inoculation-theory-using-misinformation-to-fight-misinformation-77545*

melhoria real será um processo de décadas, e muito disso será gasto tentando recuperar o atraso em um cenário tecnológico que se transforma a cada poucos meses, com disrupções como o ChatGPT surgindo aparentemente da noite para o dia. A única maneira de fazer avanços é olhar para além dos diagramas organizados e das tipologias organizadas de desinformação para ver o que realmente está acontecendo - e criar uma resposta não para o sistema de informação em si, mas para os humanos que operam nele.

Claire Wardle é cofundadora e codiretora do Information Futures Lab e professora da prática na Brown University School of Public Health.

Wardle, Claire. “Incompreensão da desinformação.” Questões de Ciência e Tecnologia 39, no. 3 (primavera de 2023): 38–40. https://doi.org/10.58875/ZAUD1691

https://www.openglobalrights.org/chatgpt-whats-left-human-rights/?lang=English

ChatGPT: O que resta de humano nos direitos humanos?

César Rodríguez-Garavito -- 25-05-2023

Abro a janela do ChatGPT e peço que escreva um ensaio de mil palavras para a Open Global Rights explicando três oportunidades e três desafios que sua inteligência artificial representa para os direitos humanos. Vacila por um momento, como se hesitasse, mas logo começa a responder com uma frase contundente: "O ChatGPT já causou um impacto significativo no mundo dos direitos humanos."

Fico impressionado com a autoconfiança dessa jovem criatura e como ela se refere a si mesma na terceira pessoa, à maneira de alguns políticos. Mas devo reconhecer o quão precisa foi a resposta. Se fosse de um aluno em um exame, com certeza eu daria uma boa nota.

Talvez essa ambigüidade – essa mistura de fascínio e estranheza, polvilhada com doses variadas de suspeita e pavor, dependendo do tema e do dia – seja o tom que domina o debate emergente sobre as implicações de direitos humanos de modelos de linguagem como o ChatGPT. É o que mostram os ensaios que inauguram a série da Open Global Rights sobre o tema: a inteligência artificial generativa pode aumentar a desinformação e as mentiras online, mas também ser uma formidável ferramenta para o exercício legal da liberdade de expressão; pode proteger ou minar os direitos de migrantes e refugiados, dependendo se é usado para monitorá-los ou para detectar padrões de abuso contra eles; e pode ser útil para grupos tradicionalmente marginalizados, mas também pode aumentar os riscos de discriminação contra a comunidade LBGBTI+, cujas identidades fluidas muitas vezes não se encaixam nas caixas algorítmicas de AIs.

Percebo uma ambigüidade semelhante na resposta do ChatGPT — o que era de se esperar, considerando que é um modelo de linguagem que sintetiza (alguns diriam que rouba)[1] ideias compartilhadas por humanos na web. Com disciplina, destaca três contribuições de sua tecnologia para os direitos humanos: maior acesso à informação, melhores traduções entre idiomas e mais análise de dados e previsão de tendências sobre temas relevantes, avanços que facilitam a investigação e denúncia de violações de direitos ao redor do mundo. E fecha com três riscos: que reproduza os enviesamentos da informação que devora, que viole a privacidade através da utilização de dados pessoais e que se preste a abusos porque "o ChatGPT pode ser difícil de compreender e analisar, tornando-o desafiador responsabilizar os responsáveis por seu uso". Note que se refere à responsabilidade dos usuários, não das corporações que criaram a tecnologia, que ainda relutam em revelar o que sabem e o que não sabem sobre como ela funciona. Talvez eles tenham programado a criança para não renegar seus pais.

Embora esses efeitos sejam importantes, acho que os analistas, tanto humanos quanto virtuais, tendem a registrar o terremoto, mas perdem de vista as placas tectônicas que estão se movendo sob a superfície. Na realidade, a IA generativa não apenas aguça os

[1] Ver *https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinating-naomi-klein*, publicado nesta edição da poliTICs.

impactos conhecidos, mas questiona as categorias básicas que dão sentido aos direitos humanos.

Basta examinar a resposta do ChatGPT para trazer algumas dessas questões à tona. “O ChatGPT tem o potencial de fornecer acesso a informações para pessoas que, de outra forma, não teriam”, diz-me como se falasse de outra pessoa. “Ao fornecer informações precisas e oportunas, o ChatGPT pode ajudar as pessoas a tomar decisões informadas e agir para proteger seus direitos.”

O que meu novo assistente virtual não menciona é que seus superpoderes são bons tanto para informação quanto para desinformação. O risco imediato da IA desregulamentada é que os mesmos atores e empresas que fabricam ou disseminam as mentiras que abalaram as democracias e os direitos humanos agora o farão em uma escala infinitamente maior, confundindo ainda mais a linha entre o que é verdadeiro ou falso. Se a esfera pública acabar inundada com textos tão impecáveis quanto falaciosos, ou imagens e vídeos tão verossímeis quanto mentirosos, a velha tática dos direitos humanos de falar a verdade ao poder também será abafada. Modelos como ChatGPT não são papagaios estocásticos, mas potenciais hackers dos códigos linguísticos humanos nos quais concebemos direitos e todas as nossas normas e crenças.[2]

O que nos leva ao outro ponto cego do debate: a transformação do humano em direitos humanos. A ênfase da discussão do ChatGPT tem sido sobre o “que”, sobre os direitos afetados pela nova tecnologia. Mas a questão mais complexa e fascinante diz respeito ao “quem”, a linha divisória entre humanos e não humanos como sujeitos de direitos.

Essa conversa já está bem estabelecida em outros campos e correntes, da filosofia da mente à teoria da informação, à cibernética, às comunicações e ao transumanismo. Um de seus analistas mais lúcidos, Meghan O'Gieblyn, resume com perspicácia o paradoxo em que nos encontramos: “à medida que a IA continua a passar por nós em benchmark após benchmark de cognição superior, reprimimos nossa ansiedade insistindo que o que distingue a verdadeira consciência é emoções, percepção, capacidade de experimentar e sentir: as qualidades, em outras palavras, que compartilhamos com os animais”.[3] Duvido que o ChatGPT possa ver tão profundamente quanto O'Gieblyn, pelo menos por enquanto.

Assim, o antropocentrismo que inspira os direitos modernos, com seu reconhecimento exclusivo dos humanos, está sendo questionado de pontos de vista muito diferentes. Enquanto alguns estão começando a propor que IAs recebam alguns direitos,[4] outros (inclusive eu)[5] têm sugerido que o círculo de direitos seja ampliado para incluir as inteligências naturais de animais, plantas, fungos e ecossistemas.

A menos que abordemos essas questões, os direitos humanos podem não ter muito a dizer a um mundo confuso e transformado não apenas pela IA, mas também pela emergência climática, mudanças geopolíticas e democracias em retrocesso. O ChatGPT, ao contrário, certamente terá muito a dizer.

---

[2]Ver *https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html*
[3]Ver *https://www.penguinrandomhouse.com/books/567075/god-human-animal-machine-by-meghan-ogieblyn/*
[4]Ver *https://link.springer.com/article/10.1007/s11948-021-00331-8*
[5]Ver *https://www.openglobalrights.org/more-than-human-rights-trees-animals-fungi/*

(*) César Rodríguez-Garavito é o editor-chefe da *OpenGlobalRights*, professor de Direito Clínico e presidente do Centro de Direitos Humanos e Justiça Global da Escola de Direito da Universidade de Nova York.

O que a Inteligência Artificial está escondendo

Microsoft e meninas vulneráveis no norte da Argentina

Tomás Balmaceda, Karina Pedace e Tobias Schleider[1]

O filme *O Mágico de Oz* 222 estreou em 1939. Um de seus atores principais foi Terry, o cachorro treinado para fazer o papel de Toto, que era então considerado "o animal mais inteligente do planeta". O assunto da inteligência animal preocupava muitos estudiosos da época, enquanto havia um interesse crescente em entender se as máquinas podiam pensar por conta própria. Tal possibilidade claramente desafiava o senso comum, que a descartava completamente, mas começou a ser contestada uma década após a estreia do filme na obra do matemático britânico Alan Turing. Durante grande parte do século XX, a ideia de que animais ou máquinas eram capaz de pensar era considerado totalmente absurdo -- muita coisa mudou desde então.

No início de 2016, o governador de Salta, na Argentina, escolheu *O Mágico de Oz* como o livro para distribuir aos alunos de sua província que estavam aprendendo a ler. As meninas descobriram no livro de L. Frank Baum que sempre há um homem por trás da "mágica". Quando se tornaram adolescentes, essa lição se estendeu a outras áreas mais concretas de suas vidas: que não é a mágica, mas os homens que estão por trás da pobreza, das promessas, das decepções e das gravidezes.

Até então, a inteligência artificial (IA) deixou de ser o teste de Turing para tornar-se a área de especialização preferida das corporações mais poderosas e influentes do mundo. Graças a aplicativos atraentes em dispositivos pessoais, como telefones celulares e plataformas de streaming, ganhou ampla popularidade.

Até alguns anos atrás, ouvíamos apenas a expressão "inteligência artificial" para nos referirmos ao HAL 9000, de *2001: Uma Odisséia no Espaço,* ou Data, o andróide de *Star Trek.* Mas hoje, poucos se surpreendem com seu uso diário. O consenso na mídia e em certa literatura acadêmica é que estamos presenciando uma das revoluções tecnológicas mais importantes da história.

No entanto, o deslumbramento inspirado por essa tecnologia – que parece saída de um conto de fadas ou de um filme de ficção científica – esconde sua verdadeira natureza: é tanto uma criação humana quanto os mecanismos que o pretenso Mágico de Oz queria que aceitassem como eventos divinos e sobrenaturais. Nas mãos do aparato estatal e das grandes corporações, a "inteligência artificial" pode ser um instrumento eficaz de controle, vigilância e dominação, e de consolidação do status quo. Isso ficou claro quando a gigante do software Microsoft aliou-se ao

[1]Artigo publicado em *State of Power 2023 – Digital Power*, Transnational Institute, fevereiro de 2022.

governo de Salta, prometendo que um algoritmo poderia ser a solução para a crise de evasão escolar e gravidez na adolescência naquela região da Argentina.

Algoritmos que preveem a gravidez na adolescência

Um ano depois de distribuir exemplares de *O Mágico de Oz* às escolas de sua província, o governador de Salta, Juan Manuel Urtubey, anunciou um acordo com a subsidiária nacional da Microsoft para implementar uma plataforma de IA projetada para evitar o que descreveu como "um dos problemas mais urgentes" da região. Ele estava se referindo ao aumento de casos de gravidez na adolescência. Segundo as estatísticas oficiais, em 2017, mais de 18% de todos os nascimentos registados na província foram de jovens com menos de 19 anos: 4.914 crianças, um ritmo superior a 13 por dia.

Ao promover sua iniciativa, o governador declarou: "Estamos lançando um programa para prevenir a gravidez na adolescência usando inteligência artificial com a ajuda de uma empresa de software de renome mundial. Com essa tecnologia, você pode prever com cinco ou seis anos de antecedência -- com o primeiro nome, o sobrenome e o endereço -- qual menina tem 86% de probabilidade de ter uma gravidez na adolescência'.

Com quase o mesmo alarde das saudações do Mágico de Oz aos visitantes que encontraram seu caminho ao longo da estrada de tijolos amarelos, a Microsoft anunciou o acordo, chamando-o de "iniciativa inovadora, única no país e um passo importante no processo de transformação digital da província".

Um terceiro integrante da aliança entre a gigante da tecnologia e o governo foi a Fundação CONIN, presidida por Abel Albino, médico e ativista que lutou contra a legalização do aborto e do uso de preservativos. Essa aliança revela os motivos políticos, econômicos e culturais por trás do programa: o objetivo era consolidar o conceito de "família" em que o sexo e os corpos das mulheres são destinados à reprodução -- supostamente o propósito último e sagrado que deve ser protegido a todo custo.

Essa conhecida visão conservadora existe há séculos na América Latina, mas aqui foi vestida com roupas de cores vivas graças à cumplicidade de uma corporação norte-americana (Microsoft) e ao uso de termos como "inteligência artificial" que aparentemente foram suficientes para garantir eficácia e modernidade.

Os anúncios também forneceram informações sobre algumas das metodologias a serem utilizadas. Por exemplo, eles disseram que os dados básicos "serão enviados voluntariamente pelos indivíduos" e permitirão que o programa "trabalhe para prevenir a gravidez na adolescência e o abandono escolar; algoritmos inteligentes são capazes de identificar características pessoais que tendem a levar a alguns desses problemas e alertar o governo". O Coordenador de Tecnologia do Ministério da Primeira Infância da Província de Salta, Pablo Abeleira, declarou que "no nível

tecnológico, o nível de precisão do modelo que estamos desenvolvendo foi próximo a 90%, de acordo com um teste piloto realizado na cidade de Salta".

O que está por trás dessas reivindicações?

O mito da inteligência artificial objetiva e neutra

A IA já foi incorporada não apenas ao discurso público, mas também em nossas vidas diárias. Às vezes parece que todos sabem o que queremos dizer com "inteligência artificial". No entanto, este termo não deixa de ser ambíguo, não apenas porque é geralmente usado como um guarda-chuva sob o qual aparecem conceitos muito semelhantes e relacionados -- mas não sinônimos -- como "aprendizado de máquina', "aprendizagem profunda" ou computação cognitiva, entre outros – mas também porque uma análise mais detalhada revela que o próprio conceito de inteligência neste contexto é controverso.

Neste ensaio, usaremos IA para nos referirmos a modelos ou sistemas de algoritmos que podem processar grandes volumes de informações e dados enquanto "aprendem" e aprimoram sua capacidade de realizar tarefas além das que foram originalmente programados para fazer. Um caso de IA, por exemplo, é um algoritmo que, após processar centenas de milhares de fotos de gatos, consegue extrair o que precisa para reconhecer um gato em uma nova foto, sem confundi-lo com um brinquedo ou almofada. Quanto mais fotografias lhe derem, mais aprenderá e menos erros cometerá.

Esses desenvolvimentos em IA estão se espalhando pelo mundo e já são usados em tecnologias cotidianas, como reconhecimento de voz de assistentes digitais como Siri e Alexa, bem como em projetos mais ambiciosos, como carros autônomos ou testes para detecção precoce de câncer e outras doenças.

Existe uma gama muito ampla de usos para essas inovações, o que afeta muitas indústrias e setores da sociedade. Na economia, por exemplo, algoritmos prometem identificar os melhores investimentos na bolsa de valores. Na arena política, houve campanhas de mídia social a favor ou contra um candidato eleitoral que foram projetadas para atrair diferentes indivíduos com base em suas preferências e uso da Internet. Em relação à cultura, as plataformas de streaming utilizam algoritmos para oferecer recomendações personalizadas de séries, filmes ou músicas.

O sucesso desses usos da tecnologia e as promessas de benefícios que até recentemente existiam apenas na ficção científica, inflaram a percepção do que a IA realmente é capaz de fazer. Hoje, é amplamente considerada como a epítome da atividade racional, livre de preconceitos, paixões e erros humanos.

Isso, porém, é apenas um mito. Não existe "IA objetiva" ou IA que não seja contaminada por valores humanos. Nossa condição humana – talvez humana demais – inevitavelmente terá um impacto na tecnologia. Uma maneira de deixar isso claro é remover alguns dos véus que escondem um termo como "algoritmo".

A filosofia da tecnologia nos permite distinguir pelo menos duas maneiras de defini-la em termos conceituais. Em sentido estrito, um algoritmo é uma *construção matemática* que é selecionada por causa de sua eficácia anterior na resolução de problemas semelhantes aos que serão resolvidos agora (como redes neurais profundas, redes bayesianas ou "cadeias de Markov"). Em sentido *amplo* , um algoritmo é *todo um sistema tecnológico* composto por várias entradas, como dados de treinamento, que produz um modelo estatístico projetado, montado e implementado para resolver uma questão prática predefinida.

Tudo começa com uma compreensão simplista dos dados. Os dados emergem de um processo de seleção e abstração e, conseqüentemente, nunca podem oferecer uma descrição objetiva do mundo. Os dados são inevitavelmente parciais e tendenciosos, pois são o resultado de decisões e escolhas humanas, como incluir certos atributos e excluir outros. O mesmo acontece com a noção de previsão baseada em banco de dados. Uma questão fundamental para o uso governamental da ciência baseada em dados em geral e do aprendizado de máquina em particular é decidir o que medir e como medir com base em uma definição do problema a ser abordado, o que leva à escolha do algoritmo, no sentido estrito, que é considerado mais eficiente para a tarefa, não importa quão mortais sejam as consequências. A contribuição humana é portanto crucial para determinar qual problema resolver.

Fica assim claro que existe uma ligação inextricável entre a IA e uma série de decisões humanas. Embora o aprendizado de máquina ofereça a vantagem de processar um grande volume de dados rapidamente e a capacidade de identificar padrões nos dados, há muitas situações em que a supervisão humana não é apenas possível, mas necessária.

Puxando a cortina da IA

Quando Dorothy, o Homem de Lata, o Leão e o Espantalho finalmente conheceram o Mágico de Oz, ficaram fascinados com a voz profunda e sobrenatural desse ser que, na versão cinematográfica de 1939, foi interpretado por Frank Morgan e apareceu em um altar atrás um misterioso fogo e fumaça. Porém, Totó, o cachorro de Dorothy, não ficou tão impressionado e abriu a cortina, expondo a farsa: havia alguém manipulando um conjunto de alavancas e botões e comandando tudo no palco. Assustado e constrangido, o pretenso bruxo tentou manter a farsa: "Não dêem atenção ao homem atrás da cortina!" Mas, quando encurralado pelos outros personagens, ele foi forçado a admitir que era tudo uma farsa. "Sou apenas um homem comum", confessou a Dorothy e seus amigos. O Espantalho, porém, corrigiu-o imediatamente: "Você é mais do que isso. Você é um embuste".

Quando tiramos as roupas e vestidos extravagantes, vemos a IA como ela realmente é: um produto da ação humana que carrega as marcas de seus criadores. Às vezes, seus processos são vistos como semelhantes ao pensamento humano, mas são tratados como isentos de erros ou preconceitos. Diante da retórica generalizada e

persuasiva sobre sua neutralidade de valor e a objetividade que a acompanha, devemos analisar a inevitável influência dos interesses humanos em vários estágios dessa tecnologia supostamente "mágica".

A promessa da Microsoft e do governo de Salta de prever "com cinco ou seis anos de antecedência, com nomes, sobrenomes e endereços, qual menina ou futura adolescente tem 86% de probabilidade de ter uma gravidez na adolescência" acabou sendo uma promessa vazia.

O fiasco começou com os dados: eles usaram um banco de dados coletado pelo governo provincial e organizações da sociedade civil em bairros de baixa renda da capital provincial em 2016 e 2017. A pesquisa atingiu pouco menos de 300.000 pessoas, das quais 12.692 eram meninas e adolescentes entre 10 e 19 anos. No caso dos menores, a informação foi recolhida após obtenção do consentimento do "chefe de família" (sic).

Esses dados foram inseridos em um modelo de aprendizado de máquina que, de acordo com seus implementadores, é capaz de prever com precisão cada vez maior quais meninas e adolescentes ficarão grávidas no futuro. Isso é um absurdo absoluto: a Microsoft estava vendendo um sistema que prometia algo que é tecnicamente impossível de alcançar. Ela recebeu uma lista de adolescentes com probabilidade de gravidez. Longe de decretar qualquer política, os algoritmos forneceram informações ao Ministério da Primeira Infância para que ele pudesse lidar com os casos identificados.

O governo de Salta não especificou o que implicaria sua abordagem, nem os protocolos utilizados, as atividades de acompanhamento planejadas, o impacto das medidas aplicadas -- se é que o impacto foi medido de alguma forma -- os critérios de seleção para as organizações não governamentais ou fundações envolvidas, nem o papel da Igreja Católica.

O projeto também teve grandes falhas técnicas: uma investigação da World Web Foundation relatou que não havia informações disponíveis sobre os bancos de dados usados, as premissas que sustentam o projeto dos modelos ou sobre os modelos finais projetados, revelando a opacidade do processo . Além disso, constatou que a iniciativa falhou em avaliar as desigualdades potenciais e não prestou atenção especial a grupos minoritários ou vulneráveis que poderiam ser afetados. Também não considerou as dificuldades de trabalhar com uma faixa etária tão ampla na pesquisa e o risco de discriminação ou mesmo criminalização.

Os especialistas concordaram que os dados da avaliação foram levemente contaminados, pois os dados usados para avaliar o sistema eram os mesmos usados para treiná-lo. Além disso, os dados não eram adequados para o propósito declarado. Eles foram retirados de uma pesquisa com adolescentes residentes na província de Salta que solicitou informações pessoais (idade, etnia, país de origem etc) e se já estiveram ou estavam grávidas. No entanto, a pergunta que eles estavam

tentando responder com base nessas informações *atuais* era se uma adolescente poderia engravidar *no futuro*, algo que parecia mais uma premonição do que uma previsão. Além disso, a informação foi tendenciosa, porque os dados sobre gravidez na adolescência tendem a ser incompletos ou ocultados, dada a natureza inerentemente sensível desse tipo de informação.

Pesquisadores do Laboratório de Inteligência Artificial Aplicada do Instituto de Ciências da Computação da Universidade de Buenos Aires constataram que, além do uso de dados não confiáveis, houve graves erros metodológicos na iniciativa da Microsoft. Além disso, eles também alertaram sobre o risco de adoção de medidas equivocadas pelos formuladores de políticas: "As técnicas de inteligência artificial são poderosas e exigem que aqueles que as utilizam ajam com responsabilidade; são apenas mais uma ferramenta, que deve ser complementada por outras, e de forma alguma substituem o conhecimento ou a inteligência de um especialista", especialmente em uma área tão sensível como a saúde pública e setores vulneráveis.

E isso levanta a questão mais séria no centro do conflito: mesmo que fosse possível *prever* a gravidez na adolescência (o que parece improvável), não está claro *para que* isso serviria. Falta *prevenção em todo o processo.* O resultado, no entanto, foi criar um risco inevitavelmente alto de estigmatizar meninas e adolescentes.

IA como instrumento de poder sobre populações vulneráveis

Desde o início, a aliança entre a Microsoft, o governo de Salta e a Fundação CONIN foi baseada em pressupostos preconcebidos que não só são questionáveis, mas também conflitantes com princípios e normas consagrados na Constituição Argentina e nas convenções internacionais incorporadas ao sistema nacional . Baseia-se inquestionavelmente na ideia de que a gravidez (infantil ou adolescente) é um desastre e, em alguns casos, a única forma de preveni-la é através de intervenções diretas. Essa premissa está ligada a uma postura muito vaga sobre a atribuição de responsabilidades.

Por um lado, aqueles que planejaram e desenvolveram o sistema parecem ver a gravidez como algo pelo qual ninguém é responsável. Mas, por outro lado, atribuem a responsabilidade exclusivamente às meninas e adolescentes grávidas. De qualquer forma, essa ambiguidade contribui, antes de tudo, para a objetificação das pessoas envolvidas e também invisibiliza aqueles que são de fato responsáveis: principalmente os homens (ou adolescentes ou meninos, mas principalmente homens) que obviamente contribuíram para a gravidez (pessoas costumam dizer, com um tom grosseiro e eufemístico, que a menina ou adolescente "ficou grávida"). Em segundo lugar, ignora o fato de que, na maioria dos casos de gravidez entre mulheres jovens e em *todos* os casos de gravidez entre meninas, não só é errado presumir que a menina ou adolescente consentiu em relações sexuais, mas essa suposição deve ser completamente descartada.

Em suma, essa postura ambígua obscurece o fato crucial de que todas as gestações de meninas e muitas gestações de mulheres jovens são resultado de estupro.

No que diz respeito ao aspecto mais negligenciado do sistema – ou seja, a previsão da taxa de abandono escolar – assume-se (e conclui-se) que uma gravidez levará inevitavelmente um aluno a abandonar a escola. Embora o custo de oportunidade que a gravidez precoce e a maternidade impõem às mulheres nunca deva ser ignorado, a interrupção ou abandono da educação formal não é inevitável. Existem exemplos de programas e políticas inclusivas que têm sido eficazes para ajudar a evitar ou reduzir as taxas de abandono escolar.

De uma perspectiva mais ampla, o sistema e seus usos afetam direitos que se enquadram em um espectro de direitos sexuais e reprodutivos, que são considerados direitos humanos. A sexualidade é uma parte central do desenvolvimento humano, independentemente de os indivíduos optarem por ter filhos. No caso dos menores, importa ter em conta as diferenças nas suas capacidades evolutivas, tendo em conta que a orientação dos pais ou tutores deve sempre privilegiar a capacidade de exercício dos direitos por conta própria e em benefício próprio . Os direitos sexuais, em particular, implicam considerações específicas. Por exemplo, é essencial respeitar as circunstâncias particulares de cada menina, menino ou adolescente, seu nível de compreensão e maturidade, saúde física e mental, relacionamento com vários membros da família e, finalmente, a situação imediata que eles enfrentam.

O uso da IA tem impactos concretos nos direitos de meninas e adolescentes (potencialmente) grávidas. Primeiro, o direito à autonomia pessoal das meninas e adolescentes foi violado. Já mencionamos a sua objetificação e a indiferença do projeto face aos seus interesses individuais na busca de um suposto interesse geral. As meninas e adolescentes sequer eram consideradas titulares de direitos e seus desejos ou preferências individuais eram completamente ignorados.

Nesse projeto da Microsoft, a IA foi usada como instrumento para gerar poder sobre meninas e adolescentes, que foram catalogadas sem seu consentimento (ou seu conhecimento, aparentemente). Segundo os promotores do sistema, as entrevistas eram feitas com os "chefes de família" (principalmente seus pais) sem ao menos convidá-los a participar. Além disso, os questionários incluíam assuntos altamente pessoais (intimidade, vida sexual etc) em que seus pais raramente seriam capazes de responder em detalhes sem invadir a privacidade de suas filhas ou -- tão grave quanto -- basear-se em suposições ou preconceitos que o estado iria então assumir como verdadeiro e legítimo.

Outras violações incluem os direitos à intimidade, privacidade e liberdade de expressão ou opinião, enquanto os direitos à saúde e à educação correm o risco de serem ignorados, apesar das declarações das autoridades e da Microsoft sobre sua intenção de cuidar das meninas e adolescentes. Por fim, cabe mencionar um direito

conexo que assume particular importância no contexto específico deste projeto: o direito à liberdade de pensamento, consciência e religião.

Não iríamos ao ponto de afirmar que esse episódio teve um final feliz, como o Mágico de Oz teve. Mas o projeto da Microsoft não durou muito. Sua interrupção não foi por críticas de ativistas, mas por um motivo muito mais mundano: em 2019, foram realizadas eleições nacionais e estaduais na Argentina e Urtubey não foi reeleito. A nova administração encerrou vários programas, incluindo o uso de algoritmos para prever a gravidez, e reduziu o Ministério da Primeira Infância, Infância e Família ao status de secretaria.

O que a IA está escondendo

A fumaça retórica e os espelhos dos desenvolvimentos objetivos e neutros da IA desmoronam quando desafiados por vozes que afirmam que isso é impossível em princípio, como argumentamos na primeira seção, dada a participação de analistas humanos em vários estágios do desenvolvimento dos algoritmos . Homens e mulheres definiram o problema a ser resolvido, projetaram e prepararam os dados, determinaram quais algoritmos de aprendizado de máquina eram os mais adequados, interpretaram criticamente os resultados da análise e planejaram a ação adequada a ser tomada com base nos insights que a análise revelou.

Há insuficiente reflexão e discussão aberta sobre os efeitos indesejáveis do avanço desta tecnologia. O que parece prevalecer na sociedade é a ideia de que o uso de algoritmos em diferentes áreas garante não apenas eficiência e rapidez, mas também a não interferência de preconceitos humanos que podem "manchar" a ação imaculada dos códigos que sustentam os algoritmos.

Como resultado, as pessoas assumem que a IA foi criada para melhorar a sociedade como um todo ou, pelo menos, certos processos e produtos. Mas quase ninguém questiona o básico – para quem isso será uma melhoria, quem se beneficiará e quem avaliará as melhorias? Cidadãos? O estado? Corporações? Meninas adolescentes de Salta? Os homens adultos que abusaram delas? Em vez disso, há uma falta de consciência real sobre a escala de seu impacto social ou a necessidade de discutir se tal mudança é inevitável.

As pessoas não se surpreendem mais com as constantes notícias sobre a introdução da IA em novos campos, exceto pelo que há de novo nela, e assim como o passar do tempo, é tratada como algo que não pode ser parado ou revisitado. A crescente automação dos processos que o ser humano realizava pode gerar alarme e preocupação, mas não desperta interesse em detê-la ou refletir sobre como será o futuro do trabalho e da sociedade quando a IA assumir grande parte do nosso trabalho. Isso levanta uma série de perguntas que raramente são feitas: isso é realmente desejável? Para quais setores sociais?

Quem se beneficiará com uma maior automação e quem sairá perdendo? O que podemos esperar de um futuro onde a maioria dos trabalhos tradicionais serão executados por máquinas? Parece não haver tempo nem espaço para discutir o assunto: a automação simplesmente acontece e tudo o que podemos fazer é reclamar do mundo que perdemos ou nos maravilhar com o que ela pode alcançar hoje.

Essa complacência com os constantes avanços da tecnologia em nossas vidas privada, pública, profissional e cívica se deve à confiança na crença de que esses desenvolvimentos são "superiores" ao que pode ser alcançado pelo mero esforço humano. Assim, como a IA é muito mais poderosa, ela é "inteligente" (o rótulo "inteligente" é usado para telefones celulares, aspiradores de pó e cafeteiras, entre outros objetos que fariam Turing corar) e livre de preconceitos e intenções. No entanto, como apontado anteriormente, a própria ideia de IA de valor neutro é uma ficção. Para colocar de forma simples e clara: existem vieses em todos os estágios de projeto, teste e aplicação do algoritmo e, portanto, é muito difícil identificá-los e até mesmo mais difícil corrigi-los. No entanto, é imprescindível fazê-lo para desmascarar sua natureza supostamente estéril, desprovida de valores e erros humanos.

Uma abordagem focada nos perigos da IA, juntamente com uma postura otimista sobre seu potencial, pode levar a uma dependência excessiva da IA como solução para nossas preocupações éticas – uma abordagem em que a IA é solicitada a responder aos problemas que a IA produziu. Se os problemas forem considerados puramente tecnológicos, eles deveriam exigir apenas soluções tecnológicas. Em vez disso, temos decisões humanas vestidas com trajes tecnológicos. Precisamos de uma abordagem diferente.

O caso dos algoritmos que deveriam prever a gravidez na adolescência em Salta expõe quão irreal é a imagem da chamada objetividade e neutralidade da inteligência artificial. Como Toto, não podemos ignorar o homem por trás da cortina: o desenvolvimento de algoritmos não é neutro, mas sim baseado em uma decisão tomada a partir de muitas escolhas possíveis. Como o projeto e a funcionalidade de um algoritmo refletem os valores de seus projetistas e seus usos pretendidos, os algoritmos inevitavelmente levam a decisões tendenciosas. As decisões humanas estão envolvidas na definição do problema, na preparação e projeto dos dados, na seleção do tipo de algoritmo, na interpretação dos resultados e no planejamento das ações com base na sua análise. Sem supervisão humana qualificada e ativa, nenhum projeto de algoritmo de IA é capaz de atingir seus objetivos e ser bem-sucedido. A ciência de dados funciona melhor quando a experiência humana e o potencial dos algoritmos funcionam em conjunto.

Algoritmos de inteligência artificial não são mágicos, mas não precisam ser uma farsa, como argumentou o Espantalho. Nós apenas temos que reconhecer que eles são humanos.

Karina Pedace ensina alunos de graduação e pós-graduação na Universidade de Buenos Aires e na Universidade Nacional de Matanza, e é pesquisadora do Instituto de Pesquisas Filosóficas da Sociedade Argentina (IIF-SADAF-CONICET). Suas áreas de pesquisa atuais incluem filosofia da tecnologia, metafísica da mente e metodologias de pesquisa. É secretária executiva da Rede Latino-Americana de Mulheres Filósofas da UNESCO e cofundadora do Grupo de Pesquisa em Inteligência Artificial, Filosofia e Tecnologia (GIFT). Em 2022, ela foi reconhecida internacionalmente como uma das 100 Mulheres Brilhantes na Ética da IA: *https://womeninaiethics.org/the-list/of-2022/*

Tomás Balmaceda é doutor em Filosofia pela Universidade de Buenos Aires. Atualmente é Pesquisador do IIF (SADAF/CONICET) e faz parte do grupo GIFT que analisa a tecnologia e a inteligência artificial pelas lentes da filosofia. Autor de vários livros, seus interesses incluem a ética da influência na rede, nova longevidade e educação financeira para a população LGBTIQ+.

Tobías J. Schleider é advogado e especialista em direito penal pela Universidade Nacional de Mar de Plata e doutor pela Universidade de Buenos Aires em Filosofia do Direito. É Professor da Universidade Nacional do Sul, onde dirige a licenciatura em Segurança Pública. Suas linhas de pesquisa atuais incluem prevenção da violência apoiada pela tecnologia, teoria da ação humana, causalidade e influência da sorte na atribuição de responsabilidades.