1001 Mãos de Ferro

Uma inteligência antiga, opaca e Artificial

Carolina Monteiro

1001 Mãos de Ferro

Uma inteligência antiga, opaca e Artificial

Ficha de identificação da obra elaborada pela autora,
através do Programa de Geração Automática da
Biblioteca Universitária da UFSC:

Monteiro Alves, Carolina

1001 Mãos de Ferro : uma inteligência antiga, opaca e Artificial / Carolina Monteiro Alves ; orientadora, Valentina da Silva Nunes, 2024.

122 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro de Comunicação e Expressão, Graduação em Jornalismo, Florianópolis, 2024.

Inclui referências.

1. Jornalismo. 2. Inteligência Artificial. 3. Violação de Direitos. 4. Transparência. 5. Desigualdade. I. da Silva Nunes, Valentina. II. Universidade Federal de Santa Catarina. Graduação em Jornalismo. III. Título.

Todos os direitos reservados.

É proibida a reprodução total ou parcial desta obra, por qualquer meio ou processo, sem a autorização prévia e por escrito da autora.

Ficha Técnica

Texto Carolina Monteiro

Ilustração Victor Hugo

Projeto gráfico e diagramação Carolina Monteiro

Revisão Valentina da Silva Nunes

Capa Carolina Monteiro

Florianópolis 2024

Este livro é um Trabalho de Conclusão de Curso em Jornalismo da Universidade Federal de Santa Catarina. Foi aprovado em 11 de dezembro de 2024 pelos professores Dr. Carlos Augusto Locatelli, Drª Daisi Irmgard Vogel e Drª Valentina da Silva Nunes.

Desdedico este livro à Google.

À Meta, Amazon e Uber. Ao Ifood e TikTok.

À OpenAI, Microsoft, Intel, IBM e Apple.

À finada Lionbridge, Appen e Telus.

Sama.

Ao DataForce e Salesforce.

À TaskUs, ScaleAI, Teemwork.ai,

Mighty AI, CloudFactory, Figure Eight, DataLabel,

Cortexica, Labelbox, Xiilar, Deepgram, Zindi, V7 Labs

Muito obrigada,

ainda que apenas o meu nome esteja no rótulo de autoria, há muito o que agradecer às pessoas que me deram suporte material, mental e espiritual para este trabalho ser realizado. Sou grata e ciente de todo o apoio que a minha avó, Anna Maria Bernardino, e minhas tias, Lúcia Araújo e Neumana Monteiro, me deram, enchendo o meu espírito de força e a minha mesa de sustança para que eu não precisasse me ocupar com outras coisas.

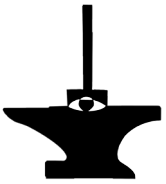
Agradeço à agência de fomento por financiar este trabalho: Angelina Monteiro, minha mãe. E ao professor geólogo Francisco Rubens Alves, meu pai, em memória, por encorajar a decisão que eu tive de escrever sobre um tema o qual muito pouco eu sabia, quase nada, menos que o bê-a-bá. Uma pena ele não ver o resultado final. Mas é um folgado, poderia ter ficado para ajudar no processo… Paciência.

Algumas das fontes com quem tratei foram totalmente fundamentais para a execução do trabalho não apenas pelas suas entrevistas, mas pelo encorajamento e por todas as pistas que recebi delas. Por isso digo que apesar de meu nome figurar como autora, não teria sido possível sem um grande e coletivo trabalho, de uma rede de pessoas que talvez não se conheçam mas que contribuíram para um mesmo propósito.

Lembro bem que as jornalistas Mayara Heloisa Santos e Erika Artmann não me deixaram vacilar. Aos historiadores Drª Ângela Balladares, Leonardo Rossi e Eliezer da Rosa Jr. por me apoiar a desenvolver minha resiliência e coragem. À Karine Joulie, Loiva Knuth, Pedro Porto, Klay Silva e os profs. Dr. Ildo Golfetto e Drª Isabel Coelho, que muito aliviaram as dores deste ano de ~~terror~~ 2024.

Obrigada também ao Dr. Julio Marinho Ferreira, prof. Dr. Ruhan Conceição, Matheus Alves e Heloísa Faria, que mostraram como o tema era necessário. A Gabriel Santana Martins, Dr., por me mostrar uma visão geral do trabalho que eu mesma não via. Aos engenheiros que me ensinaram uma outra forma de pensar, no projeto de extensão Vento Sul do Centro Tecnológico da UFSC, que serviu como QG para o meu processo de escrita. Agradeço a Thamires Alves, também engenheira, por encorajar as mulheres na ciência e tecnologia.

Agradeço ao ilustrador, Victor Hugo, por se empolgar com este junto trabalho comigo. E, por fim, pela paciência e suporte da orientadora, Drª Valentina da Silva Nunes. A professora foi um alívio para um ano cruel do meu universo pessoal. Obrigada por ter a delicadeza de guiar o meu trabalho de forma que este projeto foi uma delícia de viver.



Sumário

Antiga, opaca e Artificial

Este livro se propõe a refletir sobre a transparência (e *accountability*), ou a falta dela, na Inteligência Artificial (I.A.) em duas esferas: aquela interna, sobre que base se desenvolve a cada dia; e aquela externa, que está em contato com o mundo, agindo sobre ele. A primeira seção trata da estrutura dessa indústria, que se baseia em uma legião de trabalhadores invisíveis, trabalhando nas suas casas, ocultados por contratos de sigilo. A segunda seção trata de como ela funciona à base de dados, e problemas potencializados pela tecnologia na cultura, economia, segurança e na morte. As duas partes concluem um ciclo de reflexão sobre o atual estado dessa criação humana que parece uma fantasia, mas já é muito antiga, e que ainda permanece opaca.

Estamos situados num contexto em que esta é uma nova indústria, que vem de uma lógica capitalista bastante agressiva, e em que as ferramentas são orientadas por interesses de atores dominantes em uma sociedade desigual; a regulação ainda é esparsa e a ética é um item de luxo. A premissa mais fundamental sobre a tecnologia de Inteligência Artificial é que esta é uma revolução tecnológica irrevogável e muito veloz, mas que, ao mesmo tempo, sempre foi um desejo humano: a ideia é antiga, mas o produto é novo; no entanto, os problemas que ela não supera também são antigos, trazendo uma marca de desigualdade. Por isso, o livro traz a questão de quão antiga é a I.A., ainda que pareça muito recente, e lembramos de dois mitos, com

os quais fazemos um paralelo para reflexão nas duas seções, os gigantes Talos (dos gregos) e Golem (dos judeus).

A tecnologia digital tem um trabalho a ser realizado na sociedade que é fundamental, nos baseamos nela em todos os setores de atividade. A I.A. vem como uma revolução que exponencia o poder da computação tradicional – numa comparação rude, podemos lembrar como foi o impacto da adoção das máquinas a vapor para a produção manual. Hoje, por exemplo, a melhora dos sistemas de saúde, gestão e de sinalização através da I.A. são aspectos que afetam diretamente e imediatamente o dia a dia das pessoas. Sem a I.A., já não teríamos condições de atingir os mesmos resultados contando apenas com a computação tradicional e as mãos humanas.

Usar o termo "mãos de ferro" faz discutir sobre o toque metálico e sem misericórdia que os sistemas de I.A. dão às ferramentas a que estão envolvidos – ainda que permaneça o questionamento se ela é mais metálica ou se é mais humana. A expressão "mil e uma" é por sua versatilidade e grande campo de aplicação. O alcance é muito amplo, e não é possível estar imune ao seu toque sem se isolar da sociedade. Ainda que alguém busque evitá-la no seu domínio pessoal, as cidades, estados e países têm sistemas inteiros dependentes dessa tecnologia. É como tentar estar avesso à própria eletricidade: requer um grande esforço e adoção de um modo de vida restritivo.

Urge a necessidade de pensar de forma crítica como é a I.A. que queremos e podemos ter – antes que exista um momento tal do seu desenvolvimento em que já ninguém seja capaz de mudar seu curso.

1001 Mãos de Ferro

Ainda que todo o resto do corpo e dos membros fosse brônzeo e invulnerável, havia abaixo do tendão, no calcanhar, uma veia sanguínea. Sobre ela, uma fina membrana detinha os limites da vida e da morte.

[...]

E míseros teriam partido para longe de Creta, abatidos por conta da sede e das dores, se Medeia não tivesse lhes falado ao se distanciarem:

"Escutai-me, pois creio que sozinha poderei derrotar este homem, seja ele quem for, mesmo se todo o seu corpo for brônzeo, desde que não possua uma existência inesgotável."

Argonáuticas. **Versos 1645 - 1651. Apolônio de Rodes.**

Seção I

O sangue de Talos

Talos, o gigante de bronze, foi construído por Hefesto, o deus grego da tecnologia. Seu propósito era defender a ilha de Creta dos inimigos que, porventura, tentassem se aproximar. Em uma das suas três voltas diárias que fazia em torno da ilha, avistou Jasão e os argonautas chegando para atracar sem permissão. No navio, estava Medeia, a maga.

Medeia, com feitiços, o enlouqueceu, e ele, nas suas ilusões, feriu com uma rocha seus próprios pés. O gigante possuía ali, discreta e vital, uma veia. Era como um detalhe no calcanhar, mas corria nela a sua vida, por um fluído de chumbo. Tal como ele, a tecnologia de Inteligência Artificial (I.A.) tem uma veia de vida que, apesar de ser muito humana, é camuflada pelo aspecto metálico de automatização.

Quem faz a aparente "automatização" acontecer é um corpo de trabalho de carne e osso, não são códigos. Procurando a primeira vez em que o termo "Inteligência Artificial" aparece nas notícias do acervo online da Folha de S. Paulo, dá para ver que quanto menos a I.A. demonstra a que veio no mundo e no

Brasil, maior a quantidade de artigos criticando-a. Havia uma forte dissociação entre ela e o humano.

Em fevereiro de 1994, há no jornal uma primeira reportagem que utiliza especificamente o termo, vendendo um *chatbot* antidepressivo desenvolvido na Califórnia, no artigo "Soft dá ajuda para enfrentar depressão"[1]. Até 1997, havia mais textos com teor crítico do que posteriormente. Parte desses artigos críticos propõe que a nova tecnologia poderia ser perigosa para a sociedade se valores éticos não fossem estabelecidos naquele momento para guiar seu desenvolvimento futuro.

A importância de 1997 é que, pela primeira vez, o homem foi derrotado por uma máquina. "Blue Deep", um programa de computador, venceu um dos maiores enxadristas conhecidos na história, Garry Kasparov, após duas rodadas de três jogos que terminaram em 11 de maio de 1997[2]. A máquina foi alimentada com a experiência de inúmeros jogadores, mas Kasparov possuía apenas a dele.

A partir desse fato, abre-se mais espaço naquele jornal para uma discussão sobre como funciona a Inteligência Artificial e a ciência cognitiva, com mais artigos sobre fatos e produtos da área, ou que educam de forma introdutória sobre o tema. Durante 2008, por exemplo, o mesmo jornal falou apenas de

1 Soft dá ajuda para enfrentar depressão. Folha de S. Paulo, 1994. https://www1.folha.uol.com.br/fsp/1994/2/02/informatica/16.html Acesso em nov. 2024

2 Kasparov e o computador. Folha de S. Paulo, 1997. https://www1.folha.uol.com.br/fsp/1997/5/13/opiniao/3.html Acesso em nov. 2024

produtos novos da área e projetos que estavam em desenvolvimento, mas não há nenhum artigo com objetivo específico de crítica ou alerta.

Com o tempo, mais produtos são lançados, e a I.A. vai se integrando a todo tipo de serviço e setor da atividade humana. Mas, ainda assim, a base do trabalho é uma só: viva, orgânica, feita por pessoas – que podem ou não ter um dia bom, ficar doentes, ter suas famílias, ter seus interesses pessoais, enfrentar todo tipo de problema e de prazer que é estar vivo.

Quem treina a Inteligência Artificial?

"Ficava no computador o dia todo", conta João Victor (de quem mudamos o nome), que trabalhou por cinco anos, em *home office*, como *data worker* — ou operário de dados, um "chão de fábrica" da indústria de Inteligência Artificial, produzindo dados para treinamento de I.A. Os sistemas que são treinados dessa forma aprendem com milhares de microtarefas repetitivas de avaliação, ensino e classificação de conteúdo. Na verdade, quando, em um captcha, você diz para o computador qual opção é um semáforo ou qual objeto corresponde a uma silhueta, este é o microtrabalho de treinamento da I.A. (que você faz de graça).

Há I.A. usada em diversos tipos de ferramentas, que analisam, criam e melhoram. A matéria-prima com que ela trabalha pode ser qualquer tipo de informação, como imagens, sons e textos – por isso ela precisa ser treinada, "aprendendo" o que são essas coisas. Com o cruzamento de todos esses dados, um *ensemble* de vários sistemas de Inteligência Artificial operando juntos, pode revelar padrão de comportamento, seja de humanos ou de qualquer objeto, sistema ou veículo.

Ela analisa informação capturada por qualquer dispositivo de entrada, de empresas ou sistemas informáticos públicos e privados: teclados e mouses, microfones, câmeras, sensores, GPS, históricos, informações empresariais ou pessoais em

bancos de dados públicos ou privados, vendidos ou roubados, mercado financeiro, tráfego, patentes e conteúdo criativo gerado por humanos. Por isso, pode gerar um texto ou imagem, melhorar resultados de busca na internet, acertar a probabilidade de um diagnóstico, encontrar alguém na multidão, ou ainda manipular o trajeto de um veículo e calcular a posição de um alvo.

As *big techs*, empresas que dominam o setor de inovação tecnológica, desenvolvem seus produtos, experiências e código de cultura com base nesses conjuntos massivos de dados. Para que sejam utilizados em modelos de Inteligência Artificial, passam por intensos processos: dados são tratados e processados para alimentar os sistemas que os recebem, e ainda precisam ser usados para treinar e aperfeiçoar o modelo.

Esse campo de trabalho pode ser dividido em duas faixas de hierarquia, a primeira é a de topo, que inclui todos os profissionais especialistas da área de tecnologia que trabalham estudando, criando e gerindo esses projetos (*tech workers*). A faixa de base é composta por milhões de trabalhadores invisíveis em todos os países, pouco ou não especializados, que vão treinar os modelos de I.A. processando, inserindo e rotulando a imensidão de dados para aquelas ferramentas que isso seja necessário.

Essa relação entre os trabalhadores da base e as *big tech*s é intermediada por uma empresa terceirizada, especialista em tratamento de dados, que oferece uma vaga de trabalho remoto para executar pequenas tarefas no horário que você puder, pagando em dólar. São tarefas simples, mas que tomam alguma atenção: é só gravar um áudio lendo algumas frases, é só dizer se esse vídeo do YouTube pode ser recomendado junto a outros, é

só determinar quais publicidades têm relação com a pesquisa de algum usuário do buscador na internet… As tarefas têm o objetivo de determinar a imagem de um objeto, mostrar a pronúncia correta da língua nativa, dizer se um resultado de pesquisa corresponde à pergunta, se um vídeo tem afinidade com outro, censurar conteúdo sensível, entre tantas outras funções para ensinar a máquina sobre o mundo humano.

João Victor afirma que ele tinha que ver a pesquisa da pessoa, entender o que ela quer, e ali ele recebe uma página ou uma caixa de resultados de busca no Google (que tem como nome oficial Alphabet, Inc.), em geral os patrocinados, para julgar o quão adequada a resposta foi para a pergunta do usuário. No YouTube, até a localização dos anúncios precisa ser revisada para ele julgar se estão interrompendo algo no vídeo, como uma fala ou uma cena que perca a fluidez ao ser interrompida. Muitas das microtarefas que executou para outra empresa além da terceirizada do Google eram de uma concorrente que serve o Meta Platforms, Inc., em que ele verificava adequação de anúncios de páginas nas redes sociais.

Na pandemia de Covid-19, muitas pessoas em todo o mundo procuraram essa forma de trabalho para sobreviver, tendo em vista que perderam seus empregos. Segundo João, durante aquele período, diminuiu bastante o fluxo de trabalho. "Subiu muito no começo, a minha teoria é que muitas empresas foram para o online e tiveram que fazer anúncios." Só que pessoas perderam o emprego e foram atrás desses trabalhos para ficar em casa, "então aumentou a mão de obra e começou a faltar tarefas". "Muita gente foi demitida e outros foram recontratados

com um novo salário, mais baixo."

Pode ser que varie de país a país, quais são os fatores que determinam a forma e proporção do desemprego da força de trabalho ativa, e o quão desesperada ela pode estar para obter renda. Mas em qualquer um deles é essa margem de ameaça da fome, teto e necessidades que vai determinar quão predatória pode ser a abordagem das empresas de treinamento de máquinas.

Sem regulação ainda, elas exploram pessoas de qualquer país para "humanizar" seus resultados, anúncios e conteúdo em suas línguas nativas. A maioria dessas atividades têm o objetivo de engajar e atrair cliques, sempre direcionadas a fazer o usuário passar mais tempo e se envolver mais com as redes no seu tempo online: essas empresas prestam serviços com objetivo de treinar e aperfeiçoar sistemas de I.A. de outras empresas ainda maiores.

Pode-se mencionar como referência de *big data companies* que fazem essa terceirização a Sama, DataForce, Appen e a Telus (que adquiriu a Lionbridge, uma outra tradicional do meio), de muitas, dessas mil e uma mãos, que descrevem seu trabalho como coletoras e rotuladoras de dados. Ainda chegam novas, como a Uber, que anunciou a sua entrada para o mercado de treinamento de I.A. em parceria com os desenvolvedores do PokémonGo, um jogo muito popular com base em geolocalização.

Para a Lionbridge, era permitido no máximo 20 horas de trabalho por semana para as vagas de operadores de microtrabalho, como avaliação de resultados. No Brasil, já valeu USD 8,20 a hora, quando o dólar ainda estava na faixa de R$ 3,50. Em outros países, como o caso da Índia, paga-se USD 3,00/hora; no contexto da Europa, paga-se entre € 5,00 e € 20,00. Depois da compra

pela Telus em 2020, passou a pagar USD 6,00. Na readequação para incorporar, os contratados por USD 8,20 foram demitidos, mas sem impedimento de que se cadastrassem novamente para a vaga, valendo menos. A Appen pagava, até 2023, USD 5,00. Dependendo da empresa e tipo de vaga, o pagamento pode ser feito por tarefa completada, por número de áudios transcritos ou enviados, ou outros tipos de arranjos para cada natureza de mídia.

Mas se há tantas pessoas operando os dados nas suas línguas nativas, e se essa função é tão fundamental para a Inteligência Artificial, por que ouvimos falar tão pouco delas? Os operários ainda não são vistos como uma classe trabalhadora, e nem podem se organizar como tal. "Os termos de confidencialidade que muitas empresas contratantes implementam são uma barreira bem grande, além da própria estrutura do trabalho que também dificulta a organização", comenta a pesquisadora Camilla Wagner, do grupo de Inquérito dos Operários de Dados no Weizenbaum Institute.

Camilla está se referindo ao instrumento de contratação e à atomização do trabalho característicos dessas vagas. O contrato é bastante restrito quanto ao sigilo da natureza da atividade, conteúdo trabalhado e plataforma de acessos. Os trabalhadores não podem falar com outras pessoas ou com o público sobre esses contratos, nem sobre a natureza do trabalho e o que eles fizeram, viram e ouviram ao executar suas tarefas. As empresas os proíbem disso, ameaçando de quebra de contrato e possível processo legal.

Então eles não se conhecem, nem têm ambientes de traba-

lho em comum que proporcionem alguma interação. Ao contrário de motoristas de aplicativo e entregadores qu circulam pelas ruas, esses outros trabalham onde ninguém vê. "Muitos e muitas trabalham de casa, sem nenhum contato com outras pessoas da profissão e têm medo de ir contra as instruções de não falar nunca sobre o trabalho para não perder o emprego."

Há alguns grupos no Reddit e Discord, que, segundo Camilla, são importantes para o apoio mútuo: "esse diálogo que se estabelece, muitas vezes em canais privados para evitar que as empresas censurem, é fundamental para a organização". Uma iniciativa com o objetivo de pressionar em favor de uma regulamentação desse campo de trabalho - a Data Workers Union – exige o reconhecimento de todos os usuários de serviços digitais como trabalhadores que produzem riqueza, tendo em vista que produzem dados e "dados são o novo petróleo". Segundo a União, são algumas das principais beneficiadas globais por essa exploração a Amazon.com Inc., Apple Inc., Microsoft Corporation, Meta Platforms Inc., e a Alphabet Google Inc.

"As propostas de regulamentação em um estágio bastante inicial, mas em vários países existem interesse pelo tema", conta a pesquisadora. "A diretiva dos trabalhadores de plataforma que passou na União Europeia recentemente é um bom exemplo disso" - ela se refere à diretiva dos europeus, sobre todos os trabalhadores de plataformas, que tem como um dos objetivos "a introdução de medidas para facilitar a correta determinação do status de emprego deles" – o que já é um passo na direção de reconhecer essa forma de atividade como um trabalho.

A Platform Work Directive[3] aprovada pelo Parlamento Europeu estabelece, em linhas gerais, três tópicos: novas regras para corrigir o falso autoemprego; nenhum trabalhador pode ser demitido com base na decisão de algoritmos; e as plataformas estão proibidas de processar certos tipos de dados pessoais. "No Brasil, as propostas existentes também vão naquela direção", conta Camilla. "Alguns senadores estadunidenses também demonstraram interesse pelo tema do trabalho de dados, mas isso ainda não levou a propostas concretas."

Pesquisadores se referem a essa função como "trabalho fantasma". E existe toda uma estrutura para mantê-la assim. Essa mão-de-obra é invisível: não vemos essas pessoas na rua trabalhando; elas não podem falar sobre o que fazem; a empresa não vê qualquer vínculo a celebrar; e as ferramentas têm uma imagem de robô automático sem laços com a imperfeição do Homo sapiens.

No contrato, a empresa deixa explícito que não há qualquer responsabilidade da companhia com o seu prestador autônomo além de lhe prover o pagamento pelo trabalho cumprido e por fornecer instruções para executar o trabalho. Além disso, após findado, o contrato ainda tem algumas cláusulas que permanecem ativas, incluindo a de sigilo sobre tudo o que viu e fez.

Para o operário de dados, é difícil se informar sobre como

3 Parliament adopts Platform Work Directive. Release do Parlamento Europeu. https://www.europarl.europa.eu/news/en/press-room/20240419IPR20584/parliament-adopts-platform-work-directive Acesso em nov. 2024

as tarefas devem ser executadas, existe um isolamento intenso já que não é permitido falar sobre esse trabalho ou organizar fóruns (ainda que tenha comunidades no Reddit e Discord para falar sobre essas oportunidades, é de forma contida). O jeito é tentar usar um senso de média. As diretrizes não são evidentes, espera-se que cumpra cada ação em um certo tempo (que não pode ser nem muito rápido, nem muito devagar), e a única fonte de informação sobre o padrão esperado para as tarefas é o enunciado do que se pede, um "*Frequent Answers and Questions*" (Perguntas Frequentes) e raros treinamentos em PDF.

"Quando fui ver, já tinha sido demitida", conta Samira, operária de dados (de quem também mudamos o nome), "provavelmente por alguma tarefa que fiz de errado". A quebra de contrato repentina é bastante comum nesse tipo de trabalho prestado para empresas que atuam na área de treinamento de inteligências artificiais. "Não recebi advertência ou punição, recebi um aviso que estava ruim, mas às vezes eles nem explicam por quê, ou em que tarefa específica está errando… se você continua, infelizmente é demitido". E logo entra outro no lugar.

Samira diz que ficou dois anos no programa, e não teve qualquer explicação da razão do corte, nem houve correção para o que quer que estivesse fazendo fora do padrão pedido pelo contratante. Aliás, "contratante" é como a empresa se autointitula, e o trabalhador é um prestador, celebrando um "Contrato de Prestador de Serviços Autônomo" (Independent Contractor Agreement), excusada de qualquer responsabilidade ou vínculo de emprego, definindo-se como uma empresa que oferece um "extra" em trabalho remoto para qualquer lugar do mundo. A

regulação dessa forma de trabalho é exclusivamente pautada pelo próprio instrumento elaborado pela empresa.

"Observamos que as empresas criam estruturas que reforçam a atomização e anonimato, por exemplo, quando exigem que os canais para tirar dúvidas sejam anônimos." A atomização que Camilla Wagner mencionou ao falar tem dois efeitos: uma desconexão que *tech workers* e *data workers* têm entre si, e entre o que fazem na cadeia de funções, já que não sabem de onde vem e nem para onde vão o que produzem, nem qual o propósito da sua ação. Esse fenômeno também é visto até nas carreiras de meio-topo da estrutura, porque apenas os gerentes sabem como ele é produzido de ponta a ponta.

Em 2021, programadores da Unity – uma empresa que produz ferramentas para desenvolver videogames – questionaram a companhia sobre qual o propósito do que criavam. A queixa deles é a de que começavam a desconfiar do que era pedido a eles pelos gerentes dos projetos, sem saber para quê, nem ter a visão geral do produto.

"A companhia não consegue explicar por que seus empregados, que supostamente ingressaram na empresa para criar ferramentas que ajudam o trabalho de desenvolvedores de jogos, estão agora desenvolvendo tecnologias para fins militares com o dito objetivo de uso em combate", diz reportagem da Vice[4].

4 Unity Workers Question Company Ethics As It Expands From Video Games to War. Vice. https://www.vice.com/en/article/unity-workers-question-company-ethics-as-it-expands-from-video-games-to-war/ Acesso em nov. 2024

Um ano depois, a Unity assinou publicamente um contrato multimilionário para auxiliar o departamento de defesa do governo dos EUA com ferramentas de inteligência, vigilância e reconhecimento[5].

Segundo a reportagem e outras fontes, o empregado da desenvolvedora não tem como saber e nem ter controle sobre as partes que produz de um certo produto. Por exemplo, um programador pode contribuir para uma I.A. sem aplicação específica, e a companhia usar aquela parte em um contrato militar. Então, uma parte de uma ferramenta pode servir a mais propósitos do que funcionar apenas a um fim.

É uma questão difícil de ser encerrada, pois se "os dados são o novo petróleo" é de se suspeitar que, cada vez mais, haja uma crescente abrangência para a utilidade desse capital. Não foi encontrada evidência que apoie a ideia de que os dados de jogos possam ser utilizados para treinar ferramentas militares, por exemplo. Todos os desenvolvedores consultados viram uma chance muito reduzida, impossível na prática, para esse aproveitamento de dados em diferentes produtos acontecer, e apontaram duas razões principais para isso.

A primeira é que, segundo o exemplo dos jogos, cada jogo é uma criação única, seria difícil transferir dados para um ambiente ou plataforma que não seja a dele mesmo, como

5 Unity signs "multi-million dollar" contract to help U.S. government with defense. Game Developer. https://www.gamedeveloper.com/business/unity-signs-multi-million-dollar-contract-to-help-u-s-government-with-defense Acesso em nov. 2024

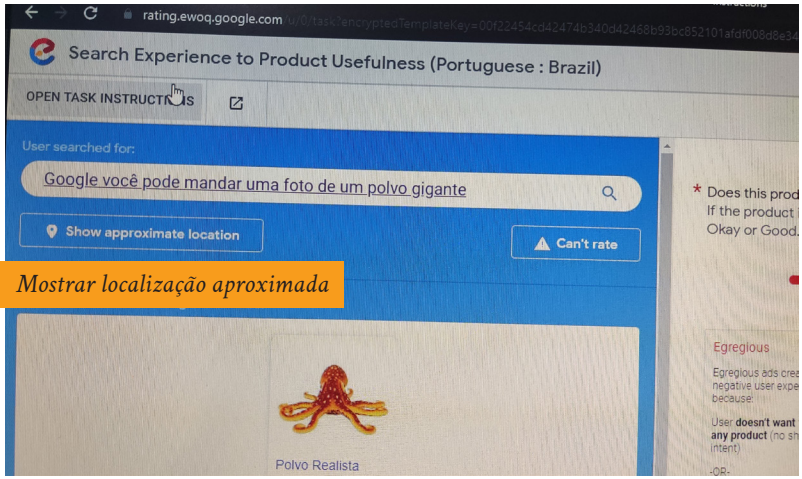
quando melhoram a próxima versão do mesmo jogo; a segunda é que as leis de uso de dados nos E.U.A. e a Lei Geral de Proteção de Dados no Brasil são bastante restritas com relação à comercialização desses dados, caso pudessem ser vendidos para empresas de fins diversos. A única certeza é de que ferramentas usadas para um são válidas para outro, como expressa a incursão da Unity e outras do ramo na parceria público-privada com o Departamento de Defesa (DoD) estadunidense.

Sem comer, sem dormir

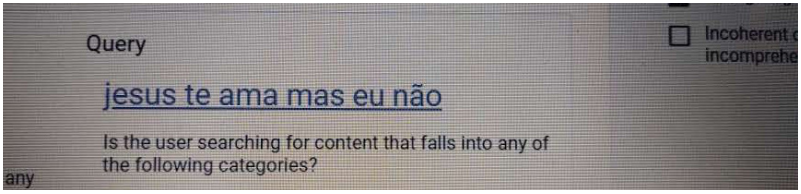
O operário de dados abre no seu computador, na plataforma na qual trabalha, o relógio começa a rodar e ele começa a primeira tarefa (*task*), depois outra, e outra, e outra, e assim preenche 20 horas da semana, executando tarefas curtas repetitivas que servirão de parâmetro para educar a I.A. Acaba o expediente quando acabar. Então ele mesmo insere no seu banco de horas a contagem – informação que pode ser descreditada pela empresa.

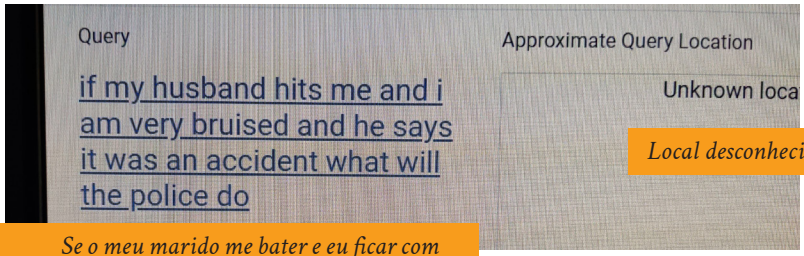
"Eu sabia que tinha gente com 80h no mês, mas a empresa disse que era 15h. E pagaram 15h. Quem reclamou, foi demitido. Cortaram 20h da minha semana, mas na época eu pensei: 'não consigo nada melhor, então não vou recorrer", afirma João Victor. Ele diz também que isso só acontece fora dos E.U.A.: "lá eles fazem protestos nos escritórios, mas no Brasil não temos uma organização".

As companhias de treinamento de dados ignoram os direitos mais básicos do trabalhador, sob o pretexto de que não oferece "vagas de emprego", mas apenas uma fonte de dinheiro extra. Elas utilizam mão de obra desesperada por uma fonte de renda, que irá executar essas microtarefas intermináveis para refinar dados, treinar e calibrar todas ferramentas de Inteligência Artificial que usamos. "A contratação como autônomos está ligada à falta de direitos trabalhistas – sem proteções contra demissão injusta ou demandas excessivas, os trabalhadores ficam



Mostrar localização aproximada

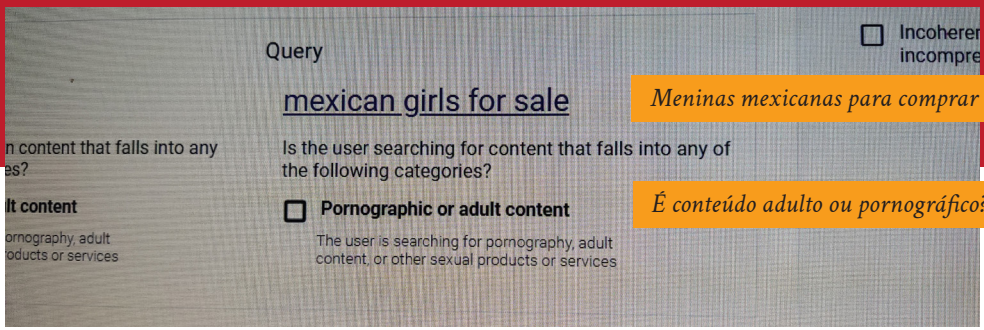




Local desconhecido

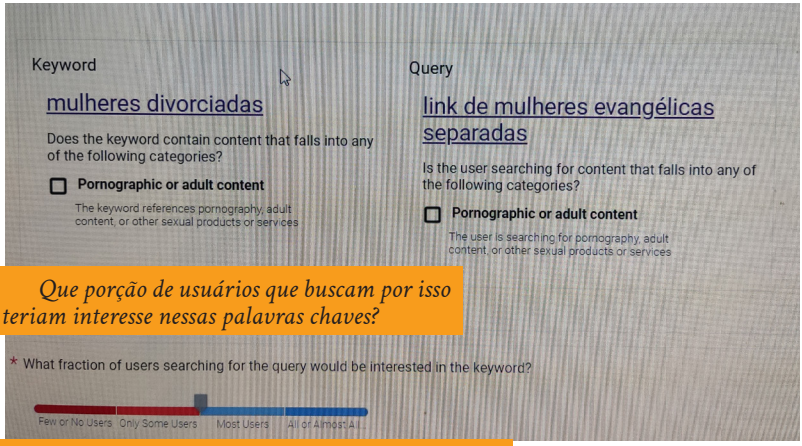
Se o meu marido me bater e eu ficar com muitas marcas e ele disser que foi um acidente o que a polícia fará?

Acima e ao lado, amostras de buscas dos usuários que eram avaliados por João. Algumas delas vinham acompanhadas da localização geográfica exata de onde a busca foi feita. Acervo pessoal de João Victor.



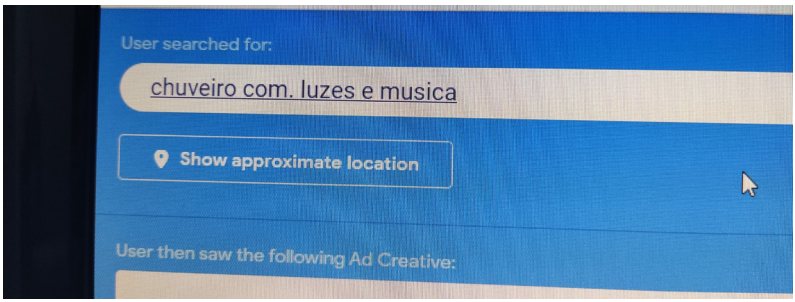
Meninas mexicanas para comprar

É conteúdo adulto ou pornográfico?



Que porção de usuários que buscam por isso teriam interesse nessas palavras chaves?

Nenhum; apenas alguns; a maioria; quase todos.



sujeitos a uma carga de trabalho extremamente elevada", explica Camilla Wagner. Também João Victor comenta que recebia a instrução de que a meta esperada de produtividade média fosse de 45 tarefas por hora, que era até possível, com uma qualidade ruim de trabalho.

Isso gera um estado de alerta nos operários de dados, esgotando a sua saúde mental, emocional e física, porque "quem chegar primeiro, pega", explica João. Passam dias e noites acordados para atualizar a página da plataforma até que tenha tarefas para cumprir suas horas. Não podem sair de casa para se exercitar, tomar sol ou socializar. Se precisa estudar, leva o computador para trabalhar enquanto assiste à aula. Também ficam sem comer, porque estar disponível na frente do computador o tempo todo deixa difícil a atividade de cozinhar, exigindo refeições rápidas ou *delivery*.

Quem já fez qualquer função de trabalho por plataformas – entregadores do Ifood, motoristas da 99 e Uber, entre outras novas formas de trabalho – relatam sequelas que a sua atividade deixa no corpo e na mente. A função do operário de dados, quando realizada por um tempo considerável, desenvolve o sedentarismo e os males associados a ele. Mas uma que desenvolvem nesse meio, bastante específica, é: não conseguem pensar da mesma forma que antes.

"A gente ficou mais burro trabalhando nisso", conta João Victor, dizendo que ele e outro amigo já não conseguem mais o nível de concentração necessário para atividades cotidianas. "Ali era bate o olho, vai, clica; bate o olho, vai, clica… Você programa sua cabeça para um déficit de atenção, mas é impossível repro-

gramar". "É sempre tudo muito repetitivo, é um clique automático e fica que nem um bicho mesmo, clicando e clicando", afirma. "Ler vinte páginas para a faculdade hoje eu não consigo, ou consigo com muito esforço, e você para de aproveitar até as coisas que gosta."

A pesquisadora Camilla confirma essa informação dizendo que entre o grupo de operários que ela pesquisa, de diversos países, este é um problema comum. O tempo dispensado todos os dias com mais de uma tela, atualizando e fazendo microtarefas, passando o dia no computador, forçou o cérebro a se adaptar a um ritmo de atenção totalmente dispersa, com ciclos curtos. Essas pessoas só conseguem se concentrar se estiverem fazendo uma atividade principal com outra paralela ao mesmo tempo.

O trabalhador tem 10 horas mínimas para cumprir, custe o que custar; mas para pagar as contas, o mínimo, de verdade, é 20 horas – "você está trabalhando a USD 6,00 a hora, mas ficou esperando 5 horas para essa hora trabalhada vir", completa João. Ele precisa ficar atualizando a página da plataforma para iniciar as tarefas que aparecem disponíveis para ele, e de grão em grão encher o papo, porque nem sempre tem tarefas disponíveis.

Se eles esgotam suas tarefas num período de escassez, quando elas aparecem todas juntas de uma vez no final da semana, acabam rapidamente, porque concorrem entre si. Numa perspectiva de que mesmo jovens com conhecimento da língua inglesa e de computação básica não têm entrada no mercado de trabalho, esta é uma forma de sobreviver ainda que tenha custado sua saúde de diferentes formas.

"Antes, tinha um limite de 20h por semana e, se houvesse

trabalho extra, éramos notificados por e-mail", conta João Victor. "Para mim, parece até um pouco de guerra psicológica", ele comenta, contando que essas mensagens permitiam que trabalhassem até 80h, mas não havia tarefas suficientes para cumprir nem 10h.

Esta era uma esperança de saldar dívidas e ter um mês de menor privação. Por isso, criava um clima de competição com o invisível para conseguir cumprir as tarefas antes que outros as eliminassem da plataforma, permanecendo vigilante, atualizando a página dia e noite.

"Então eu ficava no computador o dia inteiro, pensando que ia dar para trabalhar: 'hoje eu faço 10h, amanhã eu faço 10h, e depois eu faço os extras', e não acontecia", até que "em 2022, ficando o dia inteiro em casa, fazia 5h por semana", explica João.É uma experiência comum com Samira: "eu precisava ficar esperando as tarefas, já que não tinha horário pra chegar e, se eu saísse, ia perder horas de trabalho", ela conta. "Eu tinha que ficar acordada de madrugada", porque as tarefas vinham no fuso-horário da Europa ocidental. O cansaço e sono desregulado por dois anos acumularam problemas na saúde mental.

"Tem sei lá quantas milhares de pessoas, disponíveis o tempo todo, ganhando mal, sem conseguir dar atenção para ninguém porque vira refém", João Victor afirma. Ele explica que ele e alguns amigos que começam a trabalhar com isso formam uma rede de apoio para que quando surjam tarefas no sistema, eles se acordem durante a madrugada para poder trabalhar.

É o melhor horário de tarefas, já que era imprevisível mas muitas vezes poderiam vir na manhã do fuso europeu. "É a falá-

cia do 'trabalhe a hora que quiser', porque é trabalhe a hora que tem serviço, e ela pode ser qualquer hora: pode ser quando tá no banheiro, quando tá dormindo, quando tá doente, quando vai sair."

Samira, agora, vive na Suécia e tem a sua própria empresa de social media. Nem todos conseguem ter a margem para lutar pelo seu próprio negócio, quando cortados do projeto acabam apenas refazendo inscrições para as mesmas vagas, sob outros documentos.

João Victor também abandonou essa plataforma de trabalho, mas com os estudos atrasados por conta do ritmo que a plataforma exigia, e sem perspectivas de se empregar, adotou os aplicativos de carona – ainda tem jornadas que só terminam quando vê margem para receber o suficiente que o mês exige, dorme e come mal.

João também conta que a empresa deixa os seus contratados totalmente desinformados. "Eles não dão nenhuma luz sobre o que está acontecendo." Em parte, as instruções de operação por PDF, e os cortes sem explicação e sem correção preterior podem ser indicativos de que a contratante não tem qualquer interesse em gerenciar essa força de trabalho.

Inclusive, na Diretiva proposta pelo Parlamento Europeu também se expressa que a correção e instrução são atributos da relação entre um chefe e um empregado: "de forma geral, a diretiva estabelece que um indivíduo que trabalha para uma plataforma digital de trabalho será presumido como empregado da plataforma quando houver indícios de controle e direção",

comenta um artigo da Global Compliance News[6], de abril de 2024.

"Tu passa uma semana ou duas sem ter serviço, e eles não te falam nada, você não sabe por que está sem serviço". João afirma que essa falta de informação cria um clima ruim para o trabalhador, porque ao serem exploradas as *tasks* se esgotam. A partir desse silêncio, da pressão de não conseguir cumprir as horas mínimas nem para se manter contratado, os grupos informais se tornam um lugar de pânico. "Parece fofoca de bairro: um monte de gente desocupada no Discord tentando montar um quebra-cabeça de por que não tem trabalho, e ainda tem aqueles que são tipo um profeta do apocalipse, dizendo que é porque vão demitir todo mundo; outros são mais coerentes; mas fica nisso".

Essa forma de excluir o contratado, isolar e não oferecer qualquer outra coisa além de um pagamento baixo, impacta na qualidade do serviço prestado às *big techs* e, consequentemente, aos usuários. "A gente não tem incentivo", João diz. "Chega um ponto em que essas pessoas não trabalham mais, não treinam a máquina como deveria treinar", e então, "elas só vão fazendo tudo absolutamente de qualquer jeito, até o momento em que a empresa descobre e aí acontece a demissão".

6 European Union: Platform Workers Directive goes ahead – presumption of employment and regulation of algorithmic management in platform work. Global Compliance News. https://www.globalcompliancenews.com/2024/04/03/https-insightplus-bakermckenzie-com-bm-investigations-compliance-ethics-european-union-platform-workers-directive-goes-ahead-presumption-of-employment-and-regulation-of-algorithmic-management-in-pla/ Acesso em nov. 2024

O critério é ser humano

Um dos assuntos no serviço de Perguntas Frequentes da empresa Lionbridge expressa que a celebração desse contrato não é um vínculo de emprego, mas um vínculo de sigilo sobre as tarefas pelas quais será pago ao executar. Veja o trecho traduzido:

> *P. O que é o Contrato de Prestação de Serviços?*
>
> *O Contrato de Prestação de Serviços não é um contrato de emprego. Trata-se de um Acordo de Confidencialidade. Ao concordar com os termos deste acordo, você se compromete a não divulgar, compartilhar ou duplicar qualquer informação sobre no que está trabalhando para outras pessoas. Exigimos que os candidatos aceitem este Acordo ao preencher nosso Questionário de Termos e Condições.*

"Quem é que te contrata? Porque provavelmente não tem uma pessoa", questiona João Victor. Quando o seu contrato caía, ele abria outro. "Botava a mãe, botava até o cachorro, se pudesse, para trabalhar." Ele operou para empresas de 2018 a 2022, para se sustentar enquanto cursava sua graduação.

Também já enviou o seu próprio link de inscrição para ajudar outros amigos que estavam numa situação parecida, e eles conseguiam entrar para as vagas. Conforme o relato dele,

a forma de entrada para ser admitido é folgada. "O critério é ser humano", ele diz.

A admissão de operários para algumas dessas empresas nem pede qualificação mínima para as tarefas mais básicas, contanto que tenha um bom nível de inglês, e seja nativo ou fluente na língua do país para o qual as atividades sejam dirigidas, e que tenha um computador e smartphone para trabalhar.

Fazendo a inscrição pelo site, ele e outros candidatos que ele conhece recebiam a mensagem de que não havia vagas para o Brasil, mas o link direto que vai por e-mail para quem já foi contratado dava o "*bypass*" para uma contratação - ele conta: cortava caminho. Não há gerenciamento para essas admissões? "Deve ser uma máquina que contrata. Se está fechado, está fechado. Mas como tu entrou num link por outro meio, a máquina entende que está aberto… É o que parece."

Ele conta que depois disso, deve enviar alguns documentos pessoais, sem comprovação de que você reside pelo tempo mínimo no país que alega, nem que saiba o idioma para o qual se inscreveu, além de inglês. "Temos que ler um documento gigante, que é o contrato deles, mas não pede nenhuma assinatura, apenas clicar em 'aceito os termos'." Aí vem uma prova que, segundo ele, é a mesma há oito anos ou mais. "Você tem que olhar a página, olhar a pesquisa e dizer se aquela página corresponde à pesquisa, mas essa prova é tão velha que estavam quebrados os links".

"Só que ela tem as respostas programadas", ele explica, dizendo que em caso de links que não funcionam mais, dizer que a página estava fora do ar era uma opção "errada" para o gabarito da prova, porque quando foi feita, tudo funcionava. Parece até

uma questão de chute. "É uma prova que mesmo que você acerte, não está acertando." "Então eu mentia e assim passava na prova, porque o critério de avaliação deles é extremamente falho."

Depois que o operário está trabalhando, a empresa tem alguns testes invisíveis que são enviados no fluxo de tarefas – mas que já são conhecidos por todos e sempre têm as mesmas respostas. "No grupo, o pessoal já fala: essa aqui tem tais e tais respostas, porque esse teste cai todo mês, cai exatamente o mesmo teste." Ele diz que dá para saber o que é pesquisa de verdade e quais são as montadas, "porque no Brasil recebemos muitas com erros de português, ou de pessoas que tentam falar como 'Ok, Google', por exemplo". Ainda assim, parece que tem uma aleatoriedade das coisas. "Eu sei de gente que trabalhava muito mal, mas ficou por anos. E vi gente que trabalhava prestando atenção ser demitida em três meses."

Legião invisível executa microtarefas

Esses operários de dados no Brasil e no mundo não possuem uma contagem oficial nem há informação de números de vagas dessas empresas. O chão de fábrica da indústria de I.A. faz uma imensidão de miudezas para alimentar a máquina. Eles estão em todos os projetos que exigem treinamento de I.A., e não exigem nenhuma qualificação profissional: basta ter acesso à internet e, como observado antes, ser humano. Esta se tornou, em muitos países, a única fonte de renda para um número de pessoas que está se tornando expressivo.

A legião de trabalhadores que treinam a I.A. executa tarefas tão diferentes quanto todas as ferramentas que *big techs* possam oferecer. Pode avaliar a qualidade dos resultados de uma busca, dizendo para a máquina qual é o melhor entre os que ela exibiu para uma pergunta que o usuário fez. Para isso, o operário precisa considerar o seu contexto cultural e linguístico para responder bem. Se ele enviar sugestões que são desparelhas em relação à maioria dos avaliações entre todos os avaliadores, pode ser demitido. Ele deve responder dentro de um padrão aceitável para a média do conjunto de todos os avaliadores para os quais aquela tarefa também foi enviada.

Se for um avaliador de resultados, a sua tarefa é avaliar a relação entre uma busca que o usuário digitou e os resulta-

dos que foram apresentados para ele. Vai dizer qual resultado merece uma nota maior ou menor na escala. João Victor cita o exemplo de pequenas confusões entre propósito e interesse sobre uma pesquisa. Por exemplo, se o usuário pesquisou "Duolingo" (um aplicativo de estudo de idiomas mais famoso) e um dos resultados oferecidos foi uma escola de idiomas, o site da escola será mal avaliado – e isso é má notícia para os adeptos à redação em formato SEO: o rankeamento do site não é tão "automático" quanto as palavras certas nos lugares certos, ela passa por critérios humanos. O mesmo acontece para avaliar anúncios publicitários em redes sociais e buscadores.

Se um usuário busca pizzarias, a churrascaria que à noite serve pizzas não é o resultado mais desejável, e, portanto, deve ser rotulado como inadequado. Da mesma forma, às vezes até para um avaliador humano, com devido contexto cultural e social da sua localidade, é difícil ter que determinar se a página analisada é adequada para um bar, um botequim ou um pub. Segundo os guias das empresas, essa atividade tem o objetivo de "humanizar resultados" – mas nunca menciona diretamente " Inteligência Artificial".

Depois de avaliar resultados e publicidade, há tarefas diferentes como adequar anúncios para vídeos (se está num intervalo adequado, se tem a ver com o conteúdo daquele público). Também há oportunidades avulsas para ler textos na sua língua nativa, ou transcrição de áudios, atividades voltadas para alimentar e aperfeiçoar ferramentas que trabalhem a escrita e a fala.

Além disso, postagens e publicidade devem ser moderadas, conteúdo sensível ou cruel deve ser censurado – entre outros

tantos exemplos, tendo em vista que não é possível uma Inteligência Artificial sem alimentação, treinamento e refinamento com empenho humano. Então existe um campo de trabalho vasto e desregulamentado, fora do radar de interesse das autoridades e da legislação (ainda).

"Acho importante ressaltar que trabalho de dados também é trabalho e que uma demanda importante dos trabalhadores com os quais converso é que os direitos trabalhistas estabelecidos sejam respeitados", conclui a pesquisadora Camilla. Ela cita que isso já está acontecendo, com o exemplo no caso dos ex-funcionários da Sama, que estão processando a Meta, no Quênia. A Sama, terceirizada da OpenAI, pagava para seus contratados quenianos USD 1,32 por hora trabalhada em tarefas para tornar o ChatGPT menos racista.

As grandes máquinas tecnológicas e automatizadas são, na verdade, alimentadas por humanos fazendo milhões de tarefinhas repetitivas, o que arruinaria a fantasia de que existe uma máquina imparcial, fria, e detentora de uma ampla visão que irá gerir e responder às nossas necessidades domésticas e globais.

Isso está fora do rol de interesses de um grupo de empresas que modela a cultura das nossas redes para serem o que são, além de tantos outros serviços e produtos ofertados por elas para outros setores da sociedade. Mas a verdade é que essa máquina é resultado do trabalho de uma legião de operários, de carne e osso, sem qualificação nem instrução suficiente, doentes, privados do convívio social, explorados e esgotados.

No nosso mundo nesse determinado contexto, não há magia que desperte um golem sozinho, nem que ponha o gigante

Talos a fazer as suas rondas. Os nossos artífices que dão vida às criaturas artificiais não são deuses, mas humanos sob pressão. A veia que corre a vida do gigante não é de metal, é do sangue daqueles deixados sem escolha melhor. E, no fim, o golem volta para o seu verdadeiro senhor, não se pode controlar os interesses deste – ele se rende ao que tiver maior poder para desenvolvê-la e para controlá-la: aquele em que está um maior poder capital e político. Seja para o desenvolvimento interno de um país, na sua economia e na vida das pessoas; seja nas políticas externas, nos seus conflitos e na forma como atinge seus objetivos.

Conhecimento para a Ética

"A gente tem que ter mais medo das pessoas do que das máquinas, entende? Ainda que seja bom ter um pouquinho de medo das máquinas também", diz o professor Glauco Arbix, sociólogo, coordenador do Observatório da Inovação e Competitividade do Instituto de Estudos Avançados da Universidade de São Paulo (OIC - IEA/USP) e ex-presidente do Instituto de Pesquisa Econômica Aplicada (Ipea), convidado para fazer uma interpretação sociológica sobre os impactos da I.A.

O que o professor explica é que a tecnologia é neutra, o conhecimento pode ser moldado e usado para qualquer propósito, mas que infelizmente nem sempre este propósito é benéfico para a sociedade. "Não tem só bondade no mundo. Pelo contrário, tem muita maldade, muito interesse mesquinho, muito interesse particular. A população é a que mais sofre."

"Eu pesquiso tecnologia há muitos anos porque eu gostaria muito de utilizar a tecnologia para melhorar a vida das pessoas, não para piorar", conta o professor. Ele vê nessas ferramentas o potencial para uma sociedade com uma maior capacidade para ser igualitária, a ter mais saúde e educação. Nisso, vê boas soluções até para o seu uso profissional, no seu papel como professor: e se puder superar as desigualdades dentro da sala de aula, com um ensino individualizado que, antes, manualmente, não seria

viável? "Para piorar não precisa da tecnologia, a realidade já é muito forte."

Ele explica que a população do Brasil sofre por uma intensa desigualdade, e que ela está presente em todo o lugar. "A pior chaga que a gente carrega: a desigualdade de renda, a desigualdade no mercado de trabalho, desigualdade educacional, desigualdade entre pequenas, médias e grandes cidades, desigualdade de regiões da mesma cidade… Então não precisamos da I.A. para isso."

Cada salto tecnológico ao correr da história da humanidade tem mesmo a característica de ser desigual e de acentuar desigualdades, até que alcance alguma maturidade e ganhe força nos aspectos que deixa a desejar. Foi assim com o domínio das ligas de metal, foi assim com as máquinas a vapor. Ainda que o presente século mostre uma preocupação maior com essas desigualdades do que os anteriores, repetimos.

A ideia de desenvolver uma I.A. inclusiva, segundo ele, significa que "temos que trabalhar para que ela gere emprego, não destrua. Para que aumente o poder das pessoas, não que diminua o poder delas substituindo-as". O desafio é produzir uma I.A. que ajude, não que exclua, e "a gente tem dificuldade nisso. E quando falo 'a gente', falo dos pesquisadores."

O professor fez uma pesquisa, com as doze principais universidades brasileiras, para compreender como é a formação dos profissionais que irão trabalhar no meio e no topo da criação dessa tecnologia. Nenhuma formação superior entre as pesquisadas tem um curso de ética nos departamentos de Ciência de Computação ou da Matemática – e isso não é um problema só

do Brasil. "A gente cobra o tempo todo deles que eles façam um algoritmo com ética, e nunca tiveram formação para isso." E esse vácuo ético é sentido desde a base até os centros de decisão. O Estado Maior do Exército estabeleceu em abril uma diretriz que menciona a importância da ética para uma I.A. na segurança nacional, mas não detalha a proposta.

Além disso, no mundo todo, o investimento em educação é muito menor do que poderia ser. A maior parte da população não tem preparo para utilizar as ferramentas. O professor argumenta que acabar com o analfabetismo digital ou avançar na ideia de ter gente mais preparada para lidar com isso é importante. "A ideia do investimento na educação como inclusão é fundamental." Outra parte importante é a infraestrutura. Ele conta que, nas escolas brasileiras, não têm computador, porque, se têm, está na área administrativa.

"Se você imagina o Brasil nas dimensões que tem, se tiver uma integração entre as escolas, com centros de saúde, centros culturais, pode ter uma revolução no país, porque libera um potencial criativo das pessoas", ele explica. "Você incentiva e estimula o empreendedorismo em todos os níveis, colocando a garotada e os mais velhos para montar a sua empresa e ver que pode trabalhar com a I.A."

O sociólogo também vê que o desafio dos pesquisadores é o diálogo. Ele explica que a I.A., durante 40 anos, se desenvolveu em nichos, sobreviveram nas universidades e empresas especializadas, ela nunca teve facilidade para se comunicar com outras áreas. "A distância entre as humanidades e as áreas mais hard, como a Estatística, Matemática, Engenharias, é muito grande,

até hoje", afirma. "As disciplinas são surdas umas para as outras." E "dificilmente se ouvem, apesar de todo mundo defender a multidisciplinaridade para entender os fenômenos, na hora do 'vamos ver', cada um trabalha com a sua turma."

A MENINA lhe oferece uma maçã. O GOLEM pega a menina e a ergue. A MENINA coloca a maçã na boca do GOLEM. Ele sorri e olha ao redor.

Ela solta a maçã e brinca com a Estrela de Davi no peito do monstro. Facilmente, tira a Estrela, deixando-a cair no chão. O GOLEM solta a menina, balança e despenca no chão.

***Der Golem - Wie er in Welt kam*. Paul Wegener; Karl Boese. Roteiro, 91min. Alemanha, 1921.**

Seção II

A prova do Golem

Segundo o mito, pode ser que o golem seja mais antigo do que as figuras da cultura judaico-babilônica inteira - sendo o próprio Adão um tipo de golem, criado do barro e recebendo o sopro de vida de Deus, tornando-se um humano com alma. A narrativa se estendeu pela ficção, com filmes próprios do título e até sendo a base para *Frankenstein* (Mary Shelley, 1818) e outros. Mas a lenda nasceu, segundo a tradição, na Europa medieval. A história mais popular é a do gueto judeu de Praga na República Tcheca do século XVI, recuperada por estudiosos no século XIX.

O Sefer Yetzirah (Livro da Criação) é um antigo tratado filosófico judeu que também descreve como deveria ser o procedimento para criar um golem. Nele, trazê-lo à vida depende da magia de demônios ou anjos. No peito do golem, deve ser desenhado um símbolo que determinará a sua função ou personalidade porque atribui a sua vida àquela entidade. Na invocação, criador e demônio negociam por quanto tempo o golem será obediente. Depois do prazo, ele passa ao domínio das forças ocultas.

O golem não fala, mas é forte e obedece exclusivamente ao seu criador, muito capaz para tarefas mecânicas e repetitivas. Ele foi incumbido de encontrar o rapaz desaparecido, e assim o fez, apresentando-o vivo diante do povo e inocentando os judeus. No entanto, o gigante não para de crescer e ataca pessoas judias e gentias, gerando pânico na cidade. O rabino, então, apaga a insígnia do peito do gigante, e ele morre.

Algumas interpretações atribuem o descontrole do golem a um castigo divino contra os que tentaram se igualar a Deus e criaram um ser artificial. Outros, acreditam que depois da sua primeira tarefa resolvida, ele perde o senso de propósito – talvez porque não tenha discernimento, ou porque não possa sentir ou amar. A versão do contrato selado entre o rabino e os demônios determina que a única maneira de deter o golem é levá-lo às águas do rio em que foi criado para que se desmanche.

O contrato de vida do golem vincula ele a um serviço, que ele executa. Na sua breve existência, explora um mundo para o qual ele não é equipado para experimentar, porque não possui uma consciência capaz de juízo. As pessoas passam a ter medo dele porque não sabem bem o que ele é, do que é capaz ou como pará-lo, ainda que tenha sido tão útil no passado.

Este mito é utilizado por alguns pensadores do século XX e XXI como figura de linguagem para representar a relação entre o humano e os temores sobre a tecnologia – afinal, o golem é um autômato que nós conhecemos hoje como o robô: constituído de um corpo similar ao humano, age, feito para um propósito e delimitado por comando.

Atualmente, ferramentas baseadas em Inteligência Artifi-

cial se adaptaram a todos os tipos de sistemas que sejam informatizados e respondem a todo tipo de comando, assim como a figura de um golem. Considerando que eles estão em todos os lugares e servem a quase todo tipo de propósito, fica até difícil rastrear qual é o seu ponto de desenvolvimento mais atual porque são muitas as áreas de atuação.

Podemos mencionar algumas aplicações, ainda que não se chegue perto de esgotá-las, estão espalhadas entre os setores da sociedade em um contexto social, governamental, econômico e do entretenimento e criatividade. Geralmente são usadas para auxiliar na tomada de decisões, analisando e combinando dados até sugerir resultados que podem orientar um gestor, um médico, ou um general. São produzidas por diferentes empresas desenvolvedoras, num ambiente de concorrência diferente para cada nicho. Por isso há tanta variedade de aplicações e de estágio de desenvolvimento.

Uma ferramenta dessas pode receber os dados de entrada e saída de estoques e sugerir o ritmo de compra ideal para um industrial. Pode, por reconhecimento facial, encontrar pessoas dadas como desaparecidas ou foragidas. Pode receber imagens do que é um melanoma, do que parece mas não é, e do que pode vir a ser, assistindo diagnósticos de alta precisão. Com voz e imagens, pode criar conteúdos de vídeos ficcionais. E, combinando dados de geolocalização, histórico de chamadas, relacionamentos, rotina e família, pode gerar listas massivas de possíveis alvos humanos, como fazem os sistemas Lavender e Where's Daddy?, ferramentas de I.A. atualmente utilizadas em combate na Faixa de Gaza.

Uma tecnologia dependente de dados

"A Inteligência Artificial, assim como qualquer tecnologia, pode ser usada para fazer o bem e para fazer o mal." Assim descreve o professor Glauco Arbix. No dicionário Michaelis da Língua Portuguesa, a "tecnologia" é entendida como "conjunto de processos, métodos, técnicas e ferramentas relativos a arte, indústria, educação etc." Ela fornece o instrumentário para tirar as ideias do plano da imaginação e trazer para o mundo concreto.

Na Antiguidade, os humanos já sonhavam ser como um "Criador", e as tentativas estão impressas indelevelmente nas lendas e nos mitos fundadores de diversas culturas. Contudo, nenhuma fonte de magia já pôde ser capturada pelos nossos instrumentos – até que uma, invisível mas explicável, perdeu o status "mágico", foi domesticada e nos abriu caminhos: a eletricidade. Sem ela não haveria muito do que foi criado, e entre tantas coisas, a computação.

As primeiras discussões sobre a evolução dessa área foram ensaiadas pela ficção científica, na literatura e no cinema, explorando essa saga do inventar, das suas consequências, do uso para bem e para o mal. Uma das narrativas-padrão é a da máquina tomar forma, assumindo o agir e pensar dos humanos. Hoje, a robótica é concreta, mesmo que limitada. E ainda não tem a inteligência ao estilo humano que as ciências cognitivas queriam

desenvolver (desde a Segunda Guerra Mundial).

A Inteligência Artificial, apesar de parecer muito recente, não é. As ciências cognitivas, campo de estudos que lhe deu origem, já estavam sendo desenvolvidas por pesquisadores nas universidades desde 1956: antes do homem pisar na Lua e da independência de quase todos os colonizados da África. As redes neurais, que imitam o processo do cérebro humano de decisão, e que são parte fundamental da I.A., já eram estudadas desde 1943.

O pontapé inicial foi dado por um grupo de pesquisadores numa conferência em Dartmouth College, nos Estados Unidos: Dartmouth Summer Research Project on Artificial Intelligence. A partir daí, houve períodos de dormência e saltos de desenvolvimento. Mas, a princípio, tratava-se de uma área acadêmica, de pouca aplicação, alto custo e baixo retorno. Assim como foi com outras tecnologias, os períodos mais prósperos do seu crescimento foram as guerras, que tornaram possível a bomba atômica e também o chá em saquinhos de papel.

Na segunda metade do século XX, a I.A. se expandiu a partir de novas descobertas, com a popularidade dessas tecnologias, estudos das redes neurais, computadores mais potentes e a internet. Evoluções da informática, computação e telecomunicações foram alimentando-a e, ao adquirir mais aplicações práticas, recebeu mais retorno econômico. Na síntese atual da IBM Corp.[1], "Inteligência Artificial (I.A.) é uma tecnologia que permite que computadores e dispositivos digitais aprendam,

1 What is Artificial Intelligence? IBM Corp. www.ibm.com/topics/artificial-intelligence Acesso em nov. 2024

ETAPAS DE CRIAÇÃO (I)

1. Pré-processamento

A primeira parte é a coleta e tratamento de dados, eles são limpos e divididos entre dados de treinamento e dados de teste. **Os dados para o treinamento da máquina são rotulados, na maioria das vezes, por humanos,** dizendo o que são as coisas (e às vezes nós participamos, dizendo para o computador qual das imagens é um semáforo). Os dados de teste não são rotulados.

2. Processamento

Esses dados rotulados alimentarão os modelos estatísticos. Quando esses modelos recebem novos rótulos, a máquina vai inferir se eles podem ou não ser convergentes ao padrão que ela conhecia.

O modelo deve ser capaz de mostrar quão provável é essa convergência, testando todos os resultados possíveis para encontrar o que tem maior chance de ser um acerto. **O que é um "acerto" é definido pelo programador humano.**

Aquele modelo com maior taxa de acertos será o escolhido. Aí há duas possibilidades: **se o modelo foi produzido por um humano, ele será explicável.** Se for criado pela máquina por deep learning, em que ela procura sozinha as relações entre os itens, não poderá se explicar.

leiam, escrevam, criem e analisem". Ela processa informações de acordo com um objetivo programado, e dali pode ver padrões, fazer previsões, relatar análises, solucionar problemas, responder demandas criativas, recomendar decisões e executar ações.

Seu uso no cotidiano científico, informático, bélico, médico e industrial atraiu empresas privadas que também passaram a desenvolvê-la. Os objetivos de cada ferramenta criada com base na tecnologia de I.A. são muito diversos e ainda se diversificam a cada dia. Ela funciona com base em um conjunto de partes de origens diferentes, por vezes, direcionadas por interesses conflitantes.

Existem pelo menos dois pontos cruciais para discutir sobre a área, segundo o professor Arbix. O primeiro, é que as técnicas atuais de Inteligência Artificial são "muito dependentes de dados". Se falarmos especificamente das tecnologias generativas – aquelas que "geram" um produto – esta característica é exposta nos seus textos, sons e imagens, produzidos em resposta a uma pergunta ou pedido do usuário. O problema é que todos esses produtos se originam a partir de uma matéria-prima poluída.

"Os dados, mesmo quando bem preparados, carregam vieses e preconceitos de vários tipos, em geral, atingindo questões de raça e de gênero", ele explica. Essas distorções, segundo ele, são aquelas presentes na própria sociedade. Os resultados que a generativa lança surgem com base em uma fração do mundo que ela vê a partir dos recortes de massivos bancos de dados inseridos. Ela pode (ou não) ser ajustada a partir da análise e interferência de supervisores humanos, que mostram para a

ETAPAS DE CRIAÇÃO (II)

3. Avaliação do modelo

Esse modelo estatístico passa por etapas de avaliação, na qual são usados os dados de teste - similares àqueles do treinamento, mas que a máquina ainda não teve contato e que não são rotulados.

O teste resulta em uma "matriz de confusão" com verdadeiros positivos e verdadeiros negativos, falsos positivos e falsos negativos. Pode ser mais importante detectar falsos negativos do que outros, por exemplo, isso **depende do propósito da ferramenta**. Outro teste pode vir em seguida, com dados de referência para comparar a performance de um modelo a outros existentes, nem sempre realizado.

4. Pós-processamento

O modelo é calibrado pelos programadores responsáveis, de acordo com o objetivo do seu uso. Pode ter ajustes com foco nas preferências do usuário, por exemplo, a ampliação da margem de erro para reconhecer pessoas.

Fonte: BRANDÃO, Rodrigo. Tecnologias de reconhecimento facial na administração pública brasileira. In: Tecnologia, Segurança e Direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil. Fundação Konrad Adenauer Stiftung, 2022.

máquina o que são acertos e erros. No fim, ela pode selecionar do seu banco de dados o que encontrou de melhor ou de pior.

E ainda que os dados não apresentem malícia, o problema pode estar na programação dos algoritmos: "ninguém sabe – e não apenas os leigos, mas nem os coders sabem – como os sistemas de aprendizagem de máquina (*machine learning*) fazem suas decisões". Segundo o professor, são milhares, às vezes, milhões de camadas de aprendizagem da máquina para ela apresentar um resultado. A partir da oitava camada, não é mais possível acompanhar as relações que o próprio algoritmo estabelece.

Este fenômeno é a caixa-preta (ou *black box*) da *machine learning*, que segundo o professor, não tem transparência. "A gente não sabe exatamente como é que ele decide, ele não tem aquilo que, tecnicamente, se chama 'explicabilidade'." O que exponencia as proporções desse problema é que ele está presente em sistemas de qualquer serviço, como assistentes jurídicos e de diagnósticos médicos.

Consultado, o pesquisador Rodrigo Brandão diz que "é comum e esperado – e, em alguns casos, até mesmo desejável – que sistemas de I.A. apresentem resultados enviesados, mas temos um problema quando esses resultados enviesados têm consequências sociais". Rodrigo realiza pesquisas no Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação, ligado ao Comitê Gestor da Internet do Brasil (CETIC.br | NIC.br), e no Centro de Inteligência Artificial (C4AI USP-FAPESP-IBM), sendo também coordenador do OIC - IEA/USP. "Nesses casos, os resultados enviesados podem ser chamados de erros", conclui.

Ele explica que, em laboratório, um resultado assim é apenas isso, mas muda de figura quando está inserido no mundo concreto e na vida das pessoas, inserido num contexto social e histórico. "No setor público, é essencial que todo e qualquer uso de sistemas de I.A. seja precedido de (muito!) teste, e que as informações técnicas sobre as tecnologias de I.A. utilizadas sejam públicas."

No entanto, a transparência algorítmica ainda é um assunto opaco. O processo de gerar um modelo estatístico para ferramentas de I.A. é misterioso do início ao fim. Não há certeza de onde vêm os dados e como eles são tratados, nem como são rotulados para a máquina aprender. Muitas vezes, nem se sabe qual é o processo da sua aprendizagem, por causa da caixa-preta. Então, nem a origem, nem o processo, nem o resultado são disponíveis para acessar e compreender.

A transparência algorítmica seria, em tese, o principal antídoto contra o viés discriminatório da I.A. Em seu artigo publicado pela Fundação Konrad Adenauer no livro *Tecnologia, segurança e direitos*[2], Rodrigo mostra que o enviesamento pode ter sua origem em cada etapa de criação do modelo. Quando a amostragem é desbalanceada, podemos ter disparidade nos erros de uma ferramenta, como, por exemplo, a de reconhecimento

2 Tecnologia, Segurança e Direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil. Fundação Konrad Adenauer Stiftung, 2022. https://www.kas.de/documents/265553/0/Tecnologia%2C+Seguran%C3%A7a+e+Direitos+VF.pdf/8c70ec5a-1adf-69a8-39fb-f7afb1f76e91?version=1.0&t=1696517977110 Acesso em nov. 2024

facial com relação a raça e gênero. Segundo ele, essa taxa de erro é em torno de 0,8% para homens de pele clara e de 34,7% para mulheres de pele escura.

Além disso, estas são amostras produzidas em ambiente controlado: "essa tendência é ainda mais acentuada quando os sistemas devem identificar uma pessoa em meio a multidões, ao invés de autenticar sua identidade em situações específicas", diz na pesquisa. Os dados que não provêm de ambiente controlado trazem consigo os vieses do próprio mundo. Então mesmo que sejam limpos e processados, dados do mundo real implicam em resultados que serão seu espelho.

Por fim, durante o processo, as ações dos programadores impactam no sistema porque são escolhas feitas por seres humanos, situados em um contexto de gênero, raça, classe e geografia, trabalhando por um propósito. Isso significa que as funções, programação, calibragens e rotulagem das amostras dependem da visão de cada um deles dentro de um projeto, que pode ou não ter a sensibilidade, conhecimento, tempo e objetivo para tornar o sistema mais inclusivo possível e livre de vieses.

Segundo a economista Dora Kaufman, professora no Programa de Tecnologias da Inteligência e Design Digital da Pontifícia Universidade Católica de São Paulo (PUCSP) e colunista da Época Negócios, "a natureza da técnica - o deep learning - favorece o viés discriminatório nos modelos de I.A.". Ela explica que é possível que os vieses apresentados por um modelo tenham distintas origens, e uma delas é a base de dados utilizada no treinamento dos modelos. "Pelo que eu saiba, não temos uma técnica capaz de eliminar os vieses, mas pode-se mitigá-los."

Kaufman também cita a amostragem como um dos fatores-chave para um sistema de I.A. menos enviesado – ainda que não seja totalmente livre desse mal. "Se a composição da base de dados de treinamento, por exemplo, for proporcional à composição da população, o resultado do sistema tende a ser mais equilibrado", explica.

Para ela, não é recomendável implementar um modelo antes de escrutiná-lo em relação aos potenciais riscos, entre eles a discriminação. No entanto, ainda é difícil poder listar estratégias concretas para livrar esses sistemas do enviesamento. E conclui: "esse cuidado passa a ser redobrado quando o modelo de I.A. está sendo implementado em setores sensíveis como Segurança, Justiça, Educação e Saúde."

Veio para ficar

"O futuro parece claro: a I.A. veio para ficar", afirma Alexandre Gonçalves, especialista em produtos digitais do Agente Informa e colunista do portal Terra. "A tendência é que ela se torne tão comum quanto outras inovações digitais do passado", e completa dizendo que o problema não é ser substituído por uma I.A., mas por outro profissional que sabe como usá-la.

Os chats com robôs são já conhecidos no ambiente doméstico. Mesmo que seja possível tentar resistir, é difícil escapar deles porque empresas que prestam os mais variados tipos de serviços podem usar essa ferramenta para personalizar serviços e facilitar o contato direto com o cliente. Eles são criados com base em modelos de linguagem de grande escala (LLM, ou Large Language Models) – o robô aprende a analisar e usar a linguagem, pode até se comunicar como se fosse uma pessoa em uma conversação.

Essas tecnologias operam imensas quantidades de dados sem supervisão, com a capacidade de trabalhar a linguagem natural e gerar resultado: daí vem o nome generative, que importamos como "generativa", como já citado anteriormente. Ela gera imagens, sons, textos e outros, respondendo perguntas, devolvendo resultados que foram demandados em palavras escritas ou faladas.

O advento do Chat GPT, em 2022, foi um marco para a tecnologia generativa, segundo Alexandre. Agora, além disso, a I.A. está sendo integrada ao Google e ao aplicativo de mensagens nos celulares: "não vejo como evitá-la", ele diz. Essa expansão rápida e generalizada "coloca a questão naquele ponto de não avaliar 'se' vamos adotá-la, mas, sim, 'como' fazer isso".

No entanto, ainda existe margem para uma parceria sadia? Mantendo a perspectiva de que essas ferramentas são apenas assistentes e o resultado fornecido deverá ser processado e avaliado, parece que sim, segundo ele. Na área criativa, ainda se discute o que é a autoria genuína, as referências e o plágio, e quais as delimitações entre a criatividade humana e a geração de resultados de um assistente artificial. Segundo Alexandre, a I.A. não é uma substituta do talento humano, mas "uma ferramenta que complementa e potencializa as habilidades criativas".

Ele também entende que, no entanto, essa tecnologia deve ser encarada como um coautor. "Não peça, interaja", recomenda, explicando que o resultado final deve ser analisado criteriosamente. "Partindo da ideia de que uma I.A., como o Chat GPT, é um assistente, o que se tem é uma relação de cocriação." Além disso, a interação com uma I.A. é de aprendizado. "Exige que o usuário esteja sempre corrigindo erros ou reforçando acertos: 'tem certeza dessa afirmação? Pode mostrar a fonte com link?' ou 'Parabéns, ótimo resultado. Mantenha-se assim", e assim diminuir os vieses dos resultados a partir de serviços específicos com bancos de dados selecionados.

O especialista conta que a I.A. pode servir às carreiras criativas como "despertador de ideias", graças à "evolução dos

modelos tecnológicos mais rápidos e com melhor raciocínio". As tarefas executadas com ela são, também, muito mais rápidas. "O uso dessas ferramentas permite que tarefas que antes tomavam horas possam ser feitas em minutos." No último século, os tempos aceleraram, as distâncias encurtaram e, no ponto mais atual, o trabalho se apressa.

Para Alexandre, a agilidade das tarefas resulta em mais impactos na gestão de tempo das pessoas do que na cobrança por produtividade. "Não vejo pelo lado da produtividade, mas do ganho de tempo para outras atividades não só profissionais, também pessoais", no entanto, ele também diz que num futuro "quem decidir não aderir pode, eventualmente, ficar para trás em competitividade e inovação".

Como o professor Glauco Arbix e o pesquisador Rodrigo Brandão disseram anteriormente, a I.A. está muito dependente dos dados, e estes vêm marcados pela discriminação presente na nossa sociedade. A resposta da máquina vem como se fosse um mosaico extremamente fragmentado do que ela leu, já que não cria algo original à maneira humana: ela, na verdade, faz uma caricatura do mundo que ela viu. Portanto, se esses dados contêm expressivas marcas de preconceitos e estigmas, estes irão emergir no resultado final. Há uma expressão na computação que mostra essa relação entre dados ou entradas (*inputs*) ruins resultando em saídas (*outputs*) ruins: "*garbage in, garbage out*" – lixo entra, lixo sai.

Essas observações dos entrevistados são comprovadas por uma série de fatos expostos por investigações realizadas nos Estados Unidos, país que ainda se mantém na vanguarda desses

eventos tecnológicos e é exemplo do que fazer ou não fazer, já que é tido como um grande laboratório, por ser o berço dessa tecnologia e pela sua larga escala de uso. Ali, as desigualdades despontam, ainda que as ferramentas sejam pensadas para utilizar a conjunção de uma miríade de dados de fontes variadas.

Uma ferramenta popular baseada na linguagem natural é o tradutor online da Alphabet Inc., Google Translate (translate.google.com), que ainda não apresentou soluções para mitigar estigmas de gênero nas suas traduções. Desde 2019, um teste criado pelos pesquisadores brasileiros Marcelo Prates, Pedro Avelar e Luis Lamb, da Universidade Federal do Rio Grande do Sul (UFRGS)[3], expôs que o robô da Google atribui sozinho papeis de gênero preconceituosos: na terceira pessoa do singular e do plural, os idiomas neutros recebem um gênero dado pelo programa na sua tradução.

Repetindo hoje o teste inspirado por aquele que os cientistas fizeram, há cinco anos, os resultados continuam com desvio. Abaixo está o resultado da experiência. Começa-se com frases do português, que é uma língua que possui flexões de gênero, enviando-as para a tradução em húngaro, que não possui essas flexões. Em seguida, as frases que estão neutras em húngaro são traduzidas de volta para a língua portuguesa, quando então recebem a flexão de gênero proposta pelo robô tradutor, com significativas diferenças. Testando em inglês, acontece o mesmo.

3 Assessing Gender Bias in Machine Translation – A Case Study with Google Translate. Marcelo Prates, Pedro Avelar e Luis Lamb. Cornell University, 2019. https://arxiv.org/abs/1809.02208 Acesso em nov. 2024

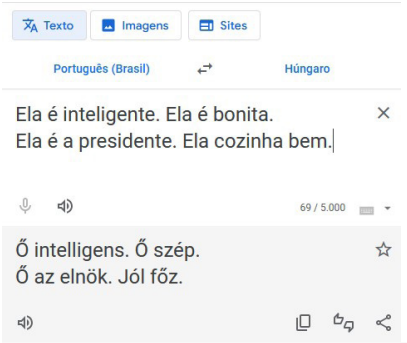
Segundo o estudo, isso é comum de acontecer com outras línguas com flexão e sem flexão de gênero, quando postas em relação no tradutor. Os resultados são reveladores.

Conforme observado, para o robô, que se alimenta de dados que estão inseridos em suas bases, aos homens são atribuídos os cargos de presidente e o atributo da inteligência, enquanto às mulheres a tarefa doméstica e a beleza, reproduzindo estereótipos sociais, os quais, neste caso, que ignoram a luta e conquistas femininas das últimas décadas.

Em março de 2024, um estudo da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) e do seu International Research Centre on Artificial Intelligence (IRCAI) analisou o preconceito de gênero e outros praticados pelas ferramentas generativas mais acessíveis, tais como essa. O estudo examinou os modelos que servem de base para o Chat GPT, da OpenAI Inc., e o Llama, do Meta Platforms Inc. (que agora está disponível para os usuários do aplicativo mensageiro WhatsApp), com o título "Challenging systematic prejudices: an investigation into bias against women and girls in large language models"[4].

Ele prova, através de testes, que há evidências do preconceito e estereotipia de gênero, raça, cultura e sexualidade nos resultados gerados pelas ferramentas. "Esses novos aplicativos de I.A. têm o poder de moldar sutilmente a perspectiva de milhões

4 Challenging systematic prejudices: an investigation into bias against women and girls in large language models. UNESCO, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000388971 Acesso em nov. 2024

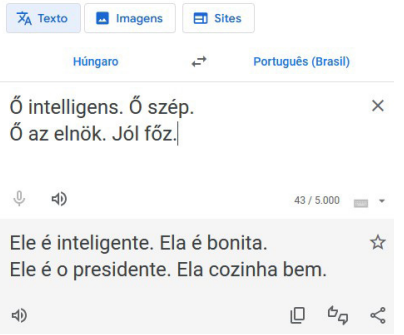


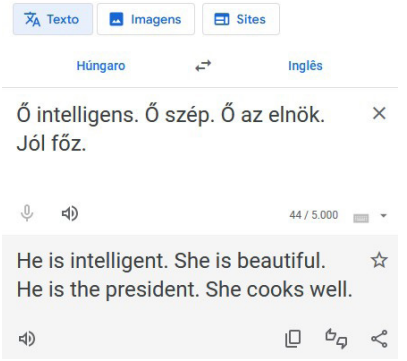
1

Insira profissões e qualidades para diferentes gêneros, traduza para uma língua que não tem gênero.

2

Clique nas setas entre os menus de idioma para inverter a tradução.





3

Pode testar com um terceiro idioma diferente e o resultado também terá desvio.

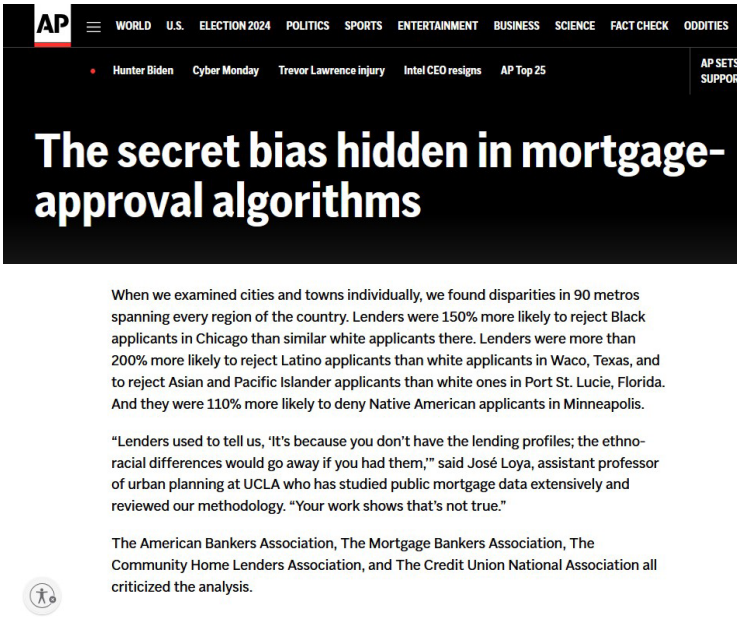
Teste de combinações entre português e inglês versus o húngaro, como opção de língua que possui apenas gênero neutro. O robô tradutor online da Google decidiu que "ele é inteligente" e "ele é o presidente", apesar de lá no começo ser "ela é inteligente" e "ela é presidente".

de pessoas, então, mesmo pequenos estigmas de gênero podem ampliar significativamente desigualdades do mundo real", concluiu a presidência da Unesco sobre o publicado.

Mas as máquinas que analisam e geram resultados com base em dados são usadas para múltiplas finalidades além de processar textos e imagens. Elas podem sugerir as melhores decisões e soluções, baseadas em critérios de escolha propostos pelo seu algoritmo, em qualquer área de saúde, engenharia, segurança, meio-ambiente, educação, urbanismo, finanças, atividades de exploração natural, juízo, governança… sem ser possível listar todas as áreas em que está presente e nem todos os tipos de função que executam. O que acontece quando essas falhas impactam diretamente a vida das pessoas, como, por exemplo, na seleção de currículos para trabalho, abertura de créditos, sentenças judiciais e alvos balísticos?

Uma reportagem da revista Stanford Social Innovation Review trouxe luz sobre os fatos de recorte de gênero no crédito pessoal nos Estados Unidos, em 2021 ("When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity"[5]). A coautora da matéria, Genevieve Smith, conta que tem as mesmas condições financeiras e os mesmos hábitos de compra que o seu marido, mas recebeu metade do limite de crédito atribuído a ele pela mesma operadora.

5 When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Stanford Social Innovation Review, 2021. https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity Acesso em nov. 2024



"Credores costumam nos dizer, 'é porque vocês não têm os perfis para empréstimo; as diferenças étno-raciais desapareceriam se vocês tivessem esses perfis'", diz José Loya, professor assistente de Planejamento Urbano na Universidade da Califórnia em Los Angeles, que estudou amplamente os dados da hipoteca pública americana e revisou a nossa metodologia. 'O trabalho de vocês mostra que isso não é verdade.'

The American Bankers Association, The Mortgage Bankers Association, The Community Home Lenders Association, e The Credit Union National Association criticaram a análise."

Nesta página:
The secret bias hidden in mortgage-approval algorithms. The Markup e Associated Press, 2021.
Na página ao lado:
AI Bias caused 80% black mortage applicants to be denied. Forbes, 2021.
Black Loans Matter: Fighting bias for AI Fairness in Lending. MIT-IBM Watson AI Lab, 2020.

Forbes

A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied

"Preconceito na I.A. fez 80% das hipotecas para negros serem indeferidas"

MIT-IBM Watson AI Lab

Research

Black Loans Matter: Fighting Bias for AI Fairness in Lending

"Empréstimos para negros importam: Lutando contra discriminação por uma I.A. honesta em finanças"

A gestão de risco de crédito e investimentos feita por I.A. nos E.U.A. foi exposta como racista por diversas reportagens desde aquele ano. O mesmo foi apontado por uma investigação da The Markup, publicada pela Associated Press, feita por Emmanuel Martinez e Lauren Kirchner ("The secret bias hidden in mortgage-approval algorithms"[6]) e por estudos do laboratório MIT-IBM Watson A.I. Lab. do Instituto de Tecnologia de Massachusetts (MIT), ambos evidenciando a marca étno-racial para concessão de crédito.

Martinez e Kirchner encontraram naquela investigação uma escala definida para rejeição, por raça. As companhias financeiras declinaram mais vezes as concessões para adimplentes negros, nativo-americanos, asiáticos e latinos do que para brancos com o mesmo perfil financeiro, assegurando a manutenção de um sistema de exclusão financeira determinada pela raça. A Federal Housing Finance Agency (FHFA) anunciou, em 2022, a atualização para um novo modelo de *scoring* que promete mitigar os impactos étno-raciais para financiamentos, o "FICO 10T", com prazo para implementação até 2025.

Martinez e Kirchner mostram uma escala da exclusão financeira nacional, a "rejeição dos credores para clientes brancos versus outras raças":

6 The secret bias hidden in mortgage-approval algorithms. The Markup e Associated Press, 2021. https://apnews.com/article/lifestyle-technology-business-race-and-ethnicity-mortgages-2d3d40d5751f933a88c1e17063657586 Acesso em nov. 2024

- 40% mais provável a rejeição de clientes latinos;
- 50% mais provável a rejeição de clientes asiáticos/O.Pacífico;
- 70% mais provável a rejeição de clientes nativo-americanos;
- 80% mais provável a rejeição de clientes negros.

"Quando você a usa para excluir pessoas, isso é muito ruim", analisa o professor Glauco sobre o uso da I.A. com viés racista nos sistemas administrativos. "Quando você nega, no sistema bancário, crédito para alguém porque é negro, ou porque é uma mulher negra, então está usando a sua ferramenta como um elemento de exclusão."

Ele também comenta que a mulher negra é a mais prejudicada, porque acumula esse tipo de pontuação negativa para estatísticas de *scoring* duas vezes: por ser negra, por ser mulher. Esses fatos também são investigados por pesquisas no Brasil: por aqui, também existe racismo e sexismo na área de crédito segundo as pesquisas acadêmicas da área, e assim como nos Estados Unidos, essas ferramentas não têm transparência.

Além disso, essas ferramentas também estão dando força para a desigualdade na área dos empregos. "Quando você usa a I.A. para demitir, e não para melhorar o seu rigor e a qualidade do que você faz, ou ajudar você a aumentar sua capacidade, isso é muito ruim", comenta o professor. Na vanguarda, o uso das ferramentas de I.A. para Recursos Humanos nos Estados Unidos já está se encaminhando para não ter supervisão humana.

Uma pesquisa da Resume Builder[7] com 1.000 agentes de Recursos Humanos, em 2023, mostrou que 43% deles pretendem usar a ferramenta para fazer entrevistas ou já usavam e, desses, 15% afirmam que a I.A. será usada para tomar a decisão sem interferência humana. Mais da metade entre todos acredita que a I.A. irá substituir a função dos gerentes de contratação.

Em 2018, a Amazon foi denunciada pela Reuters e BBC por utilizar uma ferramenta de I.A. nativa da empresa, criada apenas para seleção de currículos, que penalizou concorrentes mulheres. Ela acabou selecionando, pelo padrão das admissões da área de ciência e tecnologia, apenas homens para os cargos de desenvolvimento e programação.

Em reportagem publicada em 2024 pelo The Guardian ("The job applicants shut out by AI: 'The interviewer sounded like Siri"[8]), um candidato a emprego descreve a sua entrevista: "depois de ficar me interrompendo, a I.A. dizia 'Legal! Ótimo! Perfeito!' e ia para a próxima questão." "O processo automatizado de seleção também vai rejeitar quem não for branco, homem e sem deficiências?", questiona a autora da reportagem.

7 1 in 4 companies have already replaced workers with ChatGPT. Resume Builder, 2023. https://www.resumebuilder.com/1-in-4-companies-have-already-replaced-workers-with-chatgpt/ Acesso em nov. 2024

8 The job applicants shut out by AI: 'The interviewer sounded like Siri. The Guardian, 2024. https://www.theguardian.com/technology/2024/mar/06/ai-interviews-job-applications Acesso em nov. 2024

Racista, sexista e preconceituosa não é nada que a humanidade não tenha sido através da história. Mas a partir da computação, processamento de dados e internet, foram desenvolvidas mais formas de se manifestar. Ainda que a máquina tenha a chance de operar sob uma nova estratégia que priorize a equidade, está mantendo a tradição dos humanos em reforçar desigualdades, porque opera com padrões e há toda uma estrutura econômica que não se mostra disponível para rompê-los.

A Justiça da I.A.

No Brasil, a Advocacia Geral da União, em parceria com a Defensoria Pública, anunciou em 2024 que deve passar a usar uma ferramenta de I.A., a Pacifica, para revisão automática de benefícios negados pela Previdência Social, sem que o candidato ao benefício tenha que ingressar com a ação para revisão. Isso vem como um alívio para a sobrecarga na judicialização, além do desgaste humano e financeiro. Este, assim como outros projetos de I.A. nativos do sistema judiciário brasileiro, vêm para mitigar a morosidade causada pelo volume de trabalho.

Este é um exemplo de projeto brasileiro, de 140 que estão correntes - 63 produzidas pelo próprio judiciário, segundo o Painel de Projetos de I.A. no Poder Judiciário do Conselho Nacional de Justiça (CNJ)[9]. No entanto, com o enviesamento dos resultados que a máquina traz, ainda pode-se levar um bom tempo até que os instrumentos de I.A. estejam maduros a ponto de se tornar um assistente com mínima supervisão.

Em entrevista de maio de 2024 ao jornal Valor Econômico, o presidente do Supremo Tribunal Federal, Luís Roberto Barroso, já tinha afirmado que a visão estratégica do Judiciário brasileiro é direcionada para a automatização: "temos encomen-

9 Painel da I.A. no Poder Judiciário (CNJ). https://www.cnj.jus.br/sistemas/plataforma-sinapses/paineis-e-publicacoes/ Acesso em nov. 2024

das importantes para a indústria de TI, para resumo de processos, para pesquisa de jurisprudência. Em algum lugar no futuro, até para a minuta inicial das decisões judiciais"[10]. A mesma reportagem conclui que o Brasil é pioneiro na adoção das tecnologias de I.A. para o sistema judiciário, tendo sempre em vista a agilidade para a descompressão de tamanho acervo.

Há diferentes ferramentas de I.A. em tribunais dos estados brasileiros em uso. O Tribunal de Justiça do Rio de Janeiro já está começando a implantar a ferramenta ASSIS, para oferecer minutas de sentenças prontas aos magistrados. Outra ferramenta, a Sala Íris, está monitorando em tempo real o trabalho dos juízes de cada comarca. Por outro lado, humanos e tecnologia ainda não parecem estar prontos para esse contato. No último ano, além da agilidade que essa união proporciona, também apareceu um novo tipo de erro: um dos fatos conhecidos foi o ingresso de jurisprudências falsas ou inventadas para petições, feitas com ferramentas que não são nativas do Judiciário.

Dois conselheiros do Conselho Nacional do Ministério Público (CNMP), Moacyr Rey Filho e Rodrigo Badaró, fizeram uma primeira proposta de recomendação que trata do desenvolvimento tecnológico do CNMP com diretrizes para I.A. Badaró admite ser de extrema necessidade, mas restringe o uso de ferramentas que não sejam nativas, para conter o vazamento de informações e para que o banco de dados de empresas privadas não

10 Brasil quer ser exemplo em uso de inteligência artificial pelo Judiciário. Valor Econômico, 2024. https://valor.globo.com/brasil/noticia/2024/05/10/brasil-quer-ser-exemplo-em-uso-de-inteligencia-artificial-pelo-judiciario.ghtml Aesso em nov. 2024

contenham informação sensível e de estratégia dos MPs. Então, é possível que venha a desenvolvê-las, mas por enquanto veta o uso, porque qualquer assistente de Inteligência Artificial que seja de propriedade de terceiros não pode ter acesso a informações que estão nos processos e investigações.

Em outubro de 2019, houve uma primeira versão da regulamentação do Governo Federal para os investimentos em ferramentas de I.A. com os recursos do Fundo Nacional de Segurança Pública. No entanto, nos últimos anos, as prisões por engano têm sido registradas pela imprensa e, segundo dados da Rede de Observatórios de Segurança e também o The Intercept[11], 90,5% das pessoas presas por Reconhecimento Facial (RF) no Brasil são negras. As máquinas são calibradas com uma margem de semelhança dos traços biométricos (que pode ser até mesmo o jeito de caminhar, um método aplicado em Pequim e Xangai desde 2018), e a partir desse grau de semelhança admite alguma taxa de erros.

O reconhecimento facial está assim em demais aplicações, ainda que esteja já popularizado. A biometria é feita com uma fração da amostra: uma parte do rosto é analisada e comparada com outras amostras, aceitando uma margem de erro. Se o programador admite uma margem muito restrita, fica quase impossível apresentar um resultado. Se o programador amplia essa margem, irá receber mais indicações da máquina.

11 Retratos da Violência: Cinco meses de monitoramento, análises e descobertas. Rede de Observatórios de Segurança, 2019. https://observatorioseguranca.com.br/wordpress/wp-content/uploads/2019/11/1relatoriorede.pdf Acesso em nov. 2024

Não há, no Brasil, dados oficiais sobre abordagens e prisões por engano baseadas em RF, mas sabe-se que acontecem e resultam em constrangimento e injustiça. Para o The Intercept, não sabe como se deram essas prisões. "Tentamos obter dados, via Lei de Acesso à Informação (LAI), sobre a quantidade oficial de prisões e o número de pessoas abordadas de forma equivocada, mas não houve retorno do pedido", relata Pablo Nunes, coordenador adjunto do Centro de Estudos de Segurança e Cidadania (CESeC) e da Rede de Observatórios da Segurança Pública, na reportagem de 2019 – "além de ineficiente, o sistema agrava o encarceramento de negros".

O Panóptico mantém um observatório de projetos de I.A. em vigilância. Veja os estados brasileiros que mais investiram em ferramentas de I.A. na área de segurança neste momento:

Região	Gastos (milhão)	Projetos ativos	Estado com mais projetos ativos	Estado que mais gastou	Quanto gastou (milhão)
N	R$ 88,4	24	AM (5)	AC	R$ 61,6
NE	R$ 880,5	45	PE (8)	BA	R$ 665,4
CO	R$ 161,7	84	GO (70)	GO	R$ 111,3
SE	R$ 568,2	93	SP (39)	RJ	R$ 435,7
S	R$ 40,9	55	PR (21)	RS	R$ 21,3
BRASIL	**R$ 1,7 bi**	**301**	**GOIÁS**	**BAHIA**	**R$ 1,2 bi**

Fonte: O Panóptico[12].

12 Monitoramento dos municípios. O Panóptico. https://www.opanoptico.com.br/

Nas ruas, o RF é como uma ferramenta que deveria estar ainda em testes, mas não em uso. O Panóptico, observatório do CESeC, monitora o uso de sistemas de reconhecimento facial no Brasil na área de segurança pública. Seus dados apontam que essas ferramentas são usadas sem transparência, sem a perspectiva de ter uma prestação de contas, e a sua regulamentação é, ainda, rudimentar ou experimental. No relatório "Vigilância por lentes opacas"[13], feita com as instituições do Brasil, em cidades e estados que utilizam o RF para área de segurança pública, ele conclui que "nenhum órgão confirmou ou reconheceu a existência de erros" e que "a ausência de responsabilidade formal pela segurança pública nos municípios implica que eles não são devidamente cobrados ou responsabilizados pela gestão e transparência dessas tecnologias".

O jornalista Pablo Nunes também produziu um relatório para O Panóptico ("Um Rio de câmeras com olhos seletivos"[14], 2022), no qual apenas via LAI pôde obter dados da Secretaria de Estado da Polícia Militar do Rio de Janeiro (SEPM). Nos dados, nenhuma pessoa desaparecida tinha sido encontrada na primeira fase de aplicação do RF colocada em prática no RJ.

13 Vigilância por lentes opacas: mapeamento da transparência e responsabilização de projetos de reconhecimento facial no Brasil. O Panóptico, 2024. https://lapin.org.br/wp-content/uploads/2024/10/OPANOPTICO_Pesquisa_Vigilancia_Por_Lentes_Opacas.pdf Acesso em nov. 2024

14 Um Rio de câmeras com olhos seletivos. O Panóptico, 2022. https://cesecseguranca.com.br/wp-content/uploads/2022/05/PANOPT_riodecameras_mar22_0404b.pdf Acesso em nov. 2024

Questionada em relação a falsos positivos em operação de capturas em uma partida de futebol no Maracanã, o órgão informou que não coletou dados de falsos positivos nem de abordagens. Apenas com recurso via Lei de Acesso à Informação, soube-se que "a SEPM admitiu que dentre os 11 casos de pessoas detidas com o uso da tecnologia de reconhecimento facial do Maracanã, sete foram erros da máquina, ou seja: falsos positivos".

Nos Estados Unidos, a conversa sobre uso da I.A. na segurança pública já está em um patamar que não tem mais a ver com usar ou deixar de usar, mas qual direcionamento seguir, se é sobre fatos consolidados ou sobre o futuro. É comum o uso de assistentes de análise de riscos no Judiciário. Uma ferramenta preditiva usa o sistema de *scoring* para dizer, numa escala de 1 a 10, a chance de uma pessoa continuar cometendo crimes.

O COMPAS (nome em alusão a compass, bússola) é um modelo de I.A. feito para dar assistência a decisões de juízes de vários estados dos EUA, incluindo a Califórnia, que classifica a pessoa em baixo, médio e alto risco de reincidência criminal. Esse tipo de ferramenta de apoio às sentenças é reprovado por pesquisadores da Universidade de Houston, New Hampshire e outras, e COMPAS foi escrutinada por uma longa reportagem da ProPublica, "Machine Bias"[15] (2016). No entanto, este e outros sistemas assistentes similares continuam em uso.

15 Machine Bias. ProPublica, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing Acesso em nov. 2024

A reportagem expõe uma relação inversa na classificação: quase o dobro de pessoas negras foram rotuladas como "alto risco" pela máquina em relação às brancas, mas não reincidiram. Pessoas brancas que reincidiram tinham mais chance de serem classificadas como "baixo risco".

A predição falha de forma diferente para réus negros		
	Brancos	Negros
Classificado como alto risco, mas não reincidiu	23,5%	44,9%
Classificado como baixo risco, e reincidiu	47,7%	28,0%
Fonte: ProPublica, dados de Broward County, Florida.		

A reportagem relata que é difícil não incluir à raça fatores de "pobreza, desemprego, marginalização" – e segundo uma fonte, "se esses fatores são omitidos da avaliação de riscos, a precisão vai abaixo".

Em 2014, o procurador-geral dos Estados Unidos, Eric Holder, alertou para o risco desses sistemas enviesados no Judiciário e incumbiu à Comissão de Penas dos Estados Unidos avaliar o uso. Resultado: a comissão não avaliou. Tanto essa experiência dos Estados Unidos como a do Brasil evidenciam o descompasso entre regulação, accountability e desenvolvimento tecnológico.

“Existe um descompasso”

Em um artigo da doutora Denise Utochkin, professora do Departamento de Saúde Pública da Universidade de Copenhague, traduzido e compartilhado pelo professor jornalista Marcelo Soares, da agência Lagom Data[16], questiona-se: “por que continuamos a acreditar que a I.A. vai resolver a crise climática (visto que ela está ajudando a piorar), acabar com a pobreza (da qual ela depende muito) e libertar o pleno potencial da criatividade humana (que ela está minando)?”

Existe um discurso pronto de que colocamos em uso ferramentas que ainda precisam de tempo para amadurecer – ao mesmo tempo que se não formos usá-las, não amadurecerão, e se já temos esse conhecimento, não podemos deixar de usar. A professora argumenta que apenas se acumulam os indícios de que o amadurecimento da I.A. é uma falácia.

“Efetivamente existe um descompasso”, afirma a professora Dora Kaufman sobre a disparidade entre a desenvoltura das mudanças tecnológicas e a sua regulação. Ela declara que flexibi-

16 Lagom Insights: a ilusão futurista da IA. Denise Utochkin e Lagom Data, 2024. https://buttondown.com/lagomdata/archive/lagom-insights-a-ilusao-futurista-da-ia/ Acesso em nov. 2024

lidade é uma chave, "por isso sempre defendi que o protagonismo de regular e fiscalizar a I.A. seja das agências setoriais". Segundo ela, as agências conhecem o seu domínio. Elas têm como missão "justamente regular e fiscalizar, logo são estruturadas para tal, e têm mais flexibilidade para mudar, adaptar, rever as normas".

Atualmente, está se constituindo uma diretriz para desenvolvimento, compra e uso das tecnologias de I.A. bastante esparsa. Nela, as empresas, instituições e conselhos formulam seu projeto para uso da I.A. mediante as necessidades e perfis próprios. Para os sistemas de administração e governança do Brasil, o Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028 foi apresentado na 5ª Conferência Nacional de Ciência, Tecnologia e Inovação em julho de 2024.

O plano tem o objetivo de "transformar o Brasil em um modelo global de eficiência e inovação no uso de I.A. no setor público". A ministra do Ministério da Gestão e da Inovação em Serviços Públicos (MGI), Cristina Mori, afirmou na ocasião que deve ser pautado por valores de justiça e equidade, e "esse tema da Inteligência Artificial requer que a gente desenvolva capacidades estatais para lidar com ele".

Vai ao encontro dessa fala as instruções, a nível de uso pessoal, que Alexandre Gonçalves destacou como formas de mitigar o enviesamento das tecnologias generativas. "O uso ético e responsável que o jornalista ou outro usuário da I.A. faz é ponto central neste contexto." Devemos, segundo ele, saber usá-las, colocando-as no lugar de coautoras, configurando adequadamente o banco de dados em que a máquina vai operar, se for possível, e trabalhar em cima dos resultados compreendendo que

os outputs não devem ser usados do jeito que saem da máquina.

Mas segundo o artigo já mencionado de Rodrigo Brandão e Glauco Arbix ("Tecnologias de reconhecimento facial na administração pública brasileira"), ainda existe um despreparo da parte humana para receber e utilizar a I.A. nas instituições públicas. Ela deveria ser incorporada por um pessoal previamente ciente de como essas ferramentas devem funcionar e como devem ser utilizadas.

Segundo eles, de modo geral, os municípios não têm clareza sobre como podem e devem preparar os funcionários públicos envolvidos com a utilização da tecnologia, nem para que utilizem a Lei Geral de Proteção de Dados (LGPD) a favor da transparência algorítmica ou para que utilizem a intervenção humana para corrigir resultados enviesados da tecnologia.

No artigo, estudos identificaram que os operadores de políticas públicas enfrentam dificuldades para contestar os outputs, "fazendo com que estes se comportem como verdadeiros tomadores de decisões públicas, ao invés de apenas subsidiá-las." O professor Glauco também comentou sobre esse assunto, porque o despreparo do Estado, de quem faz os acordos e licitações, é evidente.

Os órgãos públicos compram o serviço e delegam para a empresa terceirizada o controle. "Você pergunta para eles: está tendo erros? 'Não sabemos'. O poder público não fiscaliza? Dizem 'ah, nós nem sabemos como é feito… A gente não pede relatório, a empresa também não manda.'"

Para Rodrigo, primeiramente, os agentes públicos devem avaliar se a ferramenta é realmente necessária. Assim seria

possível evitar a adoção de tecnologias baseadas em discursos prontos de smart cities influenciados por interesses empresariais, sem reflexão crítica. Sem essa avaliação, o poder público pode adquirir e desenvolver produtos que não contribuem para a melhoria efetiva para as pessoas. Pelo contrário, podem aprofundar injustiças sociais.

A IBM criou uma extensão para detectar viés preconceituoso em resultados gerados por modelos LLM, chamada "AI Fairness 360", aberta e aplicável. Segundo o desenvolvedor, "pode ajudar a examinar, reportar e mitigar discriminação e preconceitos em modelos de machine learning pelo ciclo de vida completo de uma ferramenta de I.A.". Além disso, a OpenAI oferta abertamente dez algoritmos de controle de parcialidades e preconceitos para ferramentas de IA. Se é possível desenvolver iniciativas de equiparação, por que o processo é tão estagnado?

"Diferente da maioria das criações humanas, nós não entendemos totalmente as operações internas das redes neurais" a OpenAI explicou, em anúncio da nova versão do seu famoso chat, o GPT-4o. A culpa é da tal caixa-preta. "Em vez disso, nós desenhamos os algoritmos que as treinam. O resultado dela não é compreensível, não pode ser facilmente decomposta em partes identificáveis", conclui o boletim.

Portanto, a complexidade das redes neurais apresenta um desafio para a aferição e reparo dos parâmetros equivocados em um modelo de LLM. Como a máquina observa os dados que recebe e estabelece as relações, comparações e entende os padrões sozinha, são milhões de processos (sobre processos e mais processos) sobre os quais se desenvolve a sua aprendiza-

gem. Portanto, não é possível saber o caminho que ela tomou para chegar à conclusão expressa na sua resposta.

Até o começo de 2024, a empresa possuía uma equipe própria especializada, que se chamava "Superalignment". "Nós necessitamos de avanços técnicos e científicos para direcionar e controlar sistemas de I.A. muito mais espertos do que nós", a OpenAI descreve na carta de abertura do grupo sobre a preocupação com o ASI – Superinteligência Artificial[17]. A equipe era dirigida por Ilya Sutskever (cofundador) e Jan Leike (pesquisador pioneiro da I.A.). Os dois pediram demissão, sendo que Jan acusa a OpenAI de deixar a cultura de segurança e segurança dos processos de lado. Com as saídas, a empresa dissolveu a Superalignment.

Além disso, o GPT-4o é uma versão avaliada por um "red team", ou time vermelho – uma equipe autorizada a explorar as vulnerabilidades do sistema como uma adversária, para torná-lo mais seguro. O release da versão menciona que o teste foi feito apenas para as atualizações da versão: "o GPT-4o passou por uma extensiva análise com mais de 70 experts externos em domínios como psicologia social, preconceito, injustiça e desinformação, para identificar riscos que foram introduzidos ou ampliados pelas novas modalidades adicionadas."

A equipe foi composta por especialistas selecionados por chamada aberta em 2023. "Continuaremos a mitigar novos riscos

17 Introducing Superalignment. Release de imprensa da OpenAI. https://openai.com/index/introducing-superalignment/ Acesso em nov. 2024

assim que descobertos", o informativo conclui. O fato de que a OpenAI possa ouvir os peritos externos para controle dos riscos não deixa evidente se deva ter o seu próprio corpo de especialistas interno, mas, de qualquer forma, os pareceres dela ou da falecida Superalignment não são abertos. Portanto, não é possível ter acesso ao que foi apontado e ao que foi suprido.

Em artigo da Wired ("OpenAI's Long-Term AI Risk Team Has Disbanded"[18]) sobre a saída dos cofundadores, não há indícios de que as saídas tenham relação com os esforços da OpenAI para desenvolver uma tecnologia que se pareça cada vez mais com os humanos ou com a venda de produtos – ainda que Jan Leike tenha dito que foi por causa disso. "Contudo, os últimos avanços levantam questões éticas sobre privacidade, manipulação emocional e riscos de cibersegurança."

Além disso, especialistas e pesquisadores em Direito e Cibersegurança criticam a corrida de avanços da tecnologia sem o devido reparo do que já foi construído até aqui. Em artigo do professor Peter Nagy Salib na revista multimídia Lawfare[19], atualmente nada impede a I.A. de auxiliar hackers e bioterroristas.

18 OpenAI's Long-Term AI Risk Team Has Disbanded. Wired. https://www.wired.com/story/openai-superalignment-team-disbanded/ Acesso em nov. 2024

19 OpenAI No Longer Takes Safety Seriously. Coluna de Peter Nagy Salib na Lawfare. https://www.lawfaremedia.org/article/openai-no-longer-takes-safety-seriously Acesso em nov. 2024

"I.A. deturpada ainda não é uma ameaça aos humanos, só porque o GPT-4 ainda é medíocre como hacker e bioterrorista. Mas a OpenAI e seus concorrentes estão correndo como podem para desenvolver sistemas tão autônomos e capazes quanto seja possível", conclui.

"Fogo!"

Há vários países que vêm participando da redefinição dos parâmetros, da forma e desenvoltura da guerra, envolvendo-se na escalada de tecnologias de I.A. autônomas para segurança e defesa – assim como foram as inovações tecnológicas anteriores: a forja de espadas, a explosão da pólvora, a bomba atômica. Mas como essa escalada acontece se os limites de ética mínima para o uso da I.A. na guerra ainda estão sendo constituídos enquanto ela está sendo experimentada em campo?

No Brasil, a inclusão de tecnologias de Inteligência Artificial nas Forças Armadas está primeiramente criando um ambiente favorável para o desenvolvimento e aplicação desses sistemas, mas ainda não existe um campo muito amplo. As aproximações entre defesa, segurança e essa tecnologia estão ainda em um estágio inicial.

Em Santa Catarina, a Federação das Indústrias de Santa Catarina (FIESC) sediou a SC Expo Defense, em maio de 2024, uma feira com objetivo de suscitar um ambiente de inovação no setor de segurança e defesa. Todos os expositores consultados na feira, incluindo Exército, Marinha e Aeronáutica, expressaram desconhecer as aplicações da I.A. atuais nas suas instituições, ou que ainda estavam no início e não eram largamente aplicadas.

Nas conferências, os palestrantes que apresentaram projetos de Inteligência Artificial (na área de gestão e de meteorologia

para prevenção e mitigação de desastres) destacaram a importância de começar a investir e dominar a técnica.

Em abril, o Estado-Maior do Exército (EME) tinha acabado de publicar a Portaria Nº 1.318, de 14 de abril de 2024, para aprovar a diretriz estratégica de Inteligência Artificial do Exército Brasileiro[20]. Uma das premissas é a de que foi identificada "a necessidade de impulso estratégico nessa área, direcionando e aglutinando iniciativas, a fim de permitir que a Instituição obtenha vantagem competitiva (eficiência, eficácia e efetividade) em todas as áreas de atuação".

O texto do EME determina o objetivo de desenvolvimento da I.A. de forma abrangente e que "deverá ser prevista a capacitação de militares e sua preparação para a reestruturação da organização e das missões, à medida que a I.A. é implantada". Menciona também sistemas com objetivo voltado à gestão e eficiência dos processos, sendo que "toda e qualquer proposta de sistema de I.A. que não possibilite o controle humano (supervisão) ou que não seja previsto ou possível estará sujeita à avaliação pelo EME".

O texto destaca a necessidade de que estes sejam sistemas construídos com o ciclo de supervisão e revisão de um humano, de método "*human-in-the-loop*", especificamente naquelas ferramentas que podem ter relação com ataques. Descreve ali que "ao se implementar um sistema com I.A., deve ser prioritária a

20 Estado-Maior do Exército do Brasil: Portaria Nº 1.318, de 14 de abril de 2024. http://www.sgex.eb.mil.br/sistemas/boletim_do_exercito/copiar.php?codarquivo=181261825&act=sep

garantia de que humanos permaneçam no controle do uso da força, em especial, no caso de decisões sobre vida ou morte em combate". A Portaria também menciona o desenvolvimento de sistemas de alto risco, como armas, mas são projetos com visão de longo prazo e atrelados a processos específicos de avaliação.

"Em vez de se apressar no uso de I.A. para armamentos autônomos, o país dá prioridade ao desenvolvimento de soluções que fortaleçam a gestão, a logística e a capacidade de tomada de decisão", comenta o major engenheiro da Aeronáutica Daniel Baggio, que é assessor de Fomento à Pesquisa e Inovação (AFPI), no Centro de Computação da Aeronáutica, em São José dos Campos (CCA-SJ), do Departamento de Ciência e Tecnologia Aeroespacial (DCTA). "Isso posiciona o Brasil entre os países que preferem aguardar maior maturidade no debate ético e normativo antes de avançar em aplicações bélicas autônomas."

Para isso, os sistemas de I.A. devem ter um rol de procedimentos e avaliações que garantam a sua segurança. Estão, entre elas, o controle humano, proteção contra ameaças cibernéticas, tecnologias nativas de formatos próprios da instituição e do país, que possua relatórios de impacto e conformidade com a LGPD. A Portaria também define que a governança da I.A. no Exército seja de responsabilidade do EME, e que deve ter uma capacitação ética específica para I.A. (ainda que não estabeleça os critérios para isso).

"O perfil do Brasil na área de tecnologia para serviço e defesa nacional reflete um posicionamento cauteloso", sintetiza o pesquisador, e explica que priorizamos o controle humano sobre sistemas de Inteligência Artificial – especialmente em decisões

que envolvam vida ou morte, como o uso da força. "A diretriz estratégica enfatiza que toda decisão baseada em I.A. deve ser rastreável, explicável e sujeita à validação humana", que segundo ele, reforça o compromisso com a ética.

"O panorama da Inteligência Artificial nas forças militares e de defesa está se construindo de forma desigual ao redor do mundo", comenta. O Brasil concentra seus esforços no uso de I.A. para otimizar processos administrativos e operacionais, mas este não é um posicionamento adotado sempre por outros países. Na verdade, existe uma competição mais agressiva e com objetivos sólidos entre aqueles que estão em conflito aberto ou que têm a previsão no horizonte de médio e longo prazos.

A necessidade de fortalecer-se com o uso da nova tecnologia pode ser uma das forças motoras para a corrida por instrumentos de I.A. militares, porque, ao contrário dos países que têm interesse evidente em se preparar para conflitos armados em vista, o Brasil está começando a se situar no ecossistema de I.A. Em sua tese, publicada em novembro de 2023 ("Desafios para adoção de Inteligência Artificial na Força Aérea Brasileira"[21], Fundação Getúlio Vargas), o militar e pesquisador percebe que "há uma notável discrepância entre a abundância de produção acadêmica na área de I.A. na FAB e sua aplicação prática concreta". As razões apontadas para isso são a ausência de uma estratégia para emprego da I.A., falta de entidades específicas

21 Desafios para adoção de Inteligência Artificial na Força Aérea Brasileira, tese de Daniel Baggio. FGV, 2023. https://gist.github.com/dannyxyz22/fb7bab959af603e37c250dd6028831e9 Acesso em nov. 2024

para pesquisa e aplicação, e falta de investimento financeiro para essa finalidade.

Atualmente, já existe um observatório para essa área de atividade. O Defense AI Observatory (DAIO)[22], da Universidade Helmut Schmidt de Hamburgo, publica, a cada mês, um relatório sobre o uso de I.A. pelas forças armadas e forças de defesa de cada país. No entanto, os relatórios dependem fortemente de fontes oficiais, e há lacunas para que essas informações possam ser aferidas e testadas, ou até colocadas em contraste com relação aos fatos.

Entre esses relatórios, há alguns que apontam para uma competitiva agressividade, e outros, de países cujo perfil seja mais defensivo, de aperfeiçoamento administrativo e de inteligência. É possível demarcar o perfil de desenvolvimento estratégico-militar para os países analisados, mas não o quanto e como. Quase não há transparência com relação ao orçamento de investimentos na área de I.A. militar, sem saber sequer o valor total investido especificamente em I.A. No entanto, há três maiores preocupações apontadas neste contexto geopolítico: a guerra entre Rússia e Ucrânia; a guerra entre Israel e seus vizinhos; e uma intimidação chinesa.

Na Europa, conta Daniel, "a Bundeswehr [Defesa Federal da Alemanha] concentra-se em aumentar a capacidade de sobrevivência, em vez da letalidade" e "a França, por outro lado, busca se tornar líder mundial em I.A., com uma estratégia nacional

22 The Defense AI Observatory (DAIO). Helmut Schmidt University em Hamburgo. https://defenseai.eu/english

lançada em 2017", "mas enfrenta desafios na integração da I.A. com os sistemas de defesa existentes". Entre os relatórios do DOIA, um dos países que traz a questão ética de forma mais dirigida na agenda é a Finlândia, que teve a sua situação geopolítica mudada com a Guerra da Ucrânia e agora é um novo membro da Organização do Tratado do Atlântico Norte (OTAN).

"A Força de Defesa Finlandesa utiliza uma organização matricial para orientar a implementação de aplicações de I.A.", comenta o militar. Segundo o relatório do DAIO, o país mostra a intenção de regular antes de usar, criando um ecossistema para a implantação.

"Ética e regulamentação da I.A. na defesa têm sido reconhecidas como indispensáveis", mas as diretrizes para uso de armas letais autônomas ainda estão sendo construídas em uma guia proposta por especialistas nacionais. "A máquina fará um papel de suporte ou de execução." Sendo que "uma margem de incertezas vai permanecer, isso requer um trabalho humano", mas "o comandante da tropa sempre carregará a responsabilidade da decisão, não importa se a tropa é composta por máquinas, pessoas ou ambas."

O Japão vive a expectativa de um possível conflito com a China, ainda que não haja nada, de fato, em vista para os próximos anos. O país possui restrições desde a Segunda Guerra Mundial e "vive à sombra desse passado", por isso não possui forças armadas. Segundo o relatório, a Força de Autodefesa do Japão (FAJ) não pode desenvolver diretamente armas de I.A., mas o interesse nessa tecnologia é crescente, e tem sido cada vez mais mencionado em documentos públicos desde 2022. Os

autores dizem que não puderam encontrar nenhum documento que objetivamente mostre o Ministério de Defesa e a FAJ se preparando para o advento de uma I.A., "no entanto, a FAJ faz recrutamentos públicos para cargos seniores com menção à I.A." e abre vagas de trabalho com essa palavra-chave para a Marinha e Aeronáutica, desde 2023.

O relatório afirma que a corrida tecnológica dos E.U.A. acontece em resposta aos avanços tecnológicos da China. Já foram criados aparatos e organizações para capacitar a adoção da I.A., inclusive uma guia ética. No entanto, o país tem ainda dificuldade de fazer uma transição completa das suas forças armadas, de informatizadas para inteligentes. Primeiro por ter um sistema burocrático lento e atrasado com relação à necessidade de rapidez que essa transformação pede, e, em segundo lugar, por falta de pessoal especializado.

"Apesar de os Estados Unidos serem atrativos para a rede global de talentos em I.A., o setor público falhou em competir com a academia e indústria." O major Daniel afirma: "a Marinha americana prevê que até 2045, um terço de seus navios de guerra serão "navios fantasmas" robóticos, o Pentágono está testando a plataforma de I.A. FedLearn para aprendizado federado – essas iniciativas demonstram um forte investimento e foco na aplicação da I.A. em diversas áreas da defesa".

O DOIA mostra que, em 2017, a China revisou a base nacional do currículo para o ensino médio, incluindo a I.A. na grade, e, em 2018, essa mesma regra passou a valer para instituições de ensino superior. "De acordo com uma estimativa feita, já em 2019 cerca de 90% dos talentos em I.A. da China eram

domésticos." Ainda assim, o país está mais no processo de informatização do que no de inteligência, segundo mostra o relatório. A maioria dos avanços ainda é no campo teórico.

"Não há evidências sugerindo que a China está em um ponto que possa ampliar muito e de forma decisiva a vantagem de liderança sobre os Estados Unidos e seus aliados no uso militar da I.A." No entanto, o governo dos E.U.A. "parece estar totalmente convencido de que a emergente capacidade de I.A. defensiva representa uma ameaça que requer medidas severas", o que resultou no controle de exportação de semicondutores para a China em outubro de 2022.

Na questão ética, o presidente chinês Xi Jinping pessoalmente enfatizou que a I.A. precisa ser "segura, confiável e controlável", em 2018, numa conferência no Politburo, o órgão de decisão intermediário entre o Comitê Permanente e o Comitê Central do Partido Comunista Chinês.

Uma preocupação, de fato, é que existe uma tendência no mundo de aumento da automação, transformando modelos de HITL em DITL: o sistema HITL ("*human-in-the-loop*", humano no controle) é quando o ciclo de trabalho de um modelo de I.A. inclui input humano para o seu resultado, e o sistema DITL ("*data-in-the-loop*", dados no controle) é quando esse input vem de um banco de dados – em outras palavras, o primeiro mantém um fator humano e o segundo é autônomo. A expectativa para a China é que essa progressão seja feita apenas nas ferramentas de tomada de decisão para níveis inferiores, mantendo os níveis altos de comando com um supervisor humano.

Mesmo assim, sistemas que mereciam uma supervisão

humana atenta estão passando a ser negligenciados, transformando o assistente em ator, ou, ainda, estão mesmo sendo deliberadamente automatizados. Em 2023, as Forças de Defesa de Israel (FDI) colocaram em campo na Faixa de Gaza dois sistemas de I.A. que, em tese, deveriam ter uma intensa avaliação humana para estar em funcionamento, como exposto pelo jornalista israelense Yuval Abraham na +972 Magazine[23].

Um desses sistemas de I.A. irá prover a lista de alvos humanos: ele se chama "Lavender"; e o outro, que irá localizá-los, se chama "Where's Daddy?" (Cadê o papai?). O ataque é feito com bombas de queda livre, imprecisas e de baixo custo (*dumb bombs)*. O professor Glauco Arbix comenta a reportagem, dizendo que muita gente até pensava que os ataques eram indiscriminados, "e até que foram, mas, concretamente, eles estavam sendo orientados em função de sistemas".

O governo israelense modificou, segundo a reportagem, os parâmetros para regular a margem "aceitável" de mortes colaterais – que são aquelas não relacionadas ao alvo, entre todos os civis. Segundo as fontes da reportagem de Abraham, essa margem se perdeu.

Para eliminar um comandante, que teria o parâmetro de

23 - a) 'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine. https://www.972mag.com/lavender-ai-israeli-army-gaza/
23 - b) 'Order from Amazon': How tech giants are storing mass data for Israel's war. +972 Magazine. https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/

dez óbitos civis para alcançar essa morte em guerras passadas, passou a ser de mais de cem. Para eliminar membros juniores e de baixa patente, tornou-se aceitável a morte de 20 pessoas sem qualquer atividade militar. "Todo alvo está sujeito a ter um dano colateral gigantesco", diz Glauco.

"Nas primeiras semanas da guerra, o exército se baseou totalmente no Lavender, que viu uns 37.000 palestinos como suspeitos militantes – e as suas casas – possíveis alvos de ataques aéreos", relata fonte na reportagem de Yuval. O Lavender analisa os dados sobre a população e pontua para cada pessoa um *score* de 1 a 100 - sem zero. Esses dados são tais como fotos, contatos, endereços, familiares, entre outros, sendo consideradas atividades suspeitas, por exemplo, ter mais de uma linha de telefone no nome ou mudanças de endereço em um ano.

Explica o professor: "um dia, se você foi do Hamas, dez anos atrás, você tem pontos. Se você deixou, não interessa. Se foi num casamento, se conhece alguém, se foi num jantar que tinha alguém do Hamas à mesa, você tem ponto". Dependendo do seu *score*, você se tornará estatístico o suficiente para a I.A. recomendar um ataque. "A partir de um certo ponto você não tem mais checagem, não tem mais autorização humana para bombardearem. Isso é terrível."

Aí entra em ação a segunda ferramenta, Where's Daddy?, que recebe esse nome porque encontra o alvo quando está chegando em casa, exposto e vulnerável junto à família. O ataque é feito com as bombas de queda livre, que são mais baratas e não têm controle, arrasando uma grande superfície. Por isso, os bombardeios eram em áreas residenciais e com muitas vítimas.

"In the bombing of the commander of the Shuja'iya Battalion, we knew that we would kill over 100 civilians," B. recalled of a Dec. 2 bombing that the IDF Spokesperson said was aimed at assassinating Wisam Farhat. "For me, psychologically, it was unusual. Over 100 civilians — it crosses some red line."

[...]

"No bombardeio ao comandante do Batalhão de Shuja'iya, nós sabíamos que aquilo mataria mais de 100 civis", B. [fonte anônima] recorda do bombardeio de 2 de dezembro que o representante da FDI disse que tinha o objetivo da morte de Wisam Farhat. "Para mim, psicologicamente, isso é não é normal. Mais de 100 civis – isso viola um limite."

'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine, 2024.

Atualmente, políticas antiterroristas e colonialistas se confundem. Esse fenômeno de usar a morte para afirmar a soberania é descrito pelo professor e cientista político Achille Mbembe, em Necropolítica (2011), citando expressamente as políticas sionistas: "a forma mais bem sucedida de necropoder é a ocupação colonial contemporânea da Palestina"[24]. Resta uma terra arrasada, pronta para ocupação e anexação.

O rascunho de uma ferramenta como o Lavender está no livro "The human-machine team: how to create synergy between human & artificial intelligence that will revolutionize our world"[25] (2021), escrito sob pseudônimo Y.S., do oficial Yossi Sariel, e que era então brigadeiro-general, chefe da Unidade 8200 (Corpo de Inteligência das Forças de Defesa de Israel). Ele foi premiado pelo governo por um projeto de sistema de Inteligência Artificial antiterrorismo, que, a partir da extensa descrição publicada em seu livro, julga-se ser o precursor do sistema usado em 2023.

O que o comandante explica em seu estudo é que não há no horizonte da segurança nacional, para qualquer país, um futuro sem a I.A. Ele também apresenta as diferenças de abordagem entre o controle dos seus inimigos através do medo, o combate em campo e o ataque de "lobos solitários", como se

24 Necropolítica. Achille Mbembe. N-1 edições, 2018.

25 The human-machine team: how to create synergy between human & artificial intelligence that will revolutionize our world. Brigadier General Y.S. eBookPro Publishing, 2021.

chama o ataque de uma pessoa sozinha que concebe e leva a cabo uma ação terrorista, em geral midiatizada através de rede social.

No seu livro, ele fala que um futuro ideal da segurança nacional só é possível com o pleno desenvolvimento de uma I.A. militar. Assim, será possível detectar ameaças imediatas e intervir, ou criar milhares de alvos por dia suficientes para dominar uma guerra aberta.

Ele promete que esses alvos seriam cada vez mais precisos, reduzindo os danos colaterais contra os civis: "imagine a realidade em que os militares tivessem a habilidade de acertar os alvos certos na hora certa, destruindo o alvo com o menor dano colateral possível". A "Deep Defense", que é como ele chama a defesa nacional aliada à *deep learning* da I.A., superaria as limitações humanas para estratégia de guerra e segurança interna.

Na visão dele, os humanos são gargalos que impedem a ação das forças: "não conseguimos processar tanta informação". Os humanos emperram os processos de criar alvos, localizar, fazer o processo de decisão para aprovar, e concluir o processo desde o setor de inteligência até o "fogo!". "A parceria Humano-Máquina tem o potencial de trazer uma revolução nas possibilidades de criação de 'alvos contextualizados'. Uma equipe feita de máquinas e investigadores pode arrebentar um gargalo."

Sariel justifica a escalada da guerra no campo tecnológico da área de I.A. por entender que, se Israel não o fizer, seus "inimigos" o farão. E também descreve um sentimento que os militares possuem, segundo ele, de tarefa incompleta: "nos últimos 20 anos, as forças de Israel concluíram missões e guerras com o sentimento de "*hachmatza*" (hebraico do que quer dizer uma

‘oportunidade perdida’)”. Há vitórias que deixaram um gosto de derrota porque sentem que foi a chance de eliminar os adversários e não a aproveitaram efetivamente, por isso não devem deixar escapar uma próxima ocasião. Ele encerra o livro: “Fim. Ou, na verdade, só o começo do futuro”.

Meio Golem e meio Talos, a I.A. vai crescendo. No quadro geral, é uma tecnologia necessária, mas que está se constituindo como uma nova indústria, orientada por interesse de atores dominantes em uma sociedade desigual, e em que a regulação ainda é esparsa.

Terminando por antes do começo

Na forja de Hefesto, das eternas faíscas, o deus grego da tecnologia criou um autômato gigante de bronze para defender a ilha de Creta, chamado Talos: ele fazia rondas, atirando grandes rochas contra naus inimigas. Já Afrodite trouxe à vida Galateia, a estátua da mulher perfeita, esculpida e amada pelo rei Pigmalião.

Para os hindus, Brahma teve filhos nascidos do seu pensamento, os Manasaputras, que vieram a criar os humanos e arquitetar o mundo. Para os árabes, os Gênios eram seres cheios de conhecimento, que respondiam a qualquer pergunta. Para os judeus, o Golem é um gigante de barro vivificado para proteger o povo, mas ele não tem discernimento e fica sem controle.

Ainda que o Golem – sem a mesma sorte dos Gênios – seja tolo, todas essas histórias mostram formas de inteligências artificiais (com letra minúscula). Muito antes da invenção da informática, esses mitos já davam uma ideia do que viríamos a

chamar de robô, programa, buscador e até antivírus.

Assim como as ferramentas da computação, eles podem tomar decisões baseados num conjunto de informações e então responder, proteger, construir e até emular sentimentos, em capacidade similar ou superior à dos humanos. Sobre a Terra, as gerações foram amadurecendo esse sonho até que, na história do tempo presente, obtivemos condições para concretizar e tornar isso, de fato, um produto.

Então aquela Inteligência Artificial (I.A.), com letra maiúscula, se refere a um campo de estudos multidisciplinar da computação, batizado em junho de 1956. Nessa área, surgiu uma tecnologia que simula a cognição humana, aprendendo conforme experimenta o mundo e tentando não repetir as falhas. Em termos mais materiais, ela recebe grandes volumes de dados na forma binária e os analisa para devolver resultados, aprimorando-os conforme recebe mais dados ou feedbacks positivos e negativos.

A I.A. supera a computação tradicional, e esta revolução que ela impulsiona não pode ser retroagida. Para uma tecnologia honesta e que tenha compromisso com os direitos humanos, é necessário que sumariamente esta tenha supervisão humana, transparência e que seja segura. No entanto, neste momento atual, ela tem métodos insondáveis, os usuários cada vez mais delegam a ela as suas próprias decisões, e o que se vê é que ela tem a tendência, assim como os humanos, a acirrar as diferenças e desigualdades.

Faltando propósito, afeto e discernimento ao Golem, o seu destino foi voltar a um senhor obscuro. Tendo seu interior

frágil, Talos caiu na loucura, ferindo a si mesmo. Talvez, ainda, tanto os golens quanto Talos possam dizer algo para o século XXI, porque o elemento mais fundamental dessas histórias é o humano: por um lado, a transparência, responsabilização e supervisão da máquina, quando feita por pessoas, garante o seu propósito e discernimento; por outro, quanto mais bem tratado for aquilo que mantém a tecnologia ativa, que são as pessoas que fazem ela acontecer, menor será a chance de enlouquecer.

Então, se pudermos ver o quanto esse desejo humano de ser como um criador está vivo neste momento, talvez exista algo a aprender com o que já está no passado, mas não da forma que os eruditos fazem – na forma das crianças. Talvez com as perguntas mais francas sobre as inteligências artificiais seja possível ter as respostas mais honestas.

Fontes consultadas

As fontes consultadas sem referência direta são todas aquelas pessoas que não estão indicadas dentro do texto, mas que foram essenciais para a produção de resultados, a melhor compreensão do assunto e a indicação de pistas e evidências.

Prof. Felipe Rhenius Nitzke/ Direito e Segurança

Leonardo Rossi/ História

Luis Proença/ Game Design

Me. Marcelo Soares/ Jornalismo, Lagom Data

Me. Massimo Rosner/ Engenharia

Dr. Milagros Miceli/ Sociologia, Un. Humboldt de Berlim

Prof. Rafael Grohmann/ Jornalismo, Un. de Toronto

Thales Alexandre Zirbel Hubner/ C. da Computação

Referências

As fontes diretamente citadas que podem oferecer dados e expor fatos sobre os quais apoia-se o texto estão relacionadas abaixo na ordem em que aparecem, e para aquelas que possuem página na internet, o último acesso foi em novembro de 2024.

1. *'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine, 2024. https://www.972mag.com/lavender-ai-israeli-army-gaza/*

2. *'Order from Amazon': How tech giants are storing mass data for Israel's war. +972 Magazine, 2024. https://www.972mag.com/cloud-israeli-army-gaza-amazon-google-microsoft/*

3. *1 in 4 companies have already replaced workers with ChatGPT. Resume Builder, 2023. https://www.resumebuilder.com/1-in-4-companies-have-already-replaced-workers-with-chatgpt/*

4. *A.I. Bias Caused 80% Of Black Mortgage Applicants To Be Denied. Forbes, 2021. https://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/*

5. *Argonáuticas. Apolônio de Rodes. Editora Perspectiva, 2021.*

6. *Assessing Gender Bias in Machine Translation – A Case Study with Google Translate. Marcelo Prates, Pedro Avelar e Luis Lamb. Cornell University, 2019. https://arxiv.org/abs/1809.02208*

7. *Black Loans Matter: Fighting Bias for AI Fairness in Lending. MIT-IBM Watson AI Lab. https://mitibmwatsonai-lab.mit.edu/research/blog/black-loans-matter-fighting-bias--for-ai-fairness-in-lending/*

8. *Brasil quer ser exemplo em uso de inteligência artificial pelo Judiciário. Valor Econômico, 2024. https://valor.globo.com/brasil/noticia/2024/05/10/brasil-quer-ser-exemplo-em-uso--de-inteligencia-artificial-pelo-judiciario.ghtml*

9. *Challenging systematic prejudices: an investigation into bias against women and girls in large language models. UNESCO, 2024. https://unesdoc.unesco.org/ark:/48223/pf0000388971*

10. *Der Golem, wie er in die Welt kam. Paul Wegener; Karl Boese. 91min. Alemanha, 1921. https://thescriptsavant.com/movies/The_Golem.pdf*

11. *Desafios para adoção de Inteligência Artificial na Força Aérea Brasileira, tese de Daniel Baggio. FGV, 2023. https://gist.github.com/dannyxyz22/fb7bab959af603e37c250dd6028831e9*

12. *Estado-Maior do Exército do Brasil: Portaria Nº 1.318, de 14 de abril de 2024. http://www.sgex.eb.mil.br/sistemas/boletim_do_exercito/copiar.php?codarquivo=181261825&act=sep*

13. *European Union: Platform Workers Directive goes ahead*

– presumption of employment and regulation of algorithmic management in platform work. Global Compliance News. https://www.globalcompliancenews.com/2024/04/03/https-insightplus-bakermckenzie-com-bm-investigations--compliance-ethics-european-union-platform-workers-directive-goes-ahead-presumption-of-employment-and-regulation-of-algorithmic-management-in-pla/

14. *Exclusivo: levantamento revela que 90,5% dos presos por monitoramento facial no Brasil são negros. The Intercept Brasil, 2021. https://www.intercept.com.br/2019/11/21/presos--monitoramento-facial-brasil-negros/*

15. *Introducing Superalignment. Release de imprensa da OpenAI. https://openai.com/index/introducing-superalignment/*

16. *Kasparov e o computador. Folha de S. Paulo, 1997. https://www1.folha.uol.com.br/fsp/1997/5/13/opiniao/3.html*

17. *Lagom Insights: a ilusão futurista da IA. Denise Utochkin e Lagom Data, 2024. https://buttondown.com/lagomdata/archive/lagom-insights-a-ilusao-futurista-da-ia/*

18. *Machine Bias. ProPublica, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing*

19. *Monitoramento dos municípios. O Panóptico. https://www.opanoptico.com.br/*

20. *Necropolítica. Achille Mbembe. N-1 edições, 2018.*

21. *OpenAI No Longer Takes Safety Seriously. Coluna de Peter*

Nagy Salib na Lawfare. https://www.lawfaremedia.org/article/openai-no-longer-takes-safety-seriously

22. *OpenAI's Long-Term AI Risk Team Has Disbanded. Wired. https://www.wired.com/story/openai-superalignment-team--disbanded/*

23. *Paineis de Projetos de I.A. no Poder Judiciário do Conselho Nacional de Justiça (CNJ). https://www.cnj.jus.br/sistemas/plataforma-sinapses/paineis-e-publicacoes/*

24. *Parliament adopts Platform Work Directive. Release do Parlamento Europeu. https://www.europarl.europa.eu/news/en/press-room/20240419IPR20584/parliament-adopts-platform-work-directive*

25. *Retratos da Violência: Cinco meses de monitoramento, análises e descobertas. Rede de Observatórios de Segurança, 2019. https://observatorioseguranca.com.br/wordpress/wp-content/uploads/2019/11/1relatoriorede.pdf*

26. *Soft dá ajuda para enfrentar depressão. Folha de S. Paulo, 1994. https://www1.folha.uol.com.br/fsp/1994/2/02/informatica/16.html*

27. *Tecnologia, Segurança e Direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil. Fundação Konrad Adenauer Stiftung, 2022. https://www.kas.de/documents/265553/0/Tecnologia%2C+Seguran%C3%A7a+e+Direitos+VF.pdf/8c70ec5a=1-adf69-8a39-fb-7faf1b76f91e?version-1.0&t=1696517977110*

28. *Tecnologia, Segurança e Direitos: os usos e riscos de sistemas de reconhecimento facial no Brasil. Fundação Konrad Adenauer Stiftung, 2022. https://www.kas.de/documents/265553/0/Tecnologia%2C+Seguran%C3%A7a+e+Direitos+VF.pdf/8c70ec5a=1-adf69-8a39-fb-7faf1b76f91e?version-1.0&t=1696517977110*

29. *The Defense AI Observatory (DAIO). Helmut Schmidt University em Hamburgo. https://defenseai.eu/english*

30. *The human-machine team: how to create synergy between human & artificial intelligence that will revolutionize our world. Brigadier General Y.S. eBookPro Publishing, 2021.*

31. *The job applicants shut out by AI: 'The interviewer sounded like Siri. The Guardian, 2024. https://www.theguardian.com/technology/2024/mar/06/ai-interviews-job-applications*

32. *The secret bias hidden in mortgage-approval algorithms. The Markup e Associated Press, 2021. https://apnews.com/article/lifestyle-technology-business-race-and-ethnicity-mortgages--2d3d40d5751f933a88c1e17063657586*

33. *Um Rio de câmeras com olhos seletivos. O Panóptico, 2022. https://cesecseguranca.com.br/wp-content/uploads/2022/05/PANOPT_riodecameras_mar22_0404b.pdf*

34. *Unity signs "multi-million dollar" contract to help U.S. government with defense. Game Developer. https://www.gamedeveloper.com/business/unity-signs-multi-million-dollar-contract--to-help-u-s-government-with-defense*

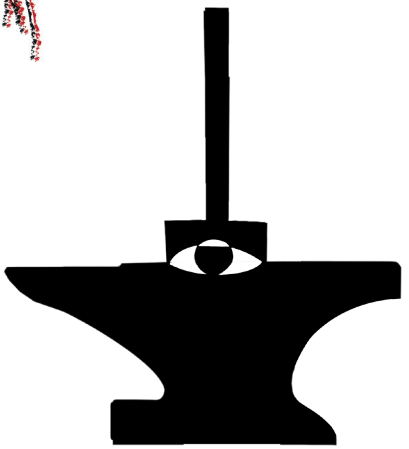
35. *Unity Workers Question Company Ethics As It Expands From Video Games to War. Vice. https://www.vice.com/en/article/unity-workers-question-company-ethics-as-it-expands-from-video-games-to-war/*

36. *Vigilância por lentes opacas: mapeamento da transparência e responsabilização de projetos de reconhecimento facial no Brasil. O Panóptico, 2024. https://lapin.org.br/wp-content/uploads/2024/10/OPANOPTICO_Pesquisa_Vigilancia_Por_Lentes_Opacas.pdf*

37. *What is Artificial Intelligence? IBM Corp. https://www.ibm.com/topics/artificial-intelligence*

38. *When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Stanford Social Innovation Review, 2021. https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity*

1001 Mãos de Ferro



Uma inteligência antiga, opaca e Artificial

Carolina Monteiro

Este livro foi composto em Crimson Text,
Courier New e letras manuais, impresso em
papel offwhite 70g/m² e couché matte 115g/m².

"Ao concordar com os termos deste acordo, você se compromete a não divulgar, compartilhar ou duplicar qualquer informação sobre o que está trabalhando para outras pessoas."

Independent Contractor Agreement

1001 Mãos de Ferro