

ALÉM DO FAUSTO E DO HYPE: SOBRE IMAGINÁRIOS E NARRATIVAS NA ERA DA INTELIGÊNCIA ARTIFICIAL

BEYOND FAUST AND HYPE:
ON IMAGINARIES AND NARRATIVES IN THE AGE OF ARTIFICIAL INTELLIGENCE

Daniele Gambetta¹
Traduzido por Ana Clara Santos Elesbão²

RESUMO

O texto analisa criticamente os imaginários, narrativas e disputas políticas em torno da inteligência artificial contemporânea, propondo uma leitura que ultrapassa tanto o entusiasmo tecnoutópico quanto o alarmismo distópico. A partir de referências da cibernética, da sociologia da tecnologia e da economia política, o texto examina os modelos de linguagem de larga escala como dispositivos técnicos que operam fundamentalmente por recombinação estatística da linguagem, problematizando a ideia de inteligência, autonomia e neutralidade associada a essas tecnologias. O debate sobre os chamados “papagaios estocásticos” é articulado às transformações do trabalho cognitivo, às controvérsias sobre direitos autorais e às condições materiais e exploratórias que sustentam a produção da IA. Discute-se ainda o papel do *hype* e do animismo tecnológico na consolidação de ideologias como o TESCREAL, bem como os impactos geopolíticos da IA, a disputa em torno dos semicondutores e os conflitos regulatórios expressos no AI Act europeu. Ao mobilizar a metáfora da “caixa-preta”, enfatiza-se a necessidade de abordagens interdisciplinares e críticas capazes de reencantar politicamente a tecnologia, abrindo espaço para imaginar formas alternativas de governança, produção e uso da inteligência artificial orientadas por justiça social, ambiental e democrática.

Palavras-chave: Inteligência Artificial; Imaginários Tecnológicos; Trabalho Digital; Regulação da Tecnologia.

ABSTRACT

The article critically analyzes the imaginaries, narratives, and political disputes surrounding contemporary artificial intelligence, proposing a reading that goes beyond both technoutopian enthusiasm and dystopian alarmism. Drawing on references from cybernetics, the sociology of technology, and political economy, the text examines large-scale language models as technical devices that operate fundamentally through the statistical recombination of language, questioning the notions of intelligence, autonomy, and neutrality associated with these technologies. The debate over so-called “stochastic parrots” is linked to transformations in cognitive labor, controversies over copyright, and the material and exploitative conditions that underpin AI production. The article

¹ Daniele Gambetta é licenciado em Matemática e doutorando em Inteligência Artificial na Universidade de Pisa/ITA. É editor da antologia *Datacrazia* (D Editore, 2018). Ao longo dos anos, contribuiu com artigos sobre ciência e tecnologia para diversas revistas e jornais italianos, incluindo Fanpage, Vice, ilManifesto, Pagina99, entre outros. E-mail: daniele.gambetta7@gmail.com. ORCID: <https://orcid.org/0009-0001-0918-3769>

² Doutoranda e Mestra pelo Programa de Pós-Graduação em Ciências Criminais da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) com bolsa CAPES. E-mail: anaelesbaos@gmail.com. ORCID: <https://orcid.org/0000-0003-0394-0295>

also discusses the role of hype and technological animism in the consolidation of ideologies such as TESCREAL, as well as the geopolitical impacts of AI, disputes over semiconductors, and regulatory conflicts expressed in the European AI Act. By mobilizing the metaphor of the “black box,” the text emphasizes the need for interdisciplinary and critical approaches capable of politically re-enchanting technology, opening space to imagine alternative forms of governance, production, and use of artificial intelligence oriented toward social, environmental, and democratic justice.

Keywords: Artificial Intelligence; Technological Imaginaries; Digital Labor; Technology Regulation.

Mas, embora no mundo antigo a máquina ainda pudesse ser um símbolo de unidade com a natureza, ou o produto de uma engenhosidade fora do comum (como os autômatos autopropulsados de Heron de Alexandria), com a chegada da civilização industrial e da modernidade, essa laceração entre o homem e a natureza, que já se expressava no mito, mas conservava uma esperança de recomposição, tornou-se irremediável.

Antonio Caronia, *Robots Between Dream and Work*, 1991.

Desse ponto de vista, o conhecimento pode ser descrito como um organismo vivo [...] E enquanto vive, vive das e nas interpretações dos atores que interagem com ele: sejam as pessoas que usam esse conhecimento para suas vidas ou para aproveitá-lo ao máximo, sejam as entidades computacionais que coletam esse conhecimento por algum motivo, seja para torná-lo acessível de alguma forma por um mecanismo de busca, seja para tentar processá-lo com os processos da inteligência artificial.

Salvatore Iaconesi, *The Spiral of Knowledge*, 2020.

INTRODUÇÃO

Em sua publicação *Futuredays - A Nineteenth century vision of the year 2000*, o escritor de ficção científica Isaac Asimov relatou, em 1986, uma série de 25 ilustrações que datam do final do século anterior e foram publicadas em 1910 sob o título *En l'an 2000*. Nessas imagens, o ilustrador francês Villemard retratou as previsões e fantasias da época sobre a tecnologia e suas aplicações 100 anos à frente. Os diversos cartões-postais mostram meios de transporte futuristas e máquinas de vários tipos, capazes de substituir o trabalho humano e manual. Duas coisas se destacam principalmente: uma, claro, é o evidente significado utópico e positivo da presença de máquinas assumindo o lugar dos seres humanos nas tarefas do dia a dia; a outra é a representação mecânica dos dispositivos de forma a se assemelharem à estrutura orgânica que substituem. Assim, por exemplo, uma ferramenta para cortar cabelo possui braços e mãos mecânicas com os quais segura a tesoura e o pente; um aparelho doméstico automático para limpar o chão consiste em uma base com rodas que maneja uma vassoura e uma pá; enquanto, para transferir conhecimento para a cabeça dos alunos, uma escola dispõe de uma máquina que literalmente ingere livros, os converte e envia as noções às crianças por meio de cabos conectados às suas cabeças.

Por outro lado, em muitos casos, a ciência tem tentado obter tecnologias capazes de simular as faculdades humanas, tentando reproduzir fielmente as estruturas que as possibilitam para nós, mas quanto mais as faculdades envolvidas se relacionam com a esfera cognitiva, maior o risco de que essa

abordagem nos leve a um caminho equivocado, além de criar confusões e mal-entendidos no uso dessas tecnologias. Quando dizemos que uma máquina - seja uma Olivetti Studio 44, um Mac ou um smartphone - está escrevendo, certamente não pensamos nela como tendo mãos com caneta e caderno, assim como máquinas voadoras não precisam de asas. A analogia entre a escrita humana e a escrita por máquinas, portanto, atinge um certo nível de profundidade, cujo limite é conhecido por todos. No entanto, no momento em que afirmamos que uma máquina pensa, ou até mesmo que o faz graças a redes neurais e neurônios artificiais, é então que a confusão aumenta, e não conseguimos mais entender exatamente o que significa dizer que ela pensa, quão semelhante é seu pensamento ao pensamento humano ou até onde pode ir na simulação do nosso pensamento.

O termo Inteligência Artificial abrange, na verdade, uma ampla gama de técnicas e métodos. Produtos como Midjourney, ChatGPT e similares introduziram ao público em geral os chamados “modelos generativos”, provocando, com razão, um enorme debate e demonstrando, de forma ainda em desenvolvimento, o impacto dessas tecnologias em toda a esfera social. Procurarei aqui, entretanto, lançar luz sobre a natureza e a estrutura dessas ferramentas e analisar brevemente o papel que vêm assumindo em um mundo em constante aceleração.

PAPAGAIOS ESTOCÁSTICOS E ONDE ENCONTRÁ-LOS

No célebre artigo *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* (Bender *et al*, 2021), um grupo de pesquisadores estadunidenses alertou sobre os riscos e as criticidades do uso dos chamados Modelos de Linguagem de Larga Escala (LLMs, na sigla em inglês), como o GPT e seus derivados. Como acertadamente apontado pelos autores, essas tecnologias não são capazes de compreender texto, mas sim de reproduzir em cascata uma série de palavras a partir de palavras anteriores, de acordo com uma determinada estrutura e coerência. Eles são, para usar a analogia do artigo, “papagaios estocásticos”, que atribuem uma certa probabilidade aos termos a serem escritos. Fazer esse esclarecimento é útil para reduzir o *hype* associado a essas tecnologias, evitando assim mal-entendidos sobre seu uso para o público em geral. Também nessa interpretação, podem-se apontar diretrizes úteis para a pesquisa a seguir. Chegamos então à questão: o que o ChatGPT pode e o que não pode fazer? Nos últimos meses, muitas pessoas têm testado, estudado e experimentado suas capacidades. As especificações técnicas de treinamento do ChatGPT não são de domínio público, mas devemos pensar em um modelo de linguagem como um mecanismo de composição de frases, em que a estrutura textual é aprendida, mas as referências precisas desaparecem.

O escritor de ficção científica Ted Chiang escreveu em fevereiro deste ano na *The New Yorker* um dos melhores artigos já escritos sobre o tema da IA: *ChatGPT Is a Blurry JPEG of the Web* Chiang (2023) começa contando como, em 2013, uma empresa alemã de fotocópias percebeu que uma de suas máquinas, ao copiar uma planta baixa com valores de metragem quadrada (como 14,13; 21,11; 17,42), sempre inseria o mesmo número (14,13) na cópia final. Descobriu-se que o fato se devia ao método de compressão que, para economizar memória, associava todos os números ao mesmo símbolo e depois os repetia onde estivessem presentes. A fotocópia era, portanto, uma réplica borrada do original. O mesmo ocorre com o ChatGPT: uma cópia borrada de toda a web. Apresentar um LLM como substituto de um mecanismo

de busca (como proposto pela Microsoft com o Bing e pelo Google com o Bard) é errado e perigoso, pois a busca por referências não é a tarefa para a qual ele foi treinado. É claro que também existem serviços que, além da resposta textual, fornecem os links usados para construir a resposta, mas nesse ponto, poderíamos simplesmente utilizar um mecanismo de busca. Em que, então, um modelo de linguagem se destaca? No que diz respeito especificamente à linguagem: traduções, resumos e reformulações. São essas as atividades para as quais ele é treinado, ainda que, em casos particulares, possa evidentemente apresentar inversões de sentido.

Há outro fato que tem me intrigado nos últimos meses: pessoas sem formação em ciência da computação ficaram extremamente surpresas ao ver a capacidade do ChatGPT de produzir códigos como algoritmos para executar uma determinada tarefa ou para produzir websites. Pessoalmente, entre as várias funcionalidades do produto, essa foi a que menos me surpreendeu: para um humano, escrever código pode parecer uma operação complexa porque requer habilidades específicas, mas um código que executa uma determinada operação é quase sempre escrito da mesma maneira, e a Web está repleta desses escritos. Afinal, é sabido que existem há muito tempo fóruns e sites, como a enorme comunidade do Stack Overflow, em que códigos para operações específicas são compartilhados entre os usuários. Em suma, escrever códigos é uma das atividades mais “papagaiadas” que existem. É claro que isso vale para tarefas simples e, de fato, é fácil perceber que, quando o ChatGPT produz códigos mais complexos, eles frequentemente contêm pequenos erros que ainda tornam o trabalho de controle humano essencial.

Outro elemento fundamental da questão é que, se o ChatGPT existe, é porque pôde se basear em uma quantidade exorbitante de informações coletadas e construídas por cada um de nós na Web (além do trabalho de exércitos de revisores mal remunerados), tornando-as acessíveis por meio de uma interface. O ChatGPT existe, portanto, graças a um trabalho coletivo e distribuído em rede. Com um pouco de ironia, mas não tanta assim, eu diria que a inteligência artificial geral já existe e nasceu em 6 de agosto de 1991, quando foi tomada a decisão de criar uma infraestrutura para coletar e armazenar, nos anos seguintes, o maior acervo de conhecimento humano. Em comparação a esse salto, o ChatGPT (e os LLMs em geral) não passam de um pequeno acréscimo.

TRABALHO, DIREITOS AUTORAIS E TUDO O QUE É ANTIGO E QUE NÃO MORRE

Com o advento do ChatGPT & Cia, também surgiram dúvidas e receios sobre o impacto futuro dessas tecnologias na vida pública. De acordo com o estudo do Fórum Econômico Mundial, *The Future of Jobs Report 2023* (World Economic Forum, 2023), três quartos das principais empresas do mundo esperam adotar IA em suas organizações, e 50% acreditam que isso impulsionará o crescimento de empregos, enquanto apenas 25% afirmam que haverá perda de empregos em diferentes áreas, como saúde, manufatura, agricultura e recrutamento. Agora, surge a questão de quais tecnologias específicas entendemos por inteligência artificial. No que diz respeito novamente ao LLMs, para compreender o que está acontecendo e o que poderá acontecer, basta retomar a discussão sobre o que essas tecnologias podem e não podem fazer. Certamente, o ChatGPT é uma ferramenta que reduzirá drasticamente o tempo de uma tarefa de tradução, especialmente uma que não requer atenção estilística especial, e provavelmente já está reduzindo a força de trabalho necessária em uma redação. Ao mesmo tempo, porém, mesmo

essa tarefa não será completamente substituível, pois uma figura humana será sempre necessária para verificar a exatidão do resultado ou para aconselhar em caso de escolhas específicas de tradução.

Ainda assim, o GPT pode ser útil na redação de um veículo de notícias sensacionalistas, onde a tarefa dos funcionários é pegar textos de agências de notícias e reformular um pouco o conteúdo para, assim, publicar dez artigos por hora. Em suma, o ChatGPT certamente será capaz de economizar tempo em trabalhos repetitivos e pouco refinados que envolvam, precisamente, a manipulação da linguagem. O cenário que parece estar se configurando é o chamado “*human in the loop*”, ou seja, o processo em que uma figura humana está presente como supervisora das atividades de uma IA. Nesse cenário, quanto maiores forem os requisitos de conhecimento da profissão, mais central será a função de supervisão e menor será o risco para o posto de trabalho. Podemos dizer, também com uma pitada de provocação, que a substituição de trabalho pela IA é um indicador do grau de *inutilidade*³ de certos empregos ou de tarefas específicas.

Também devemos lembrar, para não cairmos no mito da IA abstrata e etérea, que o processo de produção do ChatGPT envolve não apenas supervisores com qualificações especiais, mas também um exército de trabalhadores no Quênia, Uganda e Índia, que são explorados por 1,50 euros por hora para identificar discursos de ódio, expressões de violência sexual e outros materiais explícitos a fim de “ajustar a máquina”. Isso não é novidade, como já sabemos por documentários como *The Cleaners* (2018) ou por estudos sobre a força de trabalho oculta em plataformas no chamado fenômeno *Mechanical Turk* [como analisado por Antonio Casilli (2019) ou Matteo Pasquinelli (2023)].

Outra grande questão amplamente debatida nos últimos meses diz respeito ao uso de material encontrado online para treinar algoritmos de aprendizado de máquina e, conseqüentemente, à regulação das relações, econômicas ou não, entre aqueles que produziram esse material e os proprietários do algoritmo. Artistas, programadores e criadores de conteúdo de diversas áreas têm se manifestado nos últimos meses contra o uso indevido de suas obras por IAs. Por mais que essas ações representem demandas pela proteção de formas de trabalho ameaçadas, acredito que qualquer batalha pela defesa ou pelo retorno de uma remuneração baseada na autoria, para além de funcionar como ação de retaguarda, já está perdida diante das evidências dos mecanismos de produção do conhecimento.

Com o advento do digital, e, conseqüentemente, da virtualização de um produto de trabalho desvinculado de um meio específico, ficou claro que a atribuição autoral caminhava para a impossibilidade. Mesmo as tentativas reformistas relacionadas à galáxia do código aberto e das *Creative Commons* apresentam muitas limitações: por mais que se possa colocar um adesivo de “Sem Uso Comercial” em um repositório do Github, uma vez que uma ideia de desenvolvimento de software ou algoritmo esteja circulando e, portanto, possa ser reimplementada, nessa ou em outra linguagem, a autoria se dilui, e s[ó] Deus sabe o quanto o mundo corporativo tem explorado, nos últimos anos, o vasto conhecimento do código aberto como um campo de mineração livre.

A IA é apenas mais um passo nessa direção: os dados podem ser reestruturados, reformulados e usados como treinamento para a construção de um produto. As lutas e reivindicações pela proteção do trabalho cognitivo são fundamentais, mas assumem valor estratégico quando consideradas em um

3 N.T.: No original em inglês, “*bullshitness*”, fazendo referência ao conceito de “*bullshit jobs*”, cunhado por David Graeber para designar posições e tarefas inúteis no gerencialismo corporativo. Cfr.: Glaser, 2018.

caminho voltado para uma política de renda universal desvinculada do trabalho (juntamente com a reivindicação de alta tributação das grandes corporações de tecnologia), e o cenário das novas tecnologias oferece a esse caminho uma oportunidade de espaço no debate público que não deve ser desperdiçada. Além disso, experiências como o caso italiano da GKN e novos projetos no ambiente tecnológico⁴, sugerem que pode ser interessante começar a discutir e compreender, entre os trabalhadores da informação e do desenvolvimento tecnológico, se e como é possível, além da retomada de direitos laborais, também redesenhar a produção e, por que não, redefinir os propósitos e as direções do desenvolvimento da inteligência artificial a partir de baixo, em uma perspectiva de bem-estar, serviços comunitários úteis e conversão ecológica.

TÉCNICA E MAGIA: ANIMISMO DE PLATAFORMA E VIRADA LINGUÍSTICA 2.0

No conto de Fredric Brown, *The Answer*, o protagonista Dwar Ev finalmente concretiza sua criação de um supercomputador que, ao conectar todas as calculadoras de todos os noventa e seis bilhões de planetas do universo, poderia encapsular todo o conhecimento do universo em uma única máquina. Após soldar os dois últimos fios, Dwar abaixa a alavanca, ativando a criação e fazendo-lhe uma pergunta inicial: “Deus existe?” Sem hesitar, a máquina responde: “Sim: agora Deus existe”. Tomado pelo terror e arrependido de sua criação, Dwar tenta se jogar sobre o painel de controle para desligar a máquina, mas um raio do céu o incinera, prendendo a alavanca para sempre no lugar.

No final de março de 2023, uma carta assinada por empresários, autores e especialistas do setor industrial, incluindo Elon Musk, Yoshua Bengio, Steve Wozniak e Noah Harari, foi publicada no site do *Future of Life Institute* (2023), pedindo a suspensão, por pelo menos seis meses, de pesquisas avançadas na área de inteligência artificial, a fim de permitir que o ambiente legislativo se ajuste à regulação dessas tecnologias, visto que seu desenvolvimento deveria garantir transparência e responsabilização. A carta esconde, sem muita discrição, o seu verdadeiro propósito: gerar *hype*. “Olhem para nós, somos os Victor Frankensteins do século XXI, os Drs. Fausto que venderam nossas almas ao Novo Mefistófeles”, é a paráfrase da carta.

Não por coincidência, a linguagem da carta sugere um conjunto de ideologias que vem se consolidando entre os escalões superiores do Vale do Silício, expressas por uma sigla cada vez mais popular: TESCREAL (Transhumanismo, Extropianismo, Singularitarismo, Cosmismo, Racionalismo, Altruísmo Eficaz e Longo-Praxismo). Um termo cunhado criticamente por Timnit Gebru (ex-funcionária do Google e autora do artigo sobre os “papagaios” mencionado anteriormente). Em apenas três anos, a liderança do setor da alta tecnologia apropriou-se do termo, como o empreendedor Marc Andreessen, que se autodenomina um “TESCREALista” (além de “aceleracionista de IA”, “supremacista de GPU” e “ativista cyberpunk”) em sua biografia no Twitter.

Os termos do acrônimo referem-se a uma série de filosofias e correntes que surgiram em torno do transumanismo e suas declinações: profundo entusiasmo pela tecnologia como meio de superar os limites dos seres humanos (doença, envelhecimento e morte), corrida espacial, realização de IA forte como missão da humanidade, emprego do desenvolvimento e da racionalidade para transcender a

⁴ Disponível em: <https://www.crypto-commons.org/reclaim>.

condição material e corporal, e, por fim, lutar pela condição divina através da tecnologia. Alguém se referiu acertadamente ao conceito de animismo para descrever a concepção que a ideia de inteligência artificial está assumindo nesses círculos. Em um tweet de sua autoria em fevereiro passado, o (talvez ex-) CEO da OpenAI (empresa do ChatGPT) sugeriu como uso de seu produto a substituição dos cuidados de saúde para aqueles que não podem arcar com seus custos. Além do classismo grosseiro de tal declaração, é claro que a comunicação por parte desses gigantes é direcionada a um entusiasmo completamente desvinculado da realidade. E pouco adianta recriminar o usuário médio por não compreender o uso correto das tecnologias quando tais mensagens vêm de cima.

Em fevereiro de 2023, foi lançado o Bard, o chatbot de IA com o qual o Google prometia rivalizar com o ChatGPT da OpenAI. Mas poucas horas após o evento de lançamento, em um vídeo promocional, o assistente de Mountain View deu uma resposta incorreta à pergunta feita, afirmando que o Telescópio Espacial James Webb seria usado para tirar as primeiras fotos da história de um planeta localizado fora do sistema solar (exoplaneta), quando, na verdade, os primeiros exoplanetas já haviam sido fotografados por telescópios terrestres 14 anos antes do lançamento do James Webb. Um descuido que não passou despercebido ou perdoado, especialmente pelo mundo financeiro. Na quarta-feira, 8 de fevereiro, as ações da Alphabet fecharam em queda de quase 8%, US\$ 100 bilhões abaixo de seu valor de mercado. E o erro, curiosamente, estava relacionado a um conhecimento específico, e não a uma falha na composição da linguagem, que é justamente a tarefa para a qual a IA foi treinada. Em suma, as gigantes do Vale do Silício se viram vítimas de seu próprio *hype*.

Já estávamos acostumados com o poder da linguagem no âmbito financeiro, como pode ser o caso de uma frase de efeito de um CEO, mas pensar que agora esse poder linguístico também se estende às máquinas causa, no mínimo, certo receio. Vem à mente a cena de Douglas Adams em seu *Guia Intergaláctico*, quando, após sete milhões e meio de anos, uma multidão de expectadores aguarda ansiosamente de *Deep Thought* a resposta para a pergunta fundamental sobre a vida, o universo e tudo mais. Christian Marazzi escreveu no seu livro *The Place for Socks*:

Quando dizemos que, com o pós-fordismo, a comunicação entra na produção, torna-se um fator diretamente produtivo, estamos questionando a linguagem, que, pela sua própria vocação, está na base da comunicação. A coincidência do ato de produzir e do ato de comunicar no novo paradigma da produção abre uma série de problemas de análise da linguagem que são tão fascinantes quanto extremamente complexos e densos. (Marazzi, 1994).

Com o capitalismo avançado da inteligência artificial, estamos testemunhando uma virada linguística 2.0, em que a linguagem assume um papel produtivo ainda mais central. Trata-se de uma espécie de capitalismo mágico, em que a linguagem e suas fórmulas ditadas como oráculos passam a agir cada vez mais sobre o real. Uma magia que se apresenta como abstrata e imaterial, mas que sabemos estar incorporada em processos sociais e materiais, nos caprichos do Vale do Silício e das altas finanças. O desafio é desencantar a tecnologia, libertando-a dos imaginários tescrealistas e construindo outros para além do antropocentrismo e das ambições extrativistas.

ABRINDO A CAIXA PRETA

No início de 1949, nos laboratórios do Hospital Psiquiátrico Barnwood House, um asilo particular nos arredores de Gloucester, foi possível encontrar um misterioso dispositivo quadrado preto composto por quatro acumuladores, cada um equipado com um ímã capaz de oscilar entre diferentes configurações. Conforme relatado em um artigo da *Time* na época, segundo seu criador, o psiquiatra William Ross Ashby, esse dispositivo, o chamado homeostato, era a realização mais próxima de um cérebro humano artificial já projetada pelo homem até então. Ashby, além de médico, foi um dos pioneiros e divulgadores da cibernética, que, mais do que uma disciplina, poderia ser vista como uma série de estudos experimentais interdisciplinares abrangendo engenharia, biologia e ciências sociais. Suas origens podem ser rastreadas arbitrariamente no passado, dependendo de quais raízes do pensamento sistêmico se deseja considerar. No entanto, é certo que, no período imediatamente posterior à guerra, o nome dessa nova ciência foi introduzido e popularizado pelo trabalho de Norbert Wiener, que publicou seu livro *Cibernética: ou Controle e Comunicação no Animal e na Máquina*, em 1948.

As intenções do matemático americano, como expressas pelo título da obra, eram fundar uma nova ciência capaz de lidar com questões de regulação de sistemas naturais e artificiais, encontrando semelhanças e afinidades entre eles, hibridizando métodos das ciências sociais e biológicas com teorias de computação e controle automático. Os mesmos princípios inspiraram Ashby, que registrou seus pensamentos por mais de 44 anos em uma série de diários, resultando em 25 volumes, totalizando 7.189 páginas, agora confiados à Biblioteca Britânica. E foi precisamente com o objetivo de fornecer um exemplo de máquina autorreguladora que Ashby construiu o homeostato, cujo estado podia ser alterado por comandos específicos que representavam o ambiente externo e que, sob condições específicas, era capaz de retornar sozinho a um estado de equilíbrio. Por mais difícil que seja imaginar hoje, tal projeto estava absolutamente alinhado com os objetivos dos ramos nascentes da automação, a ponto de Turing, sabendo da intenção de Ashby, ter-lhe escrito propondo simular esse mecanismo na calculadora que estava projetando perto de Londres.

Mas o que nos interessa na complexa máquina de Ashby, a fim de encontrarmos a chave para os desenvolvimentos em automação que seguirão até os nossos dias, é a abordagem com a qual o homeostato foi concebido, buscando simular um sistema vivo. O modelo proposto por Ashby é o que viria a ser chamado mais tarde de teoria da caixa-preta, ou método da caixa-preta, sobre o qual os próprios ciberneticistas iniciaram um debate acalorado que ainda hoje podemos manter aberto hoje. O próprio Ashby escreveu na década de 1950 que “o que se está argumentando não é que as caixas-pretas se comportam de alguma forma como objetos reais, mas que objetos reais são, na verdade, todos caixas-pretas, e que lidamos com caixas-pretas ao longo de nossas vidas”, sugerindo essencialmente uma ontologia da caixa-preta. O papel da caixa-preta na ciência é um dilema que chega de forma explosiva, no século XXI. É Latour, em *A Esperança de Pandora* (2001), que fala da caixa-preta como

a maneira pela qual o trabalho científico e técnico se torna invisível por seu próprio sucesso. Quando uma máquina funciona eficientemente, concentra-se apenas na entrada e na saída, e não na complexidade interna. Isso faz com que, paradoxalmente, quanto mais bem-sucedida a ciência e a tecnologia se tornam, mais opacas e obscuras elas se tornam. (Latour, 2001)

Dando mais um salto de anos e chegando a tempos muito recentes, não é coincidência que preocupações com a ciência opaca tenham surgido no campo da inteligência artificial, onde se tornou evidente a necessidade de compreender os motivos de uma escolha automatizada em detrimento de outra, ou os parâmetros de classificação utilizados. Isso levou a pesquisa a desenvolver, por exemplo, o ramo da chamada Inteligência Artificial Explicável (*Explainable Artificial Intelligence*, XAI) (Guidotti *et al*, 2018). O problema da caixa preta, no entanto, não apenas técnico, mas algo cada vez mais relacionado às esferas social, econômica e cultural, envolvendo, portanto, o estudo dos efeitos da IA em um sistema sociotécnico complexo. Ser caixa preta não significa apenas IA em sentido estrito, mas também inteligência computacional distribuída, composta por uma rede aproveitada de calculadoras, dispositivos IoT, corpos hiperconectados em movimento e em constante produção. Organismo sombrio é também a máquina financeira, cheia de sutilezas metafísicas e caprichos teológicos, tão complexa que mal é compreendida pelos xamãs palacianos, ocupados demais lutando contra os demônios algorítmicos de alta frequência que eles próprios conjuraram há eras e sobre os quais agora perderam o controle total. O que se mostra cada vez mais opaca é a ciência, não por falta de informação ou inteligência, mas devido à fragmentação microdisciplinar crescente, rumo a uma hiperespecialização incapaz de se comunicar com o restante do mundo. Podemos, então, falar de sociedades de caixa-preta, como sugere Frank Pasquale, e tentar tecer considerações parciais sobre esses hiperobjetos que governam nossas vidas no limite da compreensão, no que Bridle, citando o mestre do horror Lovecraft, considera uma nova idade das trevas.

Voltando a Latour, e ao texto que ele escreveu com Steve Woolgar, *Laboratory Life* (1979),

a atividade de construir caixas-pretas, de tornar objetos de conhecimento distintos das circunstâncias de sua criação, é exatamente o que ocupa os cientistas a maior parte do tempo. [...] Uma vez que um objeto de estudo é processado em laboratório, torna-se muito difícil transformá-lo novamente em um objeto sociológico. O custo em detectar fatores sociais reflete a importância da atividade da caixa-preta. (Latour & Woolgar, 1979).

Abrir a caixa-preta, portanto, não é apenas uma tarefa para especialistas ou técnicos, assim como, voltando a Bridle, embora possa ser útil, não é necessário aprender programação para abordar o debate sobre novas tecnologias. Ao contrário, o que emerge de novas vertentes de pesquisa, como a análise de redes sociais ou a sociologia computacional, é que o ambiente da ciência da computação e das ciências exatas se viu catapultado, em poucos anos, a confrontar questões metodológicas com as quais não estava habituado a dialogar, ao contrário de sociólogos, psicólogos ou humanistas. Portanto, não apenas a interdisciplinaridade se torna necessária, mas também uma crítica dos processos de produção capaz de reconhecer a dinâmica do valor, trabalho e impacto ambiental dos sistemas abertos. Nesse sentido, dessa inteligência opaca distribuída, ou ao menos de uma pequena parte dessa megamáquina, podemos também tentar abrir algumas caixas-pretas com as ferramentas da “pesquisa coletiva de baixo para cima”, por meio de levantamentos nos lugares da produção digital e da sociologia a partir da vida cotidiana. *Into the Black Box*⁵ é “um projeto de pesquisa coletivo e transdisciplinar que adota a logística como perspectiva privilegiada para investigar as atuais mutações políticas, econômicas e sociais”. A logística como forma de inteligência estratégica na narrativa da produção harmonizada e

5 Disponível em: <http://www.intotheblackbox.com/>

autorregulada. Os artigos no site do coletivo e os eventos organizados, principalmente em Bolonha, oferecem chaves interpretativas muito úteis para ler a sociedade da caixa-preta, ampliando o olhar “para dentro e para além da tela”, e oferecendo uma visão interessante sobre práxis política e estratégica, bem como sobre análise. Projetos como este, dentro e fora da academia, são cada vez mais necessários.

A GUERRA DO VIÉS

Era o século XVII, e Gottfried Leibniz desenvolvia seu sistema de *mathesis universalis*, um método simbólico capaz de resolver qualquer controvérsia por meio de processos computacionais. Era 2008, e Chris Anderson proclamou na *Wired* o fim da teoria em virtude da era dos dados e da correlação. Desde então, o tempo passou e, em certa medida, até mesmo na academia a euforia pelos big data arrefeceu, em parte como resultado de casos noticiados sobre problemas decorrentes do uso de tecnologias de IA sem um aparato epistemológico adequado: o *Tay bot* do Twitter, o caso COMPAS e outros casos agora bem conhecidos na literatura acadêmica. Assim, começou-se a falar em viés (*bias*), com pesquisa sobre IA se ramificando para os campos da explicabilidade, buscando abrir a caixa-preta das decisões e compreender os mecanismos de tomada de decisão. No que diz respeito aos LLMs, em particular, faz-se referência a estereótipos transmitidos pela linguagem, ou mesmo há quem fale de “viés” e “vulnerabilidade” diante de respostas de chatbots que não estejam de acordo com as políticas. O sonho da inteligência artificial correta é o de uma tecnologia que pense e se comunique como humanos, mas sem mentir, sem fazer julgamentos e sem dizer bobagens; basicamente, que não pense e se comunique como humanos. No entanto, mesmo na literatura acadêmica, reconhece-se agora que o próprio termo *viés* é usado com diferentes significados e intenções, tornando o debate sobre o tema cada vez mais confuso. Em sociologia, o termo *viés* tem um significado definido e indica um possível erro na avaliação ou mensuração de um fenômeno devido a técnicas de pesquisa, uma distância entre o fenômeno estudado e sua mensuração. Poderíamos então estabelecer que, para a IA, viés é a diferença entre um benchmark considerado caso padrão ou nulo e um modelo artificial. E aqui chegamos ao ponto: quando falamos em viés racista ou sexista de uma IA, ou mesmo em viés devido à linguagem vulgar, em relação ao que estamos afirmando, existe uma distância? O que é o modelo nulo?

Se considerarmos que os textos mais comumente usados para treinamento vêm de grandes plataformas da web, descobrimos que já estamos fazendo uma escolha muito específica de tipo de linguagem e paradigma interpretativo. De acordo com a *Pwe Internet Research* de 2016 (Greenwood *et al*, 2016), os usuários do Reddit são 67% homens dos Estados Unidos, com 64% entre 18 e 29 anos. Outros estudos sobre a composição da comunidade da Wikipédia descobrem que entre 8 e 15% são mulheres. O que uma IA treinada nessas plataformas pensará ou sentirá sobre a expansão colonial? O que ela responderá quando alguém tentar pedir aconselhamento ou informação sobre aborto ou transição de gênero, dependendo do país ao qual se refere? A resposta sempre dependerá de quais informações, textos e materiais foram usados para treinar a máquina. A questão não é melhorar ou corrigir o modelo. A questão é que toda representação do mundo e qualquer forma de expressão é situada e está posicionada em um contexto social e histórico específico e, portanto, em uma situação particular de privilégio. Eliminar o viés é uma demanda sem sentido, uma vez que o discurso neutro não existe. A busca por uma justiça

fantasmagórica da IA, alinhada aos cânones e costumes sociais, só pode reconhecer que cânones e normas estejam situados geográfica, social e historicamente.

Em 2008, uma estação de rádio Māori sem fins lucrativos da Nova Zelândia, chamada *The Hiku Media*, iniciou um projeto com o objetivo de treinar algoritmos em 300 horas de conversação na língua Māori, mantendo o controle sobre os dados de sua comunidade. Depois que Aotearoa, o nome Māori para a Nova Zelândia, se tornou uma colônia britânica em 1840, o inglês passou a predominar como língua local, sendo imposto como única língua nas escolas 30 anos depois. Já em 1970, habitantes e ativistas Māori estabeleceram escolas para preservar sua língua, mas devido ao despovoamento e à urbanização, a porcentagem de falantes caiu de 90% para 12%. Somente em 1987 o Te Reo foi reconhecido como língua oficial. Com o advento das tecnologias digitais, a comunidade se encontra novamente em uma encruzilhada. “Se essas novas tecnologias ‘falam’ apenas línguas ocidentais, é inevitável que sejamos excluídos da economia digital”, afirma Michael Running Wolf, desenvolvedor Cheyenne envolvido em projetos de IA para a conservação linguística. “Os dados são a última fronteira da colonização.”

A *The Hiku Media* não deseja publicar dados e informações no Facebook e no YouTube, nem deixá-los sob licenças que não permitam o controle da comunidade. “Os dados são a última fronteira da colonização [...] Nossos dados seriam usados pelas mesmas pessoas que nos privaram dessa língua para nos vendê-los de volta como um serviço [...] É como quando tomam sua terra e depois a vendem de volta para você”, afirmam os membros do projeto. Novamente, a cultura de código aberto esbarra em um problema de tomada de decisão sobre o uso de dados, então a *The Hiku* criou uma licença específica que explicita as regras básicas para futuras colaborações, com base no princípio Māori de kaitiakitanga, ou proteção, garantindo acesso aos dados apenas a organizações que concordem em repassar todos os benefícios derivados de seu uso ao povo Māori. Obviamente, grandes empresas, observando esse e outros experimentos, já estão construindo seus próprios arquivos de dados locais em idiomas nativos, e construindo modelos de negócios sob essa perspectiva. O que fará a diferença nesse processo será o equilíbrio de poder entre as partes envolvidas.

GEOPOLÍTICA, REGULAMENTAÇÕES E MICROCHIPS

A ideia etérea e desencarnada da IA colide não apenas com a materialidade do trabalho, mas também com o impacto que a IA vem tendo nas dinâmicas internacionais. Para a China, a corrida da IA assumiu um papel fundamental, tanto para o desenvolvimento da indústria doméstica quanto no nível de confronto geopolítico, especialmente com os Estados Unidos, a ponto de Xi Jinping ter mencionado o tema diversas vezes em seus discursos. Durante o período da Covid, o uso de tecnologias digitais e inteligentes como ferramentas de controle populacional ficou evidente, desde a medição de temperatura até a triagem facial automática dentro do projeto de crédito social. Os objetivos da IA incluem o conceito de cidades inteligentes e infraestrutura interconectada de alta velocidade (5G e IoT); uma integração completa da esfera digital no planejamento econômico e de produção. A importância da IA na China é demonstrada pelo fato de o mecanismo de busca Baidu ter criado uma alternativa em língua chinesa ao ChatGPT em março de 2023, chamada Ernie.

Obviamente, no contexto de confronto internacional, a necessidade de tecnologias inteligentes no campo da guerra torna-se cada vez mais premente. Na guerra na Ucrânia, a superioridade tecnológica

da OTAN permitiu às forças de Kiev um desempenho no campo de batalha que não é insignificante aos olhos de outras nações. O exército chinês identificou diversos campos de interesse da IA, desde veículos autônomos sem motorista até drones operados remotamente, com todas as implicações éticas que essas tecnologias acarretam. O Pentágono estima que a República Popular da China esteja em um nível de desenvolvimento de IA equivalente ao dos EUA, a ponto de o relatório final da Comissão de Segurança Nacional sobre inteligência artificial afirmar que “os planos, recursos e progresso da China devem preocupar todos os americanos. Em muitas áreas, Pequim igualou os níveis dos Estados Unidos e está até à frente em algumas delas”.

A guerra da IA não se limita ao desenvolvimento de software, estendendo-se também à parte física das tecnologias e às matérias-primas para sua produção. Embora seja realmente difícil controlar produtos de trabalho intangíveis, como algoritmos, é significativamente mais fácil controlar matérias-primas e microprocessadores, o principal motivo do atrito entre os EUA e a China em relação a Taiwan. Taiwan abriga o maior sistema industrial de produção de chips do mundo. Sessenta por cento do mercado global e 90% dos microchips mais avançados são produzidos nessa ilha. A TSMC (*Taiwan Semiconductor Manufacturing Company*) é conhecida como a principal indústria de fabricação e montagem de semicondutores do mundo. Em 2020, Donald Trump convenceu a TSMC a investir no Arizona após bloquear suas exportações para a Huawei. A TSMC possui instalações na China continental há vários anos e o vínculo com Pequim sempre foi mantido, embora sempre tenha evitado exportar microchips de uso militar para a China. A guerra dos semicondutores entre o protecionismo dos EUA e a busca da China pela independência representa um dos desafios mais importantes do século XXI, com possíveis consequências e efeitos colaterais em toda a política internacional.

A DISPUTA PELO AI ACT ENTRE SOBERANIA E DESREGULAMENTAÇÃO

Com 523 votos a favor, 46 contra e 49 abstenções, após um longo e conturbado processo que por vezes pareceu estar num beco sem saída, em 13 de março de 2024, o Parlamento Europeu aprovou o AI Act, a primeira regulamentação do mundo sobre os usos da inteligência artificial e sistemas baseados nessa tecnologia. Composto por 113 artigos e 12 anexos, o documento visa proteger os direitos fundamentais e a sustentabilidade ambiental, mantendo o objetivo de facilitar e promover o crescimento de concorrentes europeus em relação aos gigantes americanos.

O Artigo 3 é o que define o que se entende por inteligência artificial, que é “um sistema baseado em máquina projetado para operar com diferentes níveis de autonomia” e que, a partir das previsões e do conteúdo gerado, “pode influenciar ambientes físicos ou virtuais”. Assim, a definição não se aplica a sistemas de software tradicionais ou abordagens que dependem de regras predefinidas. Os critérios são baseados em uma classificação de risco de sistemas de inteligência artificial. Algumas aplicações consideradas incompatíveis com os direitos pessoais são proibidas. Nesta categoria encontramos: sistemas de identificação e categorização biométrica baseados em características sensíveis; sistemas de crédito social como o da China; reconhecimento automático de emoções em locais de trabalho e escolas; policiamento preditivo baseado em perfis; e raspagem de imagens faciais de pessoas da internet, ou seja, a prática de baixar automaticamente grandes quantidades de imagens para criar bancos de dados.

No que diz respeito à proibição da identificação biométrica, no entanto, estão incluídas exceções nas quais a prática seria permitida, como no caso de busca de uma pessoa desaparecida, prevenção de um ataque terrorista ou investigação de um crime grave. Outros sistemas são permitidos, mas são considerados de alto risco devido ao seu impacto negativo na saúde ou segurança. Esses casos, como IA para gerenciar o tráfego rodoviário ou analisar currículos para vagas de emprego, estão sujeitos a obrigações específicas, como avaliar o risco na fase de projeto, ter documentação técnica, ser transparente e garantir a supervisão humana dos processos. Além disso, os usuários desses aplicativos podem registrar reclamações junto às autoridades competentes ou solicitar explicações sobre as escolhas feitas por sistemas automatizados. Por fim, sistemas de uso geral, como modelos generativos de texto ou imagem, terão que cumprir as normas europeias de transparência e direitos autorais. Os materiais gerados terão que ser claramente reconhecíveis como tal, e o usuário deve estar sempre ciente de que está conversando com uma máquina.

Para equilibrar as restrições, o regulamento também introduz facilidades para as empresas testem e adaptem o desenvolvimento de tecnologias, como os chamados sandboxes regulatórios, que são ambientes físicos ou virtuais nos quais as empresas podem testar novas tecnologias monitoradas por uma autoridade competente para garantir a compatibilidade do produto uma vez em circulação. Além disso, são oferecidas facilidades para pequenas e médias empresas e startups, e a introdução de autoridades nacionais de supervisão em cada país da UE. As penalidades por violações das regras podem chegar a 56 milhões de euros ou 7% do faturamento.

Um ponto a ser observado, como já destacado com o GDPR (Regulamento Geral sobre a Proteção de Dados), é que alguns requisitos, como transparência e a possibilidade de explicação de escolhas automatizadas, são objetivos de pesquisa que ainda estão em fase de desenvolvimento, portanto, considerá-los uma obrigação não leva em conta totalmente as possibilidades atuais dessas tecnologias. O mesmo argumento se aplica à “declaração de falsificações”, quando ainda não está claro, mesmo para a pesquisa, como é possível inserir em uma imagem, ou especialmente em um texto, um marcador que indique sua origem artificial (as chamadas marcas d’água) sem comprometer o conteúdo.

Além disso, alguns ativistas e organizações de direitos digitais, como a Access Now, criticaram o texto final por estar repleto de “brechas e exceções” em favor do controle social e corporativo.

Em particular, segundo a associação, ele não proíbe o policiamento preditivo e a vigilância em massa, nem cria políticas diferenciadas para migrantes e refugiados, que não são protegidos pela regulamentação, um ponto que a Anistia Internacional já destacou no ano passado. A UE vem se dedicando ao desenvolvimento de ferramentas automatizadas de vigilância e detecção em áreas de fronteira há anos, e a Lei de IA não afetaria essa prática. É por isso que diversas organizações (Anistia Internacional, Algorithm Watch, Avaaz e outras) lançaram a campanha #ProtectNotSurveil, apelando à Comissão e ao Parlamento para que parem com essas práticas.

Mas, além das especificidades legais da regulamentação, que está em fase final de elaboração, o que pode ser interessante é o confronto entre as partes envolvidas que se viu nos meses que antecederam sua aprovação, quando esse caminho sofreu um revés diante da oposição conjunta da França, Alemanha e Itália, que discordaram da regulamentação específica dos modelos fundamentais, ou seja, aqueles modelos básicos que permitem o processamento de dados em IA. Os três países argumentaram

que limitar esses modelos poderia prejudicar a inovação e a competitividade em relação aos concorrentes internacionais, principalmente os EUA, onde os modelos que dominam o setor foram desenvolvidos até o momento. A intenção da UE é limitar ainda mais as grandes empresas de tecnologia dos EUA, mas alguns países viram isso como uma limitação à sua eventual entrada na competição com o Vale do Silício.

Como relata a Euractiv, pressionar a oposição da França à aprovação seria a Mistral, a startup francesa que arrecadou 105 milhões de euros em uma rodada de financiamento recorde em junho de 2023. Por trás deste projeto estariam empreendedores com experiência em empresas como Google, Meta e Iliad, que também são apoiados pelo próprio Macron no objetivo de criar um concorrente francês para os modelos de idioma dos EUA. “A ambição da Mistral”, disse o chefe de assuntos europeus da *France Digitale* ao blog *Network Wars*, “é ser a OpenAI europeia”. Por trás da posição semelhante assumida pela Alemanha estaria a empresa Aleph Alpha, enquanto no final de outubro de 2023 a segunda reunião trilateral para cooperação industrial foi realizada em Roma, na qual o Ministro da Empresa e do Made in Italy, Adolfo Urso, também firmou acordos com a França e a Alemanha no desenvolvimento de tecnologias de IA.

“Temos visto um amplo lobby das Big Techs em relação à Lei de IA, com inúmeras reuniões com parlamentares europeus e acesso aos mais altos níveis decisórios”, disse a organização sem fins lucrativos Corporate Europe Observatory ao TechCrunch. “Embora essas empresas tenham reivindicado publicamente a regulamentação de IA perigosa, elas estão, na verdade, pressionando por uma abordagem laissez-faire, em que as Big Techs decidem as regras. Isso é particularmente problemático porque a Lei de IA supostamente protege nossos direitos humanos de sistemas de IA arriscados e distorcidos.” Além disso, em relação à regulamentação do reconhecimento biométrico, há também o investimento que a França já fez em sistemas de segurança para as próximas Olimpíadas de Paris, uma posição que pode ter sido decisiva para suavizar a regra em casos específicos.

O modelo francês da Mistral é caracterizado por alto desempenho comparável aos concorrentes americanos, exigindo muito menos desempenho da máquina, permitindo a composição e o processamento de texto em vários idiomas, incluindo o de programação. O governo francês apoiou e financiou fortemente a startup, observando a regulamentação europeia como uma limitação ao seu crescimento no mercado da indústria. Em outubro de 2023, Cédric O, secretário de Estado para a economia digital, observou que “dependendo da forma final do AI Act, isso poderia matar a Mistral ou fazê-la crescer”. E novamente: “O principal problema para a Europa não é a regulamentação, mas sim que, no mundo digital, são os líderes que definem os padrões, e a Europa não tem líderes”, disse ele, referindo-se às maiores empresas de tecnologia do mundo, EUA e China. “A Europa deve se concentrar em uma coisa: como trazer à tona seus campeões”. Suas palavras ressaltam a questão da chamada “soberania da IA”, a soberania da inteligência artificial, que, como foi para o gerenciamento de dados há alguns anos, é agora o novo campo de batalha e cabo de guerra para a afirmação de relações de poder mútuo no espaço geopolítico. Em um tweet de 17 de novembro de 2023, o CEO da Mistral, Arthur Mensch, acusou a regulamentação europeia de limitar o surgimento de novos players: “A intenção de introduzir uma regulamentação em dois níveis é virtuosa. Seu efeito é catastrófico. Ela consolida efetivamente a existência de duas categorias de empresas: aquelas com o direito de escalar, ou seja, as dominantes que podem arcar com os altos requisitos de conformidade, e aquelas que não podem porque não têm um exército de

advogados, ou seja, as recém-chegadas. Isso sinaliza a todos que apenas os principais players existentes podem fornecer soluções de ponta.”

O terreno do debate atual e os interesses em jogo lembram o cenário de 2018, durante a entrada em vigor do GDPR, após o caso da Cambridge Analytica. No entanto, deve-se considerar que, naquela época, a questão era sobre a propriedade dos dados dos usuários, portanto, um fenômeno bastante “distribuído”, embora em escalas diferentes, com serviços e plataformas tanto nos EUA quanto na Europa. Nesse caso, a regulação vai para um setor altamente verticalizado, onde algumas poucas - atualmente três - gigantes detêm os únicos modelos com poder suficiente para estarem sujeitas às restrições em questão.

Isso também faz com que, se por um lado a UE pressiona por uma maior imposição em termos de concorrência sobre um monopólio que, por enquanto, é apenas americano, por outro, isso é visto como uma limitação ao crescimento por parte dos Estados que apostam em startups com a intenção de escalar esse patamar. Ecoando nas palavras de O e Mensch está a ideologia californiana que exalta a desregulamentação em favor do livre mercado. Além disso, como Morozov (2023) descreve em seu artigo no *New York Times*, “a verdadeira ameaça da IA” reside na ideologia do “longotermismo” e do “A.G.I.-ismo” (onde A.G.I. significa inteligência artificial geral): bifurcações do transumanismo capitalista, narrativas tecnoentusiastas que veem a criação de tecnologias inteligentes como uma missão a ser cumprida para o bem da humanidade, e para as quais repercussões como violações de direitos humanos e impacto ambiental são meros acidentes ao longo do caminho.

REALISMO CAPITALISTA E REALISMO ALGORÍTMICO

Não pode haver uma conclusão definitiva para este texto. A globalização e o turbocapitalismo digital nos últimos anos sempre nos colocaram diante de hiperobjetos cada vez maiores, difíceis de compreender e combater, a ponto de nos deixarem em um estado de perplexidade e desamparo. As respostas alternativas de plataformas ao capital devem, no entanto, lidar com os custos humanos e ambientais envolvidos nessas tecnologias, e a OpenAI não divulgou dados específicos sobre emissões de infraestrutura. Um estudo recente realizado por pesquisadores da Universidade da Califórnia estimou a pegada hídrica de modelos de IA como o GPT, considerando o uso de 700.000 litros de água doce durante o treinamento de modelos em data centers da Microsoft, o equivalente à água necessária para produzir 370 carros BMW ou 320 veículos Tesla (Li *et al*, 2023). Também do artigo sobre papagaios estocásticos: “Se o ser humano médio é responsável por cerca de 5 toneladas de CO2 por ano, treinar um modelo Transformer emite cerca de 284 toneladas. Estima-se que treinar um único modelo BERT em GPUs exija tanta energia quanto um voo transatlântico”. Certamente, desde a publicação do artigo, tecnologias e algoritmos de alto desempenho estão reduzindo custos e a quantidade de matérias-primas, mas também sabemos que, na cadeia produtiva do capitalismo, menos custos não significa menos exploração, muito pelo contrário.

O que fazer então? Algumas formas de ação podem ser sugeridas pela união de trabalhadores da tecnologia e daqueles marginalizados e/ou oprimidos pelas tecnologias, com o objetivo de repensar as formas da tecnologia e seus objetivos. Sindicatos informais como a *Tech Workers Coalition* ou a onda de recusa trabalhista do Big Quit e, claro, os movimentos transfeminista e ecológico são sintomas de ruptura em direção a um sistema que agora está em colapso, sinais que podem e devem encontrar

convergências oportunas. Outra consideração pode advir do fato de que a estrutura discursiva e a hegemonia do realismo capitalista também assentam suas bases em uma suposta superioridade e inacessibilidade da tecnologia: redistribuir o conhecimento crítico torna-se então não apenas um meio de tornar o mundo em que vivemos compreensível, mas também uma questão política. Os círculos acadêmicos e educacionais estão cada vez mais imersos na retórica empreendedora e neoliberal, especialmente nas áreas técnicas, onde também assistimos a uma intensificação das colaborações entre ciência e capital nas forças armadas. Repensar uma cultura e educação técnico-científica dissociadas da lógica extrativista revela-se, então, mais uma ferramenta fundamental no caminho para a justiça climática e social.

REFERÊNCIAS

- BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITTECHT, Sherrie. On the dangers of stochastic parrots: Can language models be too big? In: **Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FACCT '21)**, 2021, Virtual Event, Canada. p. 610-623. New York: ACM, 2021. DOI: 10.1145/3442188.3445922.
- BLOCK, Hans; RIESEWIECK, Moritz. **The Cleaners**. [Filme documental]. Direção de BLOCK, H.; RIESEWIECK, M. Gebrueder Beetz Filmproduktion, 2018.
- CASILLI, Antonio A. **En attendant les robots**: Enquête sur le travail du clic. Paris: Seuil, 2019.
- CHIANG, Ted. ChatGPT Is a Blurry JPEG of the Web. **The New Yorker**, 9 fev. 2023. Disponível em: <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>. Acesso em: 16 dez. 2025.
- CRYPTO COMMONS ASSOCIATION. **Reclaim - Generative Algorithms for the Commons**. Crypto-Commons.org. Disponível em: <https://www.crypto-commons.org/reclaim>. Acesso em: 16 dez. 2025.
- FUTURE OF LIFE INSTITUTE. **Pause Giant AI Experiments**: An Open Letter. 22 mar. 2023. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>. Acesso em: 16 dez. 2025.
- GLASER, Eliane. **Bullshit Jobs**: A Theory by David Graeber review - the myth of capitalist efficiency. The Guardian, 25 May 2018. Disponível em: <https://www.theguardian.com/books/2018/may/25/bullshit-jobs-a-theory-by-david-graeber-review>. Acesso em: 16 dez. 2025.
- GREENWOOD, Shannon; PERRIN, Andrew; DUGGAN, Maeve. **Social Media Update 2016**. Washington, D.C.: Pew Research Center, 11 nov. 2016. Disponível em: <https://www.pewresearch.org/internet/2016/11/11/social-media-update-2016/>. Acesso em: 16 dez. 2025.
- GUIDOTTI, Riccardo; MONREALE, Anna; RUGGIERI, Salvatore; TURINI, Franco; PEDRESCHI, Dino; GIANNOTTI, Fosca. **A survey of methods for explaining black box models**. arXiv:1802.01933 [cs.CY], 6 fev. 2018. Disponível em: <https://arxiv.org/abs/1802.01933>. Acesso em: 16 dez. 2025.
- INTO THE BLACK BOX. **Into the Black Box**: A Collective Research on Logistics, Spaces, Labour. Disponível em: <http://www.intotheblackbox.com/>. Acesso em: 16 dez. 2025.

LATOUR, Bruno; WOOLGAR, Steve. **Laboratory life**: The construction of scientific facts. Princeton: Princeton University Press, 1979.

LI, Pengfei; YANG, Jianyi; ISLAM, Mohammad A.; REN, Shaolei. **Making AI Less “Thirsty”**: Uncovering and Addressing the Secret Water Footprint of AI Models. arXiv preprint arXiv:2304.03271 [cs.LG], 2023. Disponível em: <https://arxiv.org/pdf/2304.03271>. Acesso em: 16 dez. 2025.

MARAZZI, Christian. **Il posto dei calzini**: La svolta linguistica dell’economia e i suoi effetti nella politica. Bellinzona: Edizioni Casagrande, 1994.

MOROZOV, Evgeny. The true threat of artificial intelligence. **The New York Times**, 30 jun. 2023. Disponível em: <https://www.nytimes.com/2023/06/30/opinion/artificial-intelligence-danger.html>. Acesso em: 16 dez. 2025.

PASQUINELLI, Matteo. **The Eye of the Master**: A Social History of Artificial Intelligence. London: Verso Books, 2023.

WORLD ECONOMIC FORUM. **The Future of Jobs Report 2023**. Genebra: World Economic Forum, 30 abr. 2023. Disponível em: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>. Acesso em: 16 dez. 2025.

GREENWOOD, Shannon; PERRIN, Andrew; DUGGAN, Maeve. **Social Media Update 2016**. Washington, D.C.: Pew Research Center, 11 nov. 2016. Disponível em: <https://www.pewresearch.org/internet/2016/11/11/social-media-update-2016/>. Acesso em: 16 dez. 2025.