poliTICs 38 -- editorial

v. 06-06-2024

Falhas algorítmicas com resultados ou injustiças graves para as pessoas já eram descritas muito antes da era dos chatbots, como os casos descritos por Cathy O'Neil oito anos antes do anúncio do chatGPT.[1]

O texto imprescindível de Garrison Lovely faz uma detalhada resenha das visões e percalços do que se pode chamar de nova era da IA -- a era dos sistemas de IA generativa baseada em modelos linguísticos maciços (os LLMs, "large language models"), na percepção de seus criadores. Lovely descreve um dos maiores cientistas de "deep learning" expressando a visão pessimista que, tal como temos sido agentes da extinção de várias espécies, estamos agora a caminho da auto-extinção com a ajuda da IA.

Essa preocupação foi corroborada por um manifesto de alerta assinado pelos cientistas de ponta envolvidos com IA, cujo título praticamente sintetiza o documento: "A mitigação do risco de extinção provocado pela IA deve ser uma prioridade global, juntamente com outros riscos à escala social, como pandemias e guerra nuclear".[2]

Cory Doctorow menciona os riscos financeiros no campo da IA generativa. São infraestruturas de processamento, comunicação e armazenagem na escala dos grandes garimpos de criptomoedas, consumindo imensas quantidades de energia elétrica e requerendo manutenção de milhares de processadores e dispositivos de rede. É interessante lembrar que Steve Song toca no mesmo problema quanto à viabilidade econômica da rede de satélites LEO de Elon Musk.[3]

Esses riscos se somam aos problemas legais advindos de resultados falsos ou fraudes no atendimento por algoritmos.

Cory ainda lembra dos problemas na combinação entre operadores ou monitores humanos e os robôs de IA. Trata-se da IA assistida por humanos, ou seu reverso -- humanos assistidos por IA. Lembra também que a IA pode ser usada em aplicações de baixo risco -- por exemplo, design de jogos assistidos por IA. Ele descreve um exemplo de aplicação de alto risco com sérias consequências para os consumidores -- o atendimento ao cliente por chatbots da Air Canada, que fraudou centenas de passageiros --que leva a um resultado ainda pior: a maioria optou por não acionar a empresa por causa dos custos e a lentidão dos processos.

O coração da IA é a informação, em suas várias acepções. Nina Santos analisa em detalhe como tratar a ideia de integridade da informação, partindo da conceituação de documento das Nações Unidas que caracteriza a integridade da informação em oposição à "poluição" da informação. Basicamente, o conceito deve ser compreendido pelo seu antônimo, a desinformação intencional com objetivos políticos, militares ou

1 Cathy O'Neil, *Weapons of Math Destruction*, Nova York: Crown Publishers, 2016.
2 https://www.safe.ai/work/statement-on-ai-risk
3 https://politics.org.br/pt-br/acesso-news/starlink-e-desigualdade

outros. Para um documento do PNUD, "a integridade da informação é determinada pela precisão, consistência e confiabilidade do conteúdo da informação, processos e sistemas para manter um ecossistema de informação saudável".

A autora mostra que a conceituação requer ainda muito trabalho de aprofundamento, já que a literatura sobre o assunto vem se desenvolvendo há poucos anos, e que o uso no contexto de nosso idioma é problemático porque tem sido importado do inglês sem uma acumulação local do seu significado.

O texto de Sophie Nantanda é extremamente oportuno no contexto de um ano eleitoral, tanto nos EUA como no Brasil, ao analisar as diferentes formas e métodos que podem ser usados com as técnicas de "deep fake", produção e disseminação de desinformação alimentada por IA em inúmeros formatos e canais de comunicação. A autora lembra que "soluções tecnológicas, como ferramentas de verificação de conteúdo alimentadas por IA e algoritmos de detecção de deep fake, podem ajudar a detectar e mitigar o impacto de mídias falsas geradas por IA, mas não podem resolver o problema por si só e podem introduzir problemas adicionais".

Maria Farrell e Robin Benjon defendem a regeneração da Internet, em um isomorfismo conceitual com a regeneração de florestas devastadas. Ressaltam que o processo de concentração de poder econômico em poucas empresas internacionais trilionárias afeta profundamente a própria infraestrutura da Internet, tal como os desmatamentos para agropecuária maciça destroem os ecossistemas nos territórios. Fazem assim um chamado à regeneração, ou *rewilding*, da Internet.

Os pesquisadores do NIC.br fazem uma resenha primorosa do evento NETmundial+10 e de sua origem, bem como do processo de concepção da governança da Internet e do Fórum de Governança da Internet realizado anualmente pela ONU desde 2006, em texto especial para a poliTICs.

A equipe da Dataprivacy Brasil faz um apanhado dos desafios da IA no contexto das desigualdades entre os países do Sul Global e as nações desenvolvidas, com um alerta sobre as formas de exploração da força de trabalho empregada pelas empresas de IA. Enfatiza também, como Doctorow, a importância do cuidado com o impacto ambiental da infraestrutura requerida para viabilizar os sistemas computacionais e de rede que impulsionam os LLMs e outras formas de tratamento maciço e distribuido de dados. O texto é especialmente relevante pelo conjunto de recomendações ao G20 relacionadas à governança da IA.

Todos os textos traduzidos são acompanhados por anexos dos originais em inglês.

Boa leitura!

https://jacobin.com/2024/01/can-humanity-survive-ai

A humanidade pode sobreviver à IA?

Por Garrison Lovely* 22/01/2024

Com o desenvolvimento da inteligência artificial (IA) avançando a uma velocidade vertiginosa, alguns dos homens mais ricos do mundo podem estar decidindo o destino da humanidade neste momento.

O cofundador do Google, Larry Page, acredita que a IA superinteligente é "apenas o próximo passo na evolução".[1] Na verdade, Page, que vale cerca de 120 bilhões de dólares, teria alegadamente argumentado[2] que os esforços para prevenir a extinção provocada pela IA e proteger a consciência humana são "especistas" e "absurdos sentimentais".[3]

Em julho, Richard Sutton, ex-cientista sênior do Google DeepMind – um dos pioneiros do aprendizado por reforço, um importante subcampo da IA – disse que a tecnologia "poderia acabar com nossa existência" e que "não deveríamos resistir à sucessão".[4] Em uma palestra de 2015, Sutton perguntou: "É tão ruim que os humanos não sejam a forma final de vida inteligente no universo?", ao considerar a hipótese de que "tudo falhe" e a IA "mate todos nós".[5]

"Extinção biológica, esse não é o ponto", disse-me Sutton, de 66 anos. "A luz da humanidade e nossa compreensão, nossa inteligência – nossa consciência, se você preferir – pode continuar sem carne humana."

Yoshua Bengio, de 59 anos, é o segundo cientista vivo mais citado, conhecido por seu trabalho fundamental sobre aprendizagem profunda. Respondendo a Page e Sutton, Bengio me disse: "O que eles querem é jogar dados com o futuro da humanidade. Pessoalmente, acho que isso deveria ser criminalizado".[6] Um pouco surpreso, perguntei o que exatamente ele queria que fosse proibido, e ele disse que seriam esforços para construir "sistemas de IA que pudessem nos dominar e ser projetados para atuar em interesse próprio". Em maio, Bengio começou a escrever e a falar sobre como os sistemas avançados de IA podem tornar-se desonestos e representar um risco de extinção para a humanidade.[7]

Bengio postula que futuros sistemas de IA de nível genuinamente humano poderiam melhorar suas próprias capacidades, criando uma espécie nova e funcionalmente mais inteligente.[8] A humanidade extinguiu centenas de outras espécies, em grande parte por acidente. Ele teme que possamos ser os próximos – e ele não está sozinho.

Bengio compartilhou o Prêmio Turing 2018, o Prêmio Nobel da computação, com os colegas pioneiros do aprendizado profundo Yann LeCun e Geoffrey Hinton. Hinton, o cientista vivo mais citado, causou sensação em maio quando renunciou ao seu cargo sênior no Google para falar mais livremente sobre a possibilidade de futuros sistemas de IA poderem exterminar a humanidade.[9] Hinton e Bengio são os dois pesquisadores de IA mais proeminentes a se

1 https://www.vanityfair.com/news/2023/09/artificial-intelligence-industry-future/
2 https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html/
3 https://time.com/6310076/elon-musk-ai-walter-isaacson-biography/
4 https://www.youtube.com/watch?v=NgHFMolXs3U
5 https://www.youtube.com/watch?v=3l2frDNINog&t=1851s
6 https://scholar.google.com/citations?user=kukA0LcAAAAJ
7 https://yoshuabengio.org/2023/05/22/how-rogue-ais-may-arise/
8 https://yoshuabengio.org/2023/06/24/faq-on-catastrophic-ai-risks/
9 https://scholar.google.com/citations?user=JicYPdAAAAAJ&hl=en/

juntarem à comunidade do "risco-x". Às vezes referido como defensor da segurança da IA ou destruidor da mesma, este grupo pouco unido teme que a IA represente um risco existencial para a humanidade.

No mesmo mês em que Hinton demitiu-se da Google, centenas de investigadores de IA e figuras notáveis assinaram uma carta aberta afirmando: "Mitigar o risco de extinção representado pela IA deve ser uma prioridade global, juntamente com outros riscos de alcance social, como pandemias e guerra nuclear".[10] Hinton e Bengio foram os principais signatários, seguidos pelo CEO da OpenAI, Sam Altman, e pelos chefes de outros laboratórios importantes de IA.

Hinton e Bengio foram também os primeiros autores de um documento de posicionamento de outubro alertando sobre o risco de "uma perda irreversível de controle humano sobre sistemas autônomos de IA", acompanhados por acadêmicos famosos como o prêmio Nobel Daniel Kahneman e o autor de *Sapiens* Yuval Noah Harari.[11]

LeCun, que dirige a IA na Meta, concorda que a IA a nível humano está chegando, mas disse num debate público contra Bengio sobre a extinção da IA: "Se for perigosa, não a construiremos".[12]

O aprendizado profundo alimenta os sistemas de IA mais avançados do mundo, desde o modelo de enovelamento de proteínas da DeepMind até grandes modelos de linguagem (LLMs), como o ChatGPT da OpenAI. Ninguém realmente entende como funcionam os sistemas de aprendizado profundo, mas mesmo assim seu desempenho continua a melhorar. Esses sistemas não são projetados para funcionar de acordo com um conjunto de princípios bem compreendidos, mas são "treinados" para analisar padrões em grandes conjuntos de dados, com comportamentos complexos – como a compreensão da linguagem – surgindo como consequência.[13] O desenvolvedor de IA Connor Leahy me disse: "É mais como se estivéssemos cutucando algo em uma placa de Petri" do que escrevendo um trecho de código. O documento de posicionamento de outubro alerta que "atualmente ninguém sabe como alinhar de forma confiável o comportamento da IA com valores complexos".[14]

Apesar de toda esta incerteza, as empresas de IA vêem-se numa corrida para tornar estes sistemas tão poderosos quanto possível – sem um plano viável para compreender como as coisas que estão criando realmente funcionam, ao mesmo tempo que poupam esforços na segurança para ganhar mais participação de mercado.[15] A inteligência artificial geral (IAG) é o Santo Graal para o qual os principais laboratórios de IA estão trabalhando explicitamente. IAG é frequentemente definida como um sistema que é pelo menos tão bom quanto os humanos em quase todas as tarefas intelectuais. É também o que Bengio e Hinton acreditam que pode levar ao fim da humanidade.

Estranhamente, muitas das pessoas que promovem ativamente as capacidades de IA pensam que há uma chance significativa de que isso acabe causando o apocalipse. Uma pesquisa de 2022 com pesquisadores de aprendizado de máquina descobriu que quase metade deles pensava que havia pelo menos 10% de chance de que a IA avançada pudesse levar à "extinção

10 https://www.safe.ai/statement-on-ai-risk/
11 https://arxiv.org/pdf/2310.17688.pdf/
12 https://www.youtube.com/watch?v=144uOfr4SYA/
13 https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/
14 https://arxiv.org/pdf/2310.17688.pdf#page=2
15 https://www.bloomberg.com/news/features/2023-04-19/google-bard-ai-chatbot-raises-ethical-concerns-from-employees

humana ou ao enfraquecimento igualmente permanente e grave" da humanidade.[16] Poucos meses antes de tornar-se cofundador da OpenAI, Altman disse: "A IA provavelmente levará ao fim do mundo, mas, enquanto isso, haverá grandes empresas".[17]

A opinião pública sobre a IA azedou, especialmente no ano desde o lançamento do ChatGPT. Em todas as pesquisas de 2023, exceto uma, mais americanos pensaram que a IA poderia representar uma ameaça existencial para a humanidade. Nos raros casos em que os investigadores perguntaram às pessoas se queriam a IA de nível humano ou além, a grande maioria nos Estados Unidos e no Reino Unido disse que não.[18]

Até agora, quando os socialistas opinam sobre a IA, é geralmente para realçar a discriminação alimentada pela IA ou para alertar sobre o impacto potencialmente negativo da automação num mundo de sindicatos fracos e capitalistas poderosos. Mas a esquerda tem estado visivelmente calada sobre o cenário de pesadelo de Hinton e Bengio – que a IA avançada poderia matar-nos a todos.

Capacidades preocupantes

Embora grande parte da atenção da comunidade do risco-x concentre-se na ideia de que a humanidade poderá eventualmente perder o controle da IA, muitos também estão preocupados com o fato de sistemas menos capazes empoderarem maus atores em prazos muito curtos.

Felizmente, é difícil fabricar uma arma biológica. Mas isso pode mudar em breve. Anthropic, um laboratório líder de IA fundado por ex-funcionários da OpenAI, trabalhou recentemente com especialistas em biossegurança para ver o quanto um LLM poderia ajudar um aspirante a bioterrorista.[19] Testemunhando perante um subcomitê do Senado em julho, o CEO da Anthropic, Dario Amodei, relatou que certas etapas na produção de armas biológicas não podem ser encontradas em livros didáticos ou mecanismos de busca, mas que "as ferramentas de IA de hoje podem preencher algumas dessas etapas, embora de forma incompleta", e que "uma extrapolação direta dos sistemas atuais para aqueles que esperamos ver dentro de dois a três anos sugere um risco substancial de que os sistemas de IA sejam capazes de fornecer todas as peças que faltam."[20]

Em outubro de 2023 a *New Scientist* informou que a Ucrânia fez o primeiro uso no campo de batalha de armas autônomas letais (LAWs) – literalmente robôs assassinos. Os Estados Unidos, a China e Israel estão desenvolvendo as suas próprias leis.[21] A Rússia juntou-se aos Estados Unidos e a Israel na oposição ao novo direito internacional sobre as LAWs.

No entanto, a ideia mais ampla de que a IA representa um risco existencial tem muitos críticos, e o turbulento discurso da IA é difícil de analisar: pessoas igualmente credenciadas fazem afirmações opostas sobre se o risco-x da IA é real, e os capitalistas de risco estão assinando cartas abertas com eticistas de IA progressistas.[22] E embora a ideia do risco-x pareça estar ganhando terreno mais rapidamente, uma importante publicação traz um ensaio aparentemente todas as semanas argumentando que o risco-x desvia a atenção dos danos existentes. Enquanto isso, muito mais dinheiro e pessoas estão discretamente dedicados a

16 https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/#Extinction_from_AI
17 https://futureoflife.org/ai/sam-altman-investing-in-ai-safety-research/
18 https://acrobat.adobe.com/id/urn:aaid:sc:VA6C2:a01a156b-36de-4eec-929e-f085673c5b51
19 https://www.anthropic.com/news/frontier-threats-red-teaming-for-ai-safety
20 https://www.youtube.com/watch?v=IXNA-ZhJayg
21 https://www.nytimes.com/2023/11/21/us/politics/ai-drones-war-law.html
22 https://open.mozilla.org/letter/signatures/

tornar os sistemas de IA mais poderosos do que a torná-los mais seguros ou menos tendenciosos.

Alguns temem não o cenário de "ficção científica", onde os modelos de IA se tornam tão capazes que arrancam o controle de nossas mãos, mas em vez disso, que confiaremos demasiada responsabilidade a sistemas tendenciosos,[23] frágeis[24] e confabuladores,[25] abrindo uma caixa de Pandora mais pedestre, cheia de problemas terríveis, mas familiares, que aumentam de acordo com os algoritmos que os causam. Esta comunidade de investigadores e defensores – muitas vezes rotulada de "ética em IA" – tende a concentrar-se nos danos imediatos causados pela IA, explorando soluções que envolvem a responsabilização do modelo, a transparência algorítmica e a justiça da aprendizagem automática.

Falei com algumas das vozes mais proeminentes da comunidade de ética em IA, como as cientistas da computação Joy Buolamwini,[26] 33 anos, e Inioluwa Deborah Raji,[27] 27 anos. Cada uma conduziu pesquisas inovadoras sobre os danos existentes causados por modelos de IA discriminatórios e falhos, cujos impactos, na sua opinião, são obscurecidos num dia e exagerados no dia seguinte. Tal como o de muitos investigadores de ética em IA, o seu trabalho combina ciência e ativismo.

Aqueles com quem conversei no mundo da ética da IA expressaram em grande parte a opinião de que, em vez de enfrentar desafios fundamentalmente novos, como a perspectiva de desemprego tecnológico completo ou extinção,[28] o futuro da IA se parece mais com uma discriminação racial intensificada no encarceramento[29] e nas decisões de empréstimos,[30] na amazonificação dos locais de trabalho,[31] ataques aos trabalhadores pobres[32] e uma elite tecnológica ainda mais enraizada[33] e enriquecida.[34]

Um argumento frequente desta multidão é que a narrativa da extinção exagera as capacidades dos produtos da Big Tech e perigosamente "desvia a atenção" dos danos imediatos da IA.[35] Na melhor das hipóteses, dizem eles, alimentar a ideia do risco-x é uma perda de tempo e dinheiro. Na pior das hipóteses, leva a ideias políticas desastrosas.

Mas muitos dos que acreditam no risco-x realçaram que as posições "a IA causa danos agora" e "a IA pode acabar com o mundo" não são mutuamente exclusivas. Alguns investigadores tentaram explicitamente colmatar a divisão entre aqueles que se concentram nos danos existentes e aqueles que se concentram na extinção, destacando potenciais objetivos políticos partilhados.[36] O professor de IA Sam Bowman, outra pessoa cujo nome está na citada carta de

23 https://mitpress.mit.edu/9780262548328/more-than-a-glitch/
24 https://www.nature.com/articles/d41586-019-03013-5
25 https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)
26 https://www.caa.com/caaspeakers/joy-buolamwini/
27 https://en.wikipedia.org/wiki/Deborah_Raji
28 https://jacobin.com/2023/04/artificial-intelligence-chatgpt-job-loss-displacement-labor-reallocation-welfare-state
29 https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/
30 https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms
31 https://www.reuters.com/legal/legalindustry/watched-while-working-use-monitoring-ai-workplace-increases-2023-04-25/
32 https://jacobin.com/2023/07/artificial-intelligence-working-poor-australia-robodebt-welfare-exploitation
33 https://www.technologyreview.com/2023/04/18/1071727/generative-ai-risks-concentrating-big-techs-power-heres-how-to-stop-it/
34 https://theconversation.com/ai-will-increase-inequality-and-raise-tough-questions-about-humanity-economists-warn-203056
35 https://www.newscientist.com/article/mg25834453-300-the-real-reason-claims-about-the-existential-risk-of-ai-are-scary/
36 https://www.nature.com/articles/s42256-018-0003-2

extinção, fez pesquisas para revelar e reduzir o viés algorítmico e revisou os trabalhos para a principal conferência de ética em IA. Simultaneamente, Bowman apelou a que mais pesquisadores trabalhassem na segurança da IA e escreveu sobre os "perigos de subestimar" as capacidades dos LLMs.[37]

A comunidade do risco-x invoca habitualmente a defesa do clima como uma analogia, perguntando se o foco na redução dos danos a longo prazo das alterações climáticas desvia perigosamente a atenção dos danos a curto prazo da poluição atmosférica e dos derrames de petróleo.

Mas, como eles próprios admitem, nem todos do lado do risco-x foram tão diplomáticos. Em um fio de discussão agressivo de agosto de 2022 sobre políticas de IA, o cofundador da Anthropic, Jack Clark, tuitou que "Algumas pessoas que trabalham em políticas de longo prazo/estilo IAG tendem a ignorar, minimizar ou simplesmente não considerar os problemas imediatos de implantação/danos de IA."[38]

"A IA salvará o mundo"

Um terceiro grupo preocupa-se com o fato de, quando se trata de IA, não estarmos avançando suficientemente rápido. Capitalistas proeminentes como o bilionário Marc Andreessen concordam com o pessoal da segurança de que a IAG é possível, mas argumentam que, em vez de nos matar a todos, inaugurará uma idade de ouro indefinida de abundância radical e tecnologias quase mágicas.[39] Este grupo, em grande parte oriundo de Silicon Valley e comummente referido como impulsionador da IA, tende a preocupar-se muito mais com o fato de a reação regulatória exagerada sobre a IA sufocar uma tecnologia transformadora e salvadora do mundo no seu berço, condenando a humanidade à estagnação econômica.

Alguns tecno-otimistas imaginam uma utopia alimentada pela IA que faz Karl Marx parecer sem imaginação. O *Guardian* lançou recentemente um minidocumentário com entrevistas de 2016 a 2019 com o cientista-chefe da OpenAI, Ilya Sutskever, que declara resolutamente: "A IA resolverá todos os problemas que temos hoje. Resolverá o emprego, resolverá as doenças, resolverá a pobreza. Mas também criará novos problemas."[40]

Andreessen está com Sutskever – até o "mas". Em junho, Andreessen publicou um ensaio chamado "Por que a IA salvará o mundo",[41] onde explica como a IA tornará "melhor tudo o que nos interessa", desde que não a regulemos até a morte. Ele seguiu em outubro com seu "Manifesto Tecno-Otimista",[42] que, além de elogiar um fundador do fascismo italiano, nomeou como inimigas do progresso ideias como "risco existencial", "sustentabilidade", "confiança e segurança" e "ética tecnológica". Andreessen não mede palavras, escrevendo: "Acreditamos que qualquer desaceleração da IA custará vidas. Mortes que poderiam ser evitadas pela IA que foi impedida de existir são uma forma de assassinato."

Andreessen, juntamente com o "mano farmacêutico" Martin Shkreli, é talvez o mais famoso proponente do "aceleracionismo eficaz",[43] também chamado em inglês de "e/acc", uma rede majoritariamente online que mistura cientificismo de culto, hipercapitalismo e a falácia naturalista. E/acc, que se tornou viral neste verão, baseia-se na teoria do aceleracionismo do

37 https://arxiv.org/abs/2110.08300
38 https://twitter.com/jackclarkSF/status/1555990394545905665
39 https://a16z.com/ai-will-save-the-world/
40 https://www.theguardian.com/technology/ng-interactive/2023/nov/02/ilya-the-ai-scientist-shaping-the-world
41 https://a16z.com/ai-will-save-the-world/
42 https://a16z.com/the-techno-optimist-manifesto/
43 https://www.thetimes.co.uk/article/silicon-valley-tells-the-decels-to-stand-aside-for-the-ai-revolution-l2cc38cwm

escritor reacionário Nick Land, que argumenta que precisamos intensificar o capitalismo para nos impulsionarmos para um futuro pós-humano, movido pela IA. E/acc pega essa ideia e adiciona uma camada de física e memes, integrando-a a um certo subconjunto das elites do Vale do Silício. Foi formada em reação aos apelos de "desaceleração" para diminuir o ritmo de desenvolvimento da IA, que vieram significativamente da comunidade do altruísmo eficaz (AE), da qual a e/acc deriva o seu nome.

O impulsionador de IA Richard Sutton – o cientista pronto para se despedir dos "humanos de carne" – está agora trabalhando na Keen AGI, uma nova start-up de John Carmack, o lendário programador por trás do videogame Doom dos anos 1990. A missão da empresa, segundo Carmack: "IAG ou fracasso, por meio da Ciência Alucinada!"[44]

O capitalismo torna tudo pior

Em Fevereiro, Sam Altman tuitou que Eliezer Yudkowsky poderia eventualmente "merecer o Prémio Nobel da Paz".[45] Por que? Porque Altman pensava que o pesquisador autodidata e autor de *fan-fiction* de Harry Potter tinha feito "mais para acelerar a IAG do que qualquer outra pessoa". Altman citou como Yudkowsky ajudou a DeepMind a garantir o financiamento fundamental em estágio inicial de Peter Thiel, bem como o papel "crítico" de Yudkowsky "na decisão de iniciar o OpenAI".[46]

Yudkowsky era um aceleracionista antes mesmo de o termo ser cunhado. Aos dezessete anos – farto das ditaduras, da fome mundial e até da própria morte – publicou um manifesto exigindo a criação de uma superinteligência digital para "resolver" todos os problemas da humanidade.[47] Durante a década seguinte da sua vida, a sua "tecnofilia" transformou-se em fobia e, em 2008, escreveu sobre a sua história de conversão, admitindo que "dizer que quase destruí o mundo! teria sido demasiado elogioso".[48]

Yudkowsky agora é famoso por popularizar a ideia de que a IAG poderia matar todos, e ele tornou-se o mais destruidor dos destruidores da IA. Uma geração de técnicos cresceu lendo as postagens do blog de Yudkowsky, mas muitos deles (talvez mais consequentemente, Altman) internalizaram seus argumentos de que a IAG seria a coisa mais importante do que suas crenças sobre o quão difícil seria fazer com que ela não nos matasse..

Durante nossa conversa, Yudkowsky comparou a IA a uma máquina que "imprime ouro" até "acender a atmosfera". E, independentemente de inflamar ou não a atmosfera, essa máquina está imprimindo ouro mais rápido do que nunca. O boom da "IA generativa" está tornando algumas pessoas muito, muito ricas. Desde 2019, a Microsoft investiu um total acumulado de US$ 13 bilhões na OpenAI.[49] Impulsionada pelo grande sucesso do ChatGPT, a Microsoft ganhou quase US$ 1 trilhão em valor no ano seguinte ao lançamento do produto.[50] Hoje, a empresa com quase cinquenta anos vale mais do que o Google e o Meta juntos.

Os atores que maximizam os lucros continuarão avançando, externalizando riscos que o resto de nós nunca concordou em suportar, na busca de riquezas -- ou simplesmente da glória de criar superinteligência digital que, como Sutton tuitou, "será a maior conquista intelectual de todos os tempos... cujo significado está além da humanidade, além da vida, além do bem e do

44 https://twitter.com/ID_AA_Carmack/status/1560729970510422016
45 https://twitter.com/sama/status/1621621725791404032
46 https://www.penguinrandomhouse.com/books/565698/genius-makers-by-cade-metz/
47 http://www.fairpoint.net/~jpierce/staring_into_the_singularity.htm
48 https://www.lesswrong.com/posts/fLRPeXihRaiRo5dyX/the-magnitude-of-his-own-folly
49 https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html
50 https://companiesmarketcap.com/microsoft/marketcap/

mal."[51] As pressões do mercado provavelmente levarão as empresas a transferir cada vez mais poder e autonomia para os sistemas de IA à medida que estes avancem.

Um pesquisador de IA do Google me escreveu: "Acho que as grandes empresas estão com tanta pressa para ganhar participação no mercado que a segurança é vista como uma espécie de distração boba". Bengio me disse que vê "uma corrida perigosa entre empresas" que pode ficar ainda pior.

Em pânico em resposta ao mecanismo de busca Bing, baseado em OpenAI, o Google declarou um "código vermelho", "recalibrou" seu apetite ao risco e correu para liberar Bard, seu LLM, apesar da oposição da equipe.[52] Nas discussões internas, os funcionários chamavam Bard de "mentiroso patológico" e "digno de vergonha". O Google publicou de qualquer maneira.[53]

Dan Hendrycks, diretor do Centro para Segurança da IA, disse que "reduzir custos na segurança... é em grande parte o que impulsiona o desenvolvimento da IA... Na verdade, não creio que, na presença destas intensas pressões competitivas, as intenções sejam particularmente importantes."[54] Ironicamente, Hendrycks também é consultor de segurança do xAI, o mais recente empreendimento de Elon Musk.

Os três principais laboratórios de IA começaram como organizações independentes e orientadas para missões específicas, mas agora são subsidiárias integrais de gigantes da tecnologia (Google DeepMind) ou assumiram tantos bilhões de dólares em investimentos de empresas trilionárias que as suas missões altruístas podem ter sido substituidas pela busca incessante por valor para os acionistas (a Anthropic recebeu até US$ 6 bilhões do Google e da Amazon combinados,[55] e os US$13 bilhões da Microsoft compraram para a empresa 49% do braço com fins lucrativos da OpenAI[56]). O *New York Times* informou recentemente que os fundadores da DeepMind ficaram "cada vez mais preocupados com o que o Google faria com suas invenções. Em 2017, eles tentaram se separar da empresa. O Google respondeu aumentando os salários e pacotes de prêmios em ações dos fundadores da DeepMind e de sua equipe. Eles não se mexeram."[57]

Um desenvolvedor de um laboratório líder escreveu-me em outubro de 2023 que, como a liderança desses laboratórios normalmente acredita realmente que a IA evitará a necessidade de dinheiro, a busca por lucro é "em grande parte instrumental" para fins de arrecadação de fundos. Mas "então os investidores (seja uma empresa de capital de risco ou a Microsoft) exercem pressão na busca de lucro".

Entre 2020 e 2022, mais de 600 bilhões de dólares em investimento empresarial fluiram para a indústria,[58] e uma única conferência sobre IA em 2021 acolheu quase trinta mil pesquisadores.[59] Ao mesmo tempo, uma estimativa de setembro de 2022 revelou apenas

51 https://twitter.com/RichardSSutton/status/1575619655778983936?s=20
52 https://www.nytimes.com/2023/01/20/technology/google-chatgpt-artificial-intelligence.html
53 https://www.bloomberg.com/news/features/2023-04-19/google-bard-ai-chatbot-raises-ethical-concerns-from-employees
54 https://player.fm/series/future-of-life-institute-podcast/dan-hendrycks-on-catastrophic-ai-risks?t=2100
55 https://www.livemint.com/ai/artificial-intelligence/after-amazon-google-doubles-down-on-ai-set-to-invest-2-billion-in-openai-rival-startup-antropic-11698460928342.html
56 https://www.reuters.com/technology/microsoft-take-non-voting-observer-position-openais-board-2023-11-30/
57 https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html
58 https://ourworldindata.org/grapher/corporate-investment-in-artificial-intelligence-by-type
59 https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf#page=42

quatrocentos pesquisadores de segurança da IA em tempo integral,[60] e a principal conferência de ética da IA teve menos de novecentos participantes em 2023.[61]

Da mesma forma que o software "devorou o mundo", deveríamos esperar que a IA exibisse uma dinâmica semelhante em que o vencedor leva tudo, que levará a concentrações ainda maiores de riqueza e poder.[62] Altman previu que o "custo da inteligência" cairá para perto de zero como resultado da IA, e em 2021 escreveu que "ainda mais poder passará do trabalho para o capital".[63] Ele continuou: "Se as políticas públicas não se adaptarem adequadamente, a maioria das pessoas acabará em situação pior do que está hoje". Também em seu tópico "apimentado", Jack Clark escreveu: "O capitalismo de economia de escala é, por natureza, antidemocrático, e a IA com uso intensivo de investimentos é, portanto, antidemocrática".[64]

Markus Anderljung é o chefe de política do GovAI, um importante *think-tank* de segurança de IA, e o primeiro autor de um influente *white paper* focado na regulamentação da "IA de fronteira". Ele escreveu-me e disse: "Se estamos preocupados com o capitalismo na sua forma atual, deveríamos estar ainda mais preocupados com um mundo onde grande parte da economia é gerida por sistemas de IA explicitamente treinados para maximizar o lucro".

Sam Altman, por volta de junho de 2021, concordou, contando a Ezra Klein sobre a filosofia fundadora da OpenAI: "Um dos incentivos que nos deixava muito nervosos era o incentivo ao lucro ilimitado, onde mais é sempre melhor. . . . E acho que com esses sistemas de IA de uso geral muito poderosos, em particular, você não quer um incentivo para maximizar o lucro indefinidamente."[65]

Em um movimento impressionante que se tornou amplamente visto como o maior ponto crítico no debate sobre segurança da IA até agora, o conselho da organização sem fins lucrativos da OpenAI demitiu o CEO Sam Altman em 17 de novembro de 2023, sexta-feira antes do Dia de Ação de Graças. O conselho, de acordo com o estatuto incomum da OpenAI , tem um dever fiduciário para com a "humanidade", e não para com investidores ou funcionários.[66] Como justificativa, o conselho citou vagamente a falta de franqueza de Altman, mas depois, ironicamente, manteve silêncio sobre a sua decisão.

Por volta das 3h da manhã da segunda-feira seguinte, a Microsoft anunciou que Altman criaria um laboratório de pesquisa avançada com cargos para todos os funcionários da OpenAI,[67] a grande maioria dos quais assinou uma carta ameaçando aceitar a oferta da Microsoft se Altman não fosse reintegrado.[68] (Embora ele pareça ser um CEO popular, vale a pena notar que a demissão interrompeu uma venda planejada de ações de propriedade de funcionários da OpenAI avaliada em US$ 86 bilhões.) Pouco depois da 1h da manhã de quarta-feira, a OpenAI anunciou o retorno de Altman como CEO e dois novos membros do conselho: o ex-presidente do conselho do Twitter e o ex-secretário do Tesouro Larry Summers.

60 https://forum.effectivealtruism.org/posts/3gmkrj3khJHndYGNe/estimating-the-current-and-future-number-of-ai-safety
61 https://facctconference.org/2023/
62 https://a16z.com/why-software-is-eating-the-world/
63 https://moores.samaltman.com/
64 https://twitter.com/jackclarkSF/status/1555981780506722305?s=20
65 https://www.nytimes.com/2021/06/11/podcasts/transcript-ezra-klein-interviews-sam-altman.html
66 https://arstechnica.com/information-technology/2023/11/report-sutskever-led-board-coup-at-openai-that-ousted-altman-over-ai-safety-concerns/
67 https://x.com/satyanadella/status/1726509045803336122
68 https://www.nytimes.com/interactive/2023/11/20/technology/letter-to-the-open-ai-board.html

Em menos de uma semana, os executivos da OpenAI e Altman colaboraram com a Microsoft e a equipe da empresa para planejar seu retorno bem-sucedido e a remoção da maioria dos membros do conselho responsáveis por sua demissão.[69] A primeira preferência da Microsoft era ter Altman de volta como CEO. A demissão inesperada inicialmente fez com que as ações do gigante da tecnologia despencassem 5% (US$ 140 bilhões),[70] e o anúncio da reintegração de Altman levou-as a um nível mais alto de todos os tempos.[71] Relutante em ser "pega de surpresa" novamente, a Microsoft está agora ocupando um lugar sem direito a voto no conselho da organização sem fins lucrativos.[72]

Imediatamente após a demissão de Altman, X explodiu, e surgiu uma narrativa amplamente alimentada por rumores online e artigos de fontes anônimas de que altruístas eficazes focados na segurança no conselho haviam demitido Altman por sua comercialização agressiva dos modelos da OpenAI em detrimento da segurança. Capturando o teor da resposta esmagadora do e/acc, o então fundador de pseudônimo @BasedBeffJezos postou: "AEs são basicamente terroristas. Destruir 80 bilhões de valor durante a noite é um ato de terrorismo."[73]

A imagem que emergiu dos relatórios subsequentes foi que uma desconfiança fundamental em Altman, e não preocupações imediatas sobre a segurança da IA, motivou a escolha do conselho. O *Wall Street Journal* descobriu que "não houve um incidente que os levou à decisão de expulsar Altman, mas uma erosão lenta e consistente da confiança ao longo do tempo que os deixou cada vez mais inquietos".[74]

Semanas antes da demissão, Altman teria usado táticas desonestas para tentar remover a conselheira Helen Toner por causa de um artigo acadêmico de sua autoria que ele considerava crítico ao compromisso da OpenAI com a segurança da IA. No artigo, Toner, uma pesquisadora de governança de IA alinhada à AE, elogiou a Anthropic por evitar "o tipo de corte frenético que o lançamento do ChatGPT pareceu estimular".

A *New Yorker* relatou que "alguns dos seis membros do conselho consideraram Altman manipulador e conivente".[75] Dias após a demissão, um pesquisador de segurança de IA da DeepMind que trabalhava para a OpenAI escreveu que Altman "mentiu para mim em várias ocasiões" e "era enganoso, manipulador e pior para os outros", uma avaliação ecoada por reportagens recentes na *Time*.[76]

Esta não foi a primeira vez que Altman foi demitido. Em 2019, o fundador da Y Combinator, Paul Graham, removeu Altman do comando da incubadora por preocupações de que ele estava priorizando seus próprios interesses em detrimento dos da organização. Graham disse anteriormente : "Sam é extremamente bom em se tornar poderoso".[77]

O estranho modelo de governança da OpenAI foi estabelecido especificamente para evitar a influência corruptora da procura de lucro, mas, como a Atlantic proclamou corretamente, "o

69 https://www.wsj.com/tech/ai/altman-firing-openai-520a3a8c
70 https://www.thebusinessanecdote.com/post/microsoft-stock-risks-further-decline-after-openai-sacks-sam-altman
71 https://www.documentcloud.org/documents/24174939-screen-shot-2023-11-24-at-24854-pm
72 https://www.bloomberg.com/news/articles/2023-11-18/openai-altman-ouster-followed-debates-between-altman-board
73 https://twitter.com/BasedBeffJezos/status/1726654431096385978
74 https://www.wsj.com/tech/ai/altman-firing-openai-520a3a8c
75 https://www.newyorker.com/magazine/2023/12/11/the-inside-story-of-microsofts-partnership-with-openai/
76 https://twitter.com/geoffreyirving/status/1726754277618491416
77 https://www.newyorker.com/magazine/2016/10/10/sam-altmans-manifest-destiny

dinheiro vence sempre".[78] E mais dinheiro do que nunca está sendo investido no avanço das capacidades da IA.

Velocidade máxima à frente

O progresso recente da IA foi impulsionado pelo culminar de muitas tendências de décadas: aumento no poder de computação e dados utilizados para treinar modelos de IA, que foram amplificados por melhorias significativas na eficiência algorítmica.[79] Desde 2010, o poder computacional para treinar modelos de IA aumentou cerca de *cem milhões de vezes*.[80] A maioria dos avanços que vemos agora são produto de um campo que na época era muito menor e mais pobre.[81]

E embora o ano passado tenha certamente tido mais do que o seu quinhão de entusiasmo pela IA,[82] a confluência destas três tendências levou a resultados quantificáveis. O tempo que os sistemas de IA levam para atingir um desempenho de nível humano em muitas tarefas de referência diminuiu drasticamente na última década.[83]

É possível, claro, que os ganhos de capacidade da IA cheguem a um impasse. Os pesquisadores podem ficar sem bons dados para usar. A lei de Moore – a observação de que o número de transistores num microchip duplica a cada dois anos – acabará por se tornar história.[84] Os acontecimentos políticos podem perturbar as cadeias de produção e de abastecimento, aumentando os custos do setor de informática. E a ampliação dos sistemas pode não levar mais a um melhor desempenho.

Mas a realidade é que ninguém conhece os verdadeiros limites das abordagens existentes. Um clipe de uma entrevista de Yann LeCun em janeiro de 2022 reapareceu no Twitter este ano.[85] LeCun disse: "Não acho que possamos treinar uma máquina para ser inteligente puramente a partir de texto, porque acho que a quantidade de informação sobre o mundo contida no texto é pequena em comparação com o que precisamos saber". Para ilustrar seu argumento, ele deu um exemplo: "Pego um objeto, coloco em cima da mesa e empurro a mesa. É completamente óbvio para você que o objeto seria empurrado com a mesa." Porém, com "um modelo baseado em texto, se você treinar uma máquina, por mais poderosa que seja, seu 'GPT-5000'. . . nunca vai aprender sobre isso."

Mas se você der esse exemplo ao ChatGPT-3.5, ele instantaneamente fornecerá a resposta correta.

Em uma entrevista publicada quatro dias antes de sua demissão, Altman disse: "Até treinarmos esse modelo, será como um divertido jogo de adivinhação para nós. Estamos tentando melhorar nisso, porque acho importante, do ponto de vista da segurança, prever as capacidades. Mas não posso dizer exatamente o que o GPT-4 fará e o que o GPT-4 não fez."[86]

A história está repleta de previsões erradas sobre o ritmo da inovação. Um editorial do *New York Times* afirmou que poderia levar "um milhão a dez milhões de anos" para desenvolver

78 https://www.theatlantic.com/technology/archive/2023/11/sam-altman-open-ai-microsoft-investment-profit/676077/
79 https://time.com/6300942/ai-progress-charts/
80 https://asteriskmag.com/issues/03/the-great-inflection-a-debate-about-ai-and-explosive-growth/
81 https://ourworldindata.org/ai-investments
82 https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were/
83 https://contextual.ai/plotting-progress-in-ai/
84 https://physicsworld.com/a/moores-law-further-progress-will-push-hard-on-the-boundaries-of-physics-and-economics/
85 https://twitter.com/YaBoyFathoM/status/1659516423540965378
86 https://www.ft.com/content/dd9ba2f6-f509-42f0-8e97-4271c7b84ded

uma máquina voadora – sessenta e nove dias antes dos irmãos Wright voarem pela primeira vez.[87] Em 1933, Ernest Rutherford, o "pai da física nuclear", descartou com segurança a possibilidade de uma reação em cadeia induzida por nêutrons, inspirando o físico Leo Szilard a formular a hipótese de uma solução funcional no dia seguinte – uma solução que acabou sendo fundamental para a criação da bomba atômica.[88]

Uma conclusão que parece difícil de evitar é que, recentemente, as pessoas que são melhores na construção de sistemas de IA acreditam que a IAG é possível e iminente. Talvez os dois principais laboratórios de IA, OpenAI e DeepMind, tenham trabalhado em IAG desde o seu início, começando num momento em que admitir que você acreditava que a IAG era possível em breve poderia fazer você rir da sala. (Ilya Sutskever liderou um canto de "Feel the AGI" na festa de Natal de 2022 da OpenAI.[89])

Trabalhadores Perfeitos

Os empregadores já usam a IA para vigiar,[90] controlar[91] e explorar[92] os trabalhadores. Mas o verdadeiro sonho é tirar os humanos do circuito. Afinal, como escreveu Marx: "A máquina é um meio de produzir mais-valia".

A pesquisadora de risco de IA da Open Philanthropy (OP), Ajeya Cotra, escreveu-me que "o ponto final lógico de uma economia capitalista ou de mercado maximamente eficiente" não envolveria humanos porque "os humanos são apenas criaturas muito ineficientes para ganhar dinheiro". Valorizamos todas estas emoções "comercialmente improdutivas", escreve ela, "por isso, se acabarmos por nos divertir e gostar do resultado, será porque começamos com o poder e moldamos o sistema para se acomodar aos valores humanos. "

OP é uma fundação inspirada na AE, financiada pelo cofundador do Facebook, Dustin Moskovitz. É o principal financiador de organizações de segurança de IA, muitas das quais são mencionadas neste artigo.[93] A OP também concedeu US$ 30 milhões à OpenAI para apoiar o trabalho de segurança de IA dois anos antes de o laboratório criar um braço com fins lucrativos em 2019.[94] Anteriormente, recebi uma doação única para apoiar o trabalho de publicação na *New York Focus*, uma organização de notícias investigativas sem fins lucrativos que cobre a política de Nova York, dos *EA Funds*, que recebe financiamento da OP. Depois de conhecer a AE pela primeira vez em 2017, comecei a doar 10 a 20 por cento do meu rendimento para organizações sem fins lucrativos globais de saúde e anti-agroindústria, voluntariei-me como organizador de grupo local e trabalhei numa organização sem fins lucrativos global contra a pobreza. A AE foi uma das primeiras comunidades a se envolver seriamente com o risco existencial da IA, mas olhei para o pessoal da IA com alguma cautela, dada a incerteza do problema e o sofrimento imenso e evitável que está acontecendo agora.

Uma IAG complacente seria o trabalhador com que os capitalistas só podem sonhar: incansável, motivado e sem necessidade de pausas para ir ao banheiro. Gestores, de Frederick

87 https://www.snopes.com/fact-check/wright-brothers-first-flight/
88 https://blogs.scientificamerican.com/the-curious-wavefunction/leo-szilard-a-traffic-light-and-a-slice-of-nuclear-history/
89 https://www.theatlantic.com/technology/archive/2023/11/sam-altman-open-ai-chatgpt-chaos/676050/
90 https://www.newstatesman.com/spotlight/tech-regulation/cybersecurity/2023/02/amazon-workers-staff-surveillance-extreme-stress-anxiety
91 https://www.amazon.com/Algorithm-Decides-Hired-Monitored-Promoted/dp/0306827344#:~:text=Book overview&text=In The Algorithm%2C she investigates,and who receives a promotion.
92 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4331080
93 https://forum.effectivealtruism.org/posts/XdhwXppfqrpPL2YDX/an-overview-of-the-ai-safety-funding-situation
94 https://openai.com/blog/openai-lp

Taylor a Jeff Bezos, ressentem-se das diversas maneiras pelas quais os humanos não são otimizados para a produção – e, portanto, para os resultados financeiros de seus empregadores. Mesmo antes dos dias da gestão científica de Taylor, o capitalismo industrial tem procurado tornar os trabalhadores mais parecidos com as máquinas com as quais trabalham e pelas quais são cada vez mais substituídos. Como observou prescientemente o *Manifesto Comunista*, o uso extensivo de maquinaria pelos capitalistas transforma o trabalhador num "apêndice da máquina".

Mas, de acordo com a comunidade de segurança da IA, as mesmas capacidades desumanas que fariam Bezos salivar também tornam a IAG um perigo mortal para os humanos.

Explosão: O Caso de Extinção

O argumento comum do risco-x é: assim que os sistemas de IA atingirem um determinado limite, eles serão capazes de se auto-aperfeiçoar recursivamente, dando início a uma "explosão de inteligência". Se um novo sistema de IA tornar-se inteligente – ou apenas ampliado – o suficiente, será capaz de enfraquecer permanentemente a humanidade.

O documento de posicionamento de outubro afirma: "Não há razão fundamental para que o progresso da IA diminua ou pare quando atingir capacidades de nível humano... Em comparação com os humanos, os sistemas de IA podem agir mais rapidamente, absorver mais conhecimento e comunicar com uma largura de banda muito maior. Além disso, eles podem ser dimensionados para usar imensos recursos computacionais e podem ser replicados aos milhões."[95]

Esses recursos já possibilitaram habilidades sobre-humanas: os LLMs podem "ler" grande parte da Internet em meses, e o AlphaFold da DeepMind pode realizar anos de trabalho humano de laboratório em poucos dias.

Aqui está uma versão estilizada da ideia de crescimento "populacional" estimulando uma explosão de inteligência: se os sistemas de IA rivalizarem com os cientistas humanos em pesquisa e desenvolvimento, os sistemas proliferarão rapidamente, levando ao equivalente a um enorme número de novos trabalhadores altamente produtivos entrando na economia. Dito de outra forma, se o GPT-7 puder executar a maioria das tarefas de um trabalhador humano e custar apenas alguns dólares para colocar o modelo treinado para trabalhar nas tarefas diárias, cada instância do modelo seria extremamente lucrativa, disparando um ciclo de feedback positivo. Isto poderia levar a uma "população" virtual de bilhões ou mais de trabalhadores digitais, cada um valendo muito mais do que o custo da energia necessária para o seu funcionamento.[96] Sutskever acredita que é provável que "toda a superfície da Terra seja coberta por painéis solares e data centers".[97]

Estes trabalhadores digitais poderão ser capazes de melhorar os nossos projetos de IA e iniciar o seu caminho para a criação de sistemas "superinteligentes", cujas capacidades Alan Turing especulou em 1951 que em breve "ultrapassariam os nossos fracos poderes".[98] E, como argumentam alguns defensores da segurança da IA , um modelo individual de IA não precisa ser superinteligente para representar uma ameaça existencial; pode ser necessário apenas

95 https://arxiv.org/pdf/2310.17688.pdf#page=2
96 https://www.cold-takes.com/ai-could-defeat-all-of-us-combined/#fnref6
97 https://www.youtube.com/watch?si=1tZQyCeH5RxkG5rW&t=620&v=9iqn1HhFJ6c&feature=youtu.be
98 https://rauterberg.employee.id.tue.nl/lecturenotes/DDM110 CAS/Turing/Turing-1951 Intelligent Machinery-a Heretical Theory.pdf

haver cópias suficientes dele.[99] Muitas das minhas fontes compararam as corporações a superinteligências, cujas capacidades excedem claramente as dos seus membros constituintes.

"Basta desconectá-los", diz a objeção comum. Mas quando um modelo de IA for suficientemente poderoso para ameaçar a humanidade, será provavelmente a coisa mais valiosa que existe. Talvez seja mais fácil "desconectar" a Bolsa de Valores de Nova York ou a Amazon Web Services.

Uma superinteligência preguiçosa pode não representar um grande risco, e céticos como Oren Etzioni, CEO do Allen Institute for AI, a professora de estudos de complexidade Melanie Mitchell e a diretora-gerente do AI Now Institute, Sarah Myers West, me disseram que não viram evidências convincentes de que os sistemas de IA são tornando-se mais autônomos. Dario Amodei, da Anthropic, parece concordar que os sistemas atuais não apresentam um nível preocupante de agência.[100] No entanto, um sistema completamente passivo, mas suficientemente poderoso, exercido por um malfeitor é suficiente para preocupar pessoas como Bengio.

Além disso, tanto os académicos como os industriais estão a aumentar os esforços para tornar os modelos de IA mais autónomos. Dias antes da sua demissão, Altman disse ao *Financial Times*: "Tornaremos estes agentes cada vez mais poderosos... e as ações ficarão cada vez mais complexas a partir daqui... A quantidade de valor comercial que resultará da capacidade de fazer isso em todas as categorias, eu acho, é muito boa."[101]

O que há por trás do Hype?

O medo que mantém muitas pessoas do risco-x acordadas à noite não é que uma IA avançada "acorde", "torne-se má" e decida matar todos por maldade, mas sim que ela passe a nos ver como um obstáculo para quaisquer que sejam os objetivos que tenha. Em seu último livro, *Brief Answers to the Big Questions*, Stephen Hawking articulou isso, dizendo : "Você provavelmente não é um odiador de formigas ruim que pisa em formigas por maldade, mas se você é responsável pelo projeto de uma hidrelétrica verde e tem um formigueiro na região para ser inundado, azar das formigas."[102]

Comportamentos inesperados e indesejáveis podem resultar de objetivos simples, seja o lucro ou uma função de recompensa da IA. Num mercado "livre", a procura do lucro leva a monopólios, esquemas de marketing multinível, ar e rios envenenados e inúmeros outros danos.

Existem exemplos abundantes de sistemas de IA que apresentam comportamentos surpreendentes e indesejados.[103] Um programa destinado a eliminar erros de classificação em uma lista excluiu totalmente a lista.[104] Um pesquisador ficou surpreso ao encontrar um modelo de IA "fingir-se de morto" para evitar ser identificado em testes de segurança.[105]

Outros ainda veem uma conspiração da Big Tech surgindo por trás dessas preocupações. Algumas pessoas focadas nos danos imediatos da IA argumentam que a indústria está promovendo ativamente a ideia de que seus produtos podem acabar com o mundo, como

99 https://www.cold-takes.com/ai-could-defeat-all-of-us-combined/#how-ais-could-defeat-humans-without-superintelligence
100 https://www.dwarkeshpatel.com/p/dario-amodei
101 https://www.ft.com/content/dd9ba2f6-f509-42f0-8e97-4271c7b84ded
102 https://www.vox.com/future-perfect/2018/10/16/17978596/stephen-hawking-ai-climate-change-robots-future-universe-earth
103 https://arxiv.org/pdf/2206.07682.pdf
104 https://arxiv.org/pdf/1803.03453.pdf#page=7
105 https://direct.mit.edu/artl/article/26/2/274/93255/The-Surprising-Creativity-of-Digital-Evolution-A

Myers West, do *AI Now Institute*, que diz "ver as narrativas em torno do chamado risco existencial como realmente um estratagema para tirar todo o ar da sala, a fim de garantir que não haja movimento significativo no momento presente." Estranhamente, Yann LeCun[106] e o cientista-chefe da IA do Baidu, Andrew Ng,[107] afirmam concordar.

Quando apresentei a ideia aos que acreditam no risco-x, eles muitas vezes responderam com uma mistura de confusão e exasperação. Ajeya Cotra, da OP, respondeu: "Gostaria que a preocupação com o risco-x fosse menos associada à indústria, porque acho que é fundamentalmente, no mérito, uma crença muito anti-indústria... Se as empresas estão construindo coisas que vão matar todos nós, isso é muito ruim, e elas deveriam ser restringidas de forma muito rigorosa pela lei."

Markus Anderljung, do GovAI, chamou os receios de captura regulatória de "uma reação natural que as pessoas têm", mas enfatizou que as suas políticas preferidas podem muito bem prejudicar os maiores agentes da indústria.

Uma fonte compreensível de suspeita é que Sam Altman é agora uma das pessoas mais associadas à ideia do risco existencial, mas a sua empresa fez mais do que qualquer outra para avançar a fronteira da IA de uso geral.

Além disso, à medida que a OpenAI se aproximava da lucratividade e Altman se aproximava do poder, o CEO mudou seu tom público. Em uma sessão de perguntas e respostas de janeiro de 2023, quando questionado sobre seu pior cenário para IA, ele respondeu : "Luzes apagadas para todos nós".[108] Mas ao responder a uma pergunta semelhante sob juramento perante senadores em maio, Altman não menciona a extinção.[109] E, talvez em sua última entrevista antes de sua demissão, Altman disse : "Na verdade, não acho que todos seremos extintos. Eu acho que vai ser ótimo. Acho que estamos caminhando para o melhor mundo de todos."[110]

Altman implorou ao Congresso em maio que regulamentasse a indústria de IA,[111] mas uma investigação de novembro[112] descobriu que a Microsoft, empresa quase controladora da OpenAI, foi influente no lobby malsucedido para excluir "modelos fundacionais" como o ChatGPT da regulamentação da próxima Lei de IA da União Europeia.[113] E Altman fez muito lobby na UE,[114] ameaçando mesmo retirar-se da região se as regulamentações se tornassem demasiado onerosas (ameaças que ele rapidamente desconsiderou[115]). Falando num painel de CEOs em São Francisco, dias antes de sua demissão, Altman disse que "os modelos atuais estão bem. Não precisamos de regulamentação pesada aqui. Provavelmente nem mesmo nas próximas gerações."[116]

A recente ordem executiva "abrangente" do presidente Biden sobre IA parece concordar: os seus requisitos de repasse de informações de testes de segurança afetam apenas modelos

106 https://twitter.com/ylecun/status/1718670073391378694?s=20
107 https://www.afr.com/technology/google-brain-founder-says-big-tech-is-lying-about-ai-human-extinction-danger-20231027-p5efnz
108 https://www.youtube.com/watch?si=WmG9zM-4nzfXqyoK&t=1340&v=ebjkD1Om4uw&feature=youtu.be
109 https://www.techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai/
110 https://www.nytimes.com/2023/11/20/podcasts/hard-fork-sam-altman-transcript.html
111 https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html
112 https://corporateeurope.org/en/2023/11/byte-byte
113 https://time.com/6338602/eu-ai-regulation-foundation-models/
114 https://www.politico.eu/article/openai-ceo-to-meet-commission-president-in-brussels/
115 https://twitter.com/sama/status/1661975237280567297
116 https://sfstandard.com/2023/11/17/openai-sam-altman-fired-apec-talk

maiores do que qualquer outro que provavelmente tenha sido treinado até agora.[117] Myers West chamou esses tipos de "limiares de escala" de "exclusão massiva". Anderljung escreveu-me que a regulamentação deveria ser escalonada de acordo com as capacidades e a utilização de um sistema, e disse que "gostaria de alguma regulamentação dos modelos mais capazes e amplamente utilizados de hoje", mas acha que "será muito mais politicamente viável impor requisitos a sistemas que ainda não foram desenvolvidos."

Inioluwa Deborah Raji arriscou que se os gigantes da tecnologia "souberem que têm que ser os bandidos em alguma dimensão... eles prefeririam que fosse abstrato e de longo prazo." Isso me parece muito mais plausível do que a ideia de que a Big Tech realmente queira promover a ideia de que seus produtos têm uma chance razoável de *literalmente matar todo mundo* .

Quase setecentas pessoas assinaram a carta de extinção, a maioria delas acadêmicas.[118] Apenas uma delas dirige uma empresa de capital aberto: o financiador do OP, Moskovitz, que também é cofundador e CEO da Asana, um aplicativo de produtividade. Não havia nenhum funcionário da Amazon, Apple, IBM ou de qualquer empresa líder de hardware de IA. Nenhum executivo da Meta assinou.

Se os chefes das grandes empresas tecnológicas queriam amplificar a narrativa da extinção, por que não acrescentaram os seus nomes à lista?

Por que construir a "Máquina do Fim dos Tempos?"

Se a IA realmente salvar o mundo, quem quer que a tenha criado poderá esperar ser elogiado como um Júlio César moderno. E mesmo que isso não aconteça, quem primeiro construir "a última invenção que o homem alguma vez precisou fazer" não terá de se preocupar em ser esquecido pela história – a menos, claro, que a história termine abruptamente após a sua invenção.[119]

Connor Leahy pensa que, no nosso caminho atual, o fim da história seguirá em breve o advento da IAG. Com seu cabelo esvoaçante e cavanhaque despenteado, ele provavelmente se sentiria em casa usando uma placa onde se lê "O fim está próximo" – embora isso não o tenha impedido de ser convidado para discursar na Câmara dos Lordes britânica ou na CNN. O CEO da Conjecture, de 28 anos, e cofundador da EleutherAI, um influente coletivo de código aberto, me disse que grande parte da motivação para construir IA se resume a: "'Oh, você está construindo a máquina de destruição definitiva que rende bilhões de dólares e também o torna rei-imperador da terra ou mata todo mundo?' Sim, isso é como o sonho masculino. Você fica tipo, 'Porra, sim. Eu sou o rei da desgraça.'" Ele continua: "Tipo, eu entendi. Isso está muito presente na estética do Vale do Silício."

Leahy também transmitiu algo que não surpreenderá as pessoas que passaram um tempo significativo na Bay Area ou em certos cantos da Internet:

Existem empresários e tecnólogos tecnoutópicos reais, completamente irresponsáveis, não eleitos, vivendo principalmente em São Francisco, que estão dispostos a arriscar a vida de vocês, de seus filhos, de seus netos e de toda a humanidade futura, só porque podem ter uma chance. viver para sempre.

117 https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/

118 https://www.safe.ai/statement-on-ai-risk

119 https://quoteinvestigator.com/2022/01/04/ultraintelligent/

Em março, o MIT Technology Review informou que Altman "diz que esvaziou sua conta bancária para financiar dois... objetivos: energia ilimitada e prolongar a vida."[120]

Tendo em conta tudo isto, seria de esperar que a comunidade da ética visse a comunidade de segurança como um aliado natural numa luta comum para controlar as elites tecnológicas irresponsáveis que estão consrtruindo unilateralmente produtos arriscados e prejudiciais. E, como vimos anteriormente, muitos defensores da segurança fizeram aberturas aos especialistas em ética da IA. Também é raro que pessoas da comunidade de risco-x ataquem publicamente a ética da IA (embora o inverso seja... não verdadeiro[121]), mas a realidade é que os defensores da segurança têm sido por vezes difíceis de engolir.

Os especialistas em ética da IA, tal como as pessoas que ekes defendem, muitas vezes relatam que se sentem marginalizadas e afastadas do poder real, travando uma batalha difícil com empresas de tecnologia que as vêem como uma forma de se protegerem e não como uma verdadeira prioridade. Para dar crédito a esse sentimento está a destruição das equipes de ética de IA em muitas grandes empresas de tecnologia nos últimos anos (ou dias). E, em vários casos, estas empresas retaliaram contra denunciantes[122] e organizadores sindicais orientados pela ética.[123]

Isso não implica necessariamente que essas empresas estejam priorizando seriamente o risco-x. O conselho de ética do Google DeepMind, que incluía Larry Page e o proeminente pesquisador de risco existencial Toby Ord, teve sua primeira reunião em 2015, mas nunca teve uma segunda.[124] Um pesquisador de IA do Google me escreveu que "não falam sobre riscos de longo prazo... no escritório", continuando, "o Google está mais focado na construção de tecnologia e na segurança no sentido de legalidade e agressividade".

A engenheira de software Timnit Gebru co-liderou a equipe ética de IA do Google até ser forçada a deixar a empresa no final de 2020, após uma disputa sobre um rascunho de documento – agora uma das publicações de aprendizado de máquina mais famosas de todos os tempos. No artigo sobre "papagaios estocásticos",[125] Gebru e seus co-autores argumentam que os LLMs prejudicam o meio ambiente, amplificam os preconceitos sociais e usam estatísticas para unir a linguagem "ao acaso" "sem qualquer referência ao significado".

Gebru, que não é fã da comunidade de segurança de IA, pediu proteções aprimoradas de denunciantes para pesquisadores de IA,[126] o que também é uma das principais recomendações feitas no white paper do GovAI.[127] Desde que Gebru foi expulsa do Google, quase 2.700 funcionários assinaram uma carta solidária,[128] mas o "googler" Geoff Hinton não era um deles. Quando questionado na CNN por que ele não apoiava um colega denunciante, Hinton respondeu que as críticas de Gebru à IA "eram preocupações bastante diferentes das

120 https://www.technologyreview.com/2023/03/08/1069523/sam-altman-investment-180-million-retro-biosciences-longevity-death/
121 https://thedigradio.com/podcast/ai-hype-machine-w-meredith-whittaker-ed-ongweso-and-sarah-west/
122 https://www.theguardian.com/technology/2021/oct/08/tech-whistleblowers-facebook-frances-haugen-amazon-google-pinterest
123 https://www.nytimes.com/2019/04/22/technology/google-walkout-employees-retaliation.html
124 https://www.nytimes.com/2023/12/03/technology/ai-openai-musk-page-altman.html
125 https://dl.acm.org/doi/epdf/10.1145/3442188.3445922
126 https://mitsloan.mit.edu/ideas-made-to-matter/ex-google-researcher-ai-workers-need-whistleblower-protection
127 https://arxiv.org/pdf/2307.03718.pdf
128 https://googlewalkout.medium.com/standing-with-dr-timnit-gebru-isupporttimnit-believeblackwomen-6dadc300d382

minhas" que "não são tão existencialmente sérias quanto a ideia de que essas coisas se tornem mais inteligentes do que nós e assumam o controle."[129]

Raji disse-me que "muitos dos motivos de frustração e animosidade" entre os campos da ética e da segurança é que "um lado tem muito mais dinheiro e poder do que o outro lado", o que "permite que eles impulsionem a sua agenda de forma mais direta."

De acordo com uma estimativa, a quantidade de dinheiro transferida para startups e organizações sem fins lucrativos de segurança de IA em 2022 quadruplicou desde 2020, atingindo US$144 milhões.[130] É difícil encontrar um número equivalente para a comunidade ética da IA. Contudo, a sociedade civil de ambos os campos é ofuscada pelos gastos da indústria. Apenas no primeiro trimestre de 2023, a OpenSecrets relatou que cerca de US$94 milhões foram gastos em lobby de IA nos Estados Unidos.[131] A LobbyControl estimou que as empresas tecnológicas gastaram €113 milhões em 2023 em lobby junto da UE,[132] e recordemos que centenas de bilhões de dólares estão sendo investidos na indústria da IA neste momento.

Algo que pode impulsionar a animosidade ainda mais do que qualquer diferença percebida em poder e dinheiro é a linha de tendência. Após livros amplamente elogiados como *Weapons of Math Destruction* de 2016, da cientista de dados Cathy O'Neil, e descobertas bombásticas de preconceito algorítmico, como o artigo "Gender Shades" de 2018 de Buolamwini e Gebru, a perspectiva da ética da IA capturou a atenção e o apoio do público.[133]

Em 2014, a causa do risco-x da IA teve o seu próprio best-seller surpresa, *Superinteligência* do filósofo Nick Bostrom, que argumentou que a IA além-humana poderia levar à extinção -- e recebeu elogios de figuras como Elon Musk e Bill Gates. Mas Yudkowsky disse-me que, antes do ChatGPT, fora de certos círculos do Vale do Silício, considerar seriamente a tese do livro faria as pessoas olharem para você de forma engraçada. Os primeiros proponentes da segurança da IA, como Yudkowsky, ocuparam a estranha posição de manter laços estreitos com a riqueza e o poder através dos técnicos da Bay Area, permanecendo, ao mesmo tempo, marginalizados no discurso mais amplo.

No mundo pós-ChatGPT, os ganhadores do prêmio Turing e os que ganharam o Nobel estão saindo do armário de segurança da IA e abraçando argumentos popularizados por Yudkowsky, cuja publicação mais conhecida é uma peça de *fan-fiction* de Harry Potter, totalizando mais de 660.000 palavras.[134]

Talvez o presságio mais chocante deste novo mundo tenha sido transmitido em novembro, quando os apresentadores de um podcast de tecnologia do *New York Times*, *Hard Fork*, perguntaram à presidente da Comissão Federal de Comércio: "Qual é a sua estimativa, Lina Khan? Qual é a sua probabilidade de que a IA mate todos nós?" A conversa sobre bebedouros da AE se tornou popular. (Khan disse que é "otimista" e deu uma estimativa "baixa" de *15%*.)[135]

---

129 https://www.youtube.com/watch?si=LvVptspe2TbGHV2F&t=80&v=FAbsoxQtUwM&feature=youtu.be
130 https://docs.google.com/spreadsheets/d/1C_QDlzZynG00u7XVHy91Tii9qOl-dk8KtxiYcrd_ZYc/edit#gid=0
131 https://www.opensecrets.org/news/2023/05/big-tech-lobbying-on-ai-regulation-as-industry-races-to-harness-chatgpt-popularity/
132 https://corporateeurope.org/en/2023/09/lobbying-power-amazon-google-and-co-continues-grow
133 http://gendershades.org/
134 https://managing-ai-risks.com/
135 https://www.nytimes.com/2023/11/10/podcasts/hardfork-chatbot-ftc.html?showTranscript=1

Seria fácil observar todas as cartas abertas e círculos mediáticos e pensar que a maioria dos investigadores em IA estão mobilizados contra o risco existencial. Mas quando perguntei a Bengio sobre como o risco-x é percebido hoje na comunidade de aprendizado de máquina, ele disse: "Ah, mudou muito. Costumava ser cerca de 0,1% das pessoas que prestavam atenção à questão. E talvez agora seja 5%."

Probabilidades

Como muitos outros preocupados com o risco-x da IA, o renomado filósofo da mente David Chalmers apresentou um argumento probabilístico durante nossa conversa: "Esta não é uma situação em que você tenha que estar 100% certo de que teremos uma IA de nível humano para se preocupar. sobre isso. Se for 5%, temos que nos preocupar com isso."

Esse tipo de pensamento estatístico é popular na comunidade de AE e é um motivo relevante que levou seus membros a se concentrarem na IA em primeiro lugar. Se você aceitar argumentos de especialistas, poderá ficar mais confuso. Mas se você tentar calcular a média da preocupação dos especialistas a partir de um punhado[136] de pesquisas,[137] poderá acabar pensando que há pelo menos uma pequena chance de que a extinção pela IA possa acontecer, o que poderia ser suficiente para torná-la a coisa mais importante do mundo. E se atribuirmos algum valor a todas as gerações futuras que possam existir, a extinção humana é categoricamente pior do que catástrofes passíveis de sobrevivência.

No entanto, no debate sobre IA, abundam as alegações de arrogância. Céticos como Melanie Mitchell e Oren Etzioni disseram-me que não havia provas que apoiassem o argumento do risco-x, enquanto crentes como Bengio e Leahy apontam para ganhos de capacidade surpreendentes e perguntam: e se o progresso não parar? Um amigo pesquisador acadêmico de IA comparou o advento da IAG a jogar a economia e a política globais em um liquidificador.

Mesmo que, por alguma razão, a IAG só consiga igualar e não exceder a inteligência humana, a perspectiva de partilhar a Terra com um número quase arbitrariamente grande de agentes digitais de nível humano é assustadora, especialmente quando provavelmente estarão tentando ganhar dinheiro de alguém.

Existem demasiadas ideias políticas sobre como reduzir o risco existencial da IA para serem discutidas adequadamente aqui. Mas uma das mensagens mais claras provenientes da comunidade de segurança da IA é que devemos "desacelerar". Os defensores de tal desaceleração esperam que ela dê aos decisores políticos e à sociedade em geral a oportunidade de recuperar o atraso e decidir ativamente como uma tecnologia potencialmente transformadora é desenvolvida e implementada.[138]

Cooperação internacional

Uma das respostas mais comuns a qualquer esforço para regulamentar a IA é a objeção "mas e a China!?" Altman, por exemplo, disse numa comissão do Senado em maio que "queremos que a América lidere" e reconheceu que o perigo de abrandar é que "a China ou qualquer outra pessoa faça progressos mais rápidos".[139]

Anderljung escreveu-me que esta "não é uma razão suficientemente forte para não regulamentar a IA".

---

136 https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/#Extinction_from_AI
137 https://www.alignmentforum.org/posts/QvwSr5LsxyDeaPK5s/existential-risk-from-ai-survey-results
138 https://pauseai.info/
139 https://www.techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai/

Num artigo da *Foreign Affairs* de junho, Helen Toner e dois cientistas políticos relataram que os investigadores chineses de IA que entrevistaram pensavam que os LLM chineses estão pelo menos dois a três anos atrás dos modelos de última geração americanos.[140] Além disso, os autores argumentam que, uma vez que os avanços da IA chinesa "dependem muito da reprodução e do ajuste da investigação publicada no estrangeiro", um abrandamento unilateral "provavelmente desaceleraria" também o progresso chinês. A China também agiu mais rapidamente do que qualquer outro grande país para regulamentar significativamente a IA,[141] como observou o chefe de políticas da Anthropic, Jack Clark.[142]

Yudkowsky diz: "Na verdade, não é do interesse da China cometer suicídio juntamente com o resto da humanidade".

Se a IA avançada realmente ameaça o mundo inteiro, a regulamentação interna por si só não será suficiente. Mas restrições nacionais robustas poderiam sinalizar de forma credível a outros países a seriedade com que encaramos os riscos. O proeminente especialista em ética em IA, Rumman Chowdhury, pediu supervisão global. Bengio diz que "temos que fazer as duas coisas".[143]

Yudkowsky, sem surpresa, assumiu uma posição maximalista, dizendo-me que "a direção correta parece mais colocar todo o hardware de IA num número limitado de centros de dados sob supervisão internacional por órgãos com um tratado simétrico segundo o qual ninguém – incluindo os militares, governos , a China ou a CIA – podem fazer qualquer uma das coisas realmente terríveis, incluindo a construção de superinteligências."[144]

Num controverso artigo de opinião da *Time* de março de 2024, Yudkowsky defendeu "encerrar tudo", propondo uma moratória internacional sobre "novas grandes sessões de treino" apoiadas pela ameaça da força militar.[145] Dadas as fortes crenças de Yudkowsky de que a IA avançada seria muito mais perigosa do que qualquer arma nuclear ou biológica, esta posição radical surge naturalmente.

Todos os vinte e oito países presentes na recente Cúpula de Segurança da IA, incluindo os Estados Unidos e a China, assinaram a Declaração de Bletchley, que reconheceu os danos existentes da IA e o fato de que "riscos substanciais podem surgir do potencial uso indevido intencional ou de problemas de controle não intencionais relacionados com o alinhamento com a intenção humana."[146]

Na Cúpula, o governo britânico anfitrião encarregou Bengio de liderar a produção do primeiro relatório "Estado da Ciência" sobre as "capacidades e riscos da IA de fronteira", num passo significativo em direção a um órgão especializado permanente como o Painel Intergovernamental sobre Alterações Climáticas.[147]

A cooperação entre os Estados Unidos e a China será imperativa para uma coordenação internacional significativa no desenvolvimento da IA. E quando se trata de IA, os dois países

140 https://www.foreignaffairs.com/china/illusion-chinas-ai-prowess-regulation-helen-toner
141 https://time.com/6304831/china-ai-regulations/
142 https://twitter.com/jackclarkSF/status/1555980989297410049?s=20
143 https://www.wired.com/story/ai-desperately-needs-global-oversight/
144 https://twitter.com/ESYudkowsky/status/1719777049576128542
145 https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/
146 https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
147 https://www.gov.uk/government/publications/ai-safety-summit-2023-chairs-statement-state-of-the-science-2-november/state-of-the-science-report-to-understand-capabilities-and-risks-of-frontier-ai-statement-by-the-chair-2-november-2023

não estão exatamente nas melhores condições. Com os controles de exportação da Lei CHIPS de 2022, os Estados Unidos tentaram limitar as capacidades de IA da China, algo que um analista da indústria teria anteriormente considerado um "ato de guerra".[148] Tal como a Jacobin relatou em maio de 2023,[149] alguns pesquisadsores de políticas orientados para o risco-x provavelmente desempenharam um papel na aprovação dos onerosos controles. Em outubro, os Estados Unidos reforçaram as restrições da Lei CHIPS para preencher lacunas.

No entanto, num sinal encorajador, Biden e Xi Jinping discutiram a segurança da IA e a proibição da IA em sistemas de armas letais em novembro. Um comunicado de imprensa da Casa Branca declarou: "Os líderes afirmaram a necessidade de abordar os riscos dos sistemas avançados de IA e melhorar a segurança da IA através de conversações governamentais entre os EUA e a China".[150]

As armas letais autônomas são também uma área de relativo acordo nos debates sobre IA. Em seu novo livro *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*, Joy Buolamwini defende a campanha "Stop Killer Robots", ecoando uma preocupação de longa data de muitos defensores da segurança da IA. O *Future of Life Institute*, uma organização sobre o risco-x, reuniu opositores ideológicos para assinar uma carta aberta de 2016 apelando à proibição de leis ofensivas, incluindo Bengio, Hinton, Sutton, Etzioni, LeCun, Musk, Hawking e Noam Chomsky.[151]

Um assento à mesa

Após anos de inação, os governos mundiais estão finalmente voltando a sua atenção para a IA.[152] Mas ao não se envolverem seriamente naquilo que os sistemas futuros poderão fazer, os socialistas estão cedendo o seu lugar à mesa.

Em grande parte devido aos tipos de pessoas que se sentiram atraídas pela IA, muitos dos primeiros partidários sérios da ideia do risco-x decidiram envolver-se em investigação extremamente teórica sobre como controlar a IA avançada ou criaram empresas de IA.[153] Mas para um outro tipo de pessoa, a resposta ao acreditar que a IA pode acabar com o mundo é tentar *fazer com que as pessoas parem de construí-la*.

Os defensores continuam dizendo que o desenvolvimento da IA é inevitável – e se um número suficiente de pessoas acreditar, isso se tornará verdade. Mas "não há nada inevitável na inteligência artificial", escreve o *AI Now Institute*.[154] O diretor administrativo Myers West repetiu isso, mencionando que a tecnologia de reconhecimento facial parecia inevitável em 2018, mas desde então foi proibida em muitos lugares.[155] E, como aponta Katja Grace, pesquisadora do risco-x , não deveríamos sentir a necessidade de construir todas as tecnologias simplesmente porque podemos.[156]

148 https://www.nytimes.com/2023/07/12/magazine/semiconductor-chips-us-china.html
149 https://jacobin.com/2023/05/longtermism-new-cold-war-biden-administration-china-semiconductors-ai-policy
150 https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/15/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-2/
151 https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/
152 https://www.ft.com/content/59b9ef36-771f-4f91-89d1-ef89f4a2ec4e
153 https://intelligence.org/
154 https://ainowinstitute.org/general/2023-landscape-executive-summary
155 https://www.aclu.org/news/privacy-technology/2019-was-the-year-we-proved-face-recognition-surveillance-isnt-inevitable
156 https://worldspiritsockpuppet.substack.com/p/lets-think-about-slowing-down-ai#§restraint-is-not-radical

Além disso, muitos legisladores estão observando os avanços recentes da IA e *enlouquecendo*. O senador Mitt Romney está "mais aterrorizado com a IA" do que otimista,[157] e seu colega Chris Murphy diz: "As consequências de tantas funções humanas serem terceirizadas para a IA são potencialmente desastrosas".[158] Os congressistas Ted Lieu[159] e Mike Johnson[160] estão literalmente "assustados" com a IA. Se certos técnicos forem as únicas pessoas dispostas a reconhecer que as capacidades da IA melhoraram dramaticamente e poderão representar uma ameaça a nível de espécie no futuro, serão a eles que os decisores políticos darão ouvidos desproporcionalmente. Em maio, o professor e especialista em ética em IA Kristian Lum tuitou: "Há um risco existencial que tenho certeza que os LLMs representam e é sobre a credibilidade do campo da FAccT[161] /IA Ética, se continuarmos promovendo a narrativa do óleo de cobra sobre eles".[162]

Mesmo que a ideia da extinção impulsionada pela IA lhe pareça mais fictícia do que científica, ainda poderá haver um enorme impacto na influência sobre a forma como uma tecnologia transformadora é desenvolvida e quais os valores que ela representa. Assumir que podemos conseguir que uma IAG hipotética faça o que queremos levanta talvez a questão mais importante que a humanidade alguma vez enfrentará: O que deveríamos *querer* que ela quisesse?

Quando perguntei a Chalmers sobre isto, ele disse: "Em algum momento recapitulamos todas as questões da filosofia política: que tipo de sociedade realmente queremos e que realmente valorizamos?"

Uma forma de pensar sobre o advento da IA a nível humano é que seria como criar a constituição de um novo país (a "IA constitucional" da Anthropic interpreta esta ideia literalmente,[163] e a empresa recentemente experimentou incorporar elementos democráticos no documento fundamental do seu modelo[164]). Os governos são sistemas complexos que exercem um enorme poder. A base sobre a qual estão estabelecidas pode influenciar a vida de milhões de pessoas agora e no futuro. Os americanos vivem sob o jugo de homens mortos que tinham tanto medo do público que construíram medidas antidemocráticas, que continuam a atormentar o nosso sistema político mais de dois séculos depois.

A IA pode ser mais revolucionária do que qualquer inovação anterior. É também uma tecnologia exclusivamente normativa, dado o quanto a construímos para refletir as nossas preferências. Como Jack Clark disse recentemente à Vox: "É realmente estranho que este não seja um projeto governamental".[165] Chalmers me disse: "Quando de repente tivermos as empresas de tecnologia tentando incorporar esses objetivos em sistemas de IA, teremos que realmente confiar nas empresas de tecnologia para acertar essas questões sociais e políticas

157 https://www.romney.senate.gov/romney-leads-senate-hearing-on-addressing-potential-threats-posed-by-ai-quantum-computing-and-other-emerging-technology/#:~:text=I'm in the camp of being more terrified about AI than I am of the camp of those thinking this is going to make everything better for the world%2C

158 https://twitter.com/ChrisMurphyCT/status/1640347945018032129

159 https://lieu.house.gov/media-center/editorials/new-york-times-op-ed-i-m-congressman-who-codes-ai-freaks-me-out/

160 https://www.cnbc.com/2023/05/16/openai-ceo-woos-lawmakers-ahead-of-first-testimony-before-congress.html/

161 Do inglês "fairness, accountability and transparency." Uma conferência sobre o assunto está programada pela ACM no Rio de Janeiro em junho de 2024: *https://facctconference.org/2024/*

162 https://twitter.com/KLdivergence/status/1653843497932267520?s=20

163 https://www.anthropic.com/news/constitutional-ai-harmlessness-from-ai-feedback

164 https://www.nytimes.com/2023/10/17/technology/ai-chatbot-control.html

165 https://www.vox.com/future-perfect/23794855/anthropic-ai-openai-claude-2/

muito profundas". Não tenho certeza se sim. Ele enfatizou: “Você não está fazendo apenas uma reflexão técnica sobre isso, mas uma reflexão social e política”.

Falsas escolhas

Talvez não precisemos esperar para encontrar sistemas superinteligentes que não priorizem a humanidade. Agentes sobre-humanos otimizam implacavelmente uma recompensa às custas de qualquer outra coisa que nos interesse.[166] Quanto mais capaz for o agente e mais implacável for o otimizador,[167] mais extremos serão os resultados.

Soa familiar? Se sim, você não está sozinho. O AI Objectives Institute (AOI) considera tanto o capitalismo quanto a IA como exemplos de otimizadores desalinhados. Co-fundado pela ex-apresentadora de um programa de rádio público Brittney Gallagher[168] e pelo “herói da privacidade” Peter Eckersley pouco antes de sua morte inesperada[169], o laboratório de pesquisa examina o espaço entre a aniquilação e a utopia, “uma continuação das tendências existentes de concentração de poder em menos mãos – reforçadas pelo avanço da IA – em vez de uma ruptura brusca com o presente.”[170] O presidente da AOI, Deger Turan, me disse: “Risco existencial é a falha na coordenação diante de um risco”. Ele diz que “precisamos criar pontes" entre a segurança da IA e a ética da IA.

Uma das ideias mais influentes nos círculos de risco-x é a maldição do unilateralista, um termo para situações em que um ator solitário pode arruinar as coisas para todo o grupo.[171] Por exemplo, se um grupo de biólogos descobre uma maneira de tornar uma doença mais mortal, basta apenas um para publicá-la. Nas últimas décadas, muitas pessoas ficaram convencidas de que a IA poderia acabar com a humanidade, mas apenas os mais ambiciosos e tolerantes ao risco deram início às empresas que estão agora avançando na fronteira das capacidades da IA, ou, como Sam Altman disse recentemente, afastando o “véu da ignorância”.[172] Como alude o CEO, não temos como saber verdadeiramente o que está além do limite tecnológico.

Alguns de nós entendemos perfeitamente os riscos, mas seguimos em frente mesmo assim. Com a ajuda de cientistas de ponta, a ExxonMobil descobriu conclusivamente em 1977 que o seu produto causava o aquecimento global.[173] Eles então mentiram ao público sobre isso, enquanto construíam suas plataformas de petróleo cada vez maiores.

A ideia de que a queima de carbono poderia aquecer o clima foi levantada pela primeira vez no final do século XIX,[174] mas o consenso científico sobre as alterações climáticas demorou quase cem anos a formar-se. A ideia de que poderíamos perder permanentemente o controle para as máquinas é mais antiga do que a computação digital, mas permanece longe de um consenso científico. E se o progresso recente da IA continuar a ritmo acelerado, poderemos não ter décadas para formar um consenso antes de agirmos de forma efetiva.

O debate que se desenrola em praça pública pode levar-nos a acreditar que temos de escolher entre abordar os danos imediatos da IA e os seus riscos existenciais inerentemente especulativos. E há certamente prós e contras que requerem uma consideração cuidadosa.

---

166 https://www.currentaffairs.org/2017/11/a-public-option-for-food#:~:text=My friend Sarah,destroyed the world.
167 https://www.vox.com/the-highlight/23447596/artificial-intelligence-agi-openai-gpt3-existential-risk-human-extinction/
168 https://digitalvillage.org/about#:~:text=Digital Village Host,have inspired creators.
169 https://www.wired.com/story/peter-eckersley-ai-objectives-institute/
170 https://ai.objectives.institute/whitepaper/
171 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4959137/
172 https://sfstandard.com/2023/11/17/openai-sam-altman-fired-apec-talk/
173 https://insideclimatenews.org/news/16092015/exxons-own-research-confirmed-fossil-fuels-role-in-global-warming/
174 https://scied.ucar.edu/learning-zone/how-climate-works/history-climate-science-research/

Mas quando olhamos para as forças materiais em jogo, surge uma imagem diferente: num canto estão empresas de bilionárias tentando tornar os modelos de IA mais poderosos e lucrativos; em ooutro, encontramos grupos da sociedade civil tentando fazer com que a IA reflita valores que rotineiramente colidem com a maximização do lucro.

Em suma, é capitalismo versus humanidade.

(*) Garrison Lovely é um escritor freelancer e apresentador do podcast "The Most Interesting People I Know": https://www.garrisonlovely.com/about-me

https://pluralistic.net/2024/04/23/maximal-plausibility/#reverse-centauros

"Humanos no circuito" devem encontrar os erros mais difíceis de detectar, em velocidade sobre-humana

Cory Doctorow* -- 23/04/2024

Se a inteligência artificial (IA) tiver futuro (um grande "se"), terá de ser economicamente viável. Uma indústria não pode gastar 1.700% mais em chips Nvidia do que ganha indefinidamente – nem mesmo sendo a Nvidia o principal investidor em seus maiores clientes.[1]

Uma empresa que paga US$0,36/consulta por eletricidade e água (escassa e doce) não pode distribuir indefinidamente essas consultas aos milhões para pessoas que devem revisar essas consultas dezenas de vezes antes de obter a versão perfeita de "instruções para remover um sanduíche de queijo grelhado de um videocassete no estilo da Bíblia do Rei James".[2]

Eventualmente, a indústria terá de descobrir alguma combinação de aplicações que cubra os seus custos operacionais, nem que seja para manter as luzes acesas face à desilusão dos investidores (isto não é opcional – a desilusão dos investidores é uma parte inevitável de cada bolha).

Agora, existem *muitas* aplicações de baixo risco para IA que podem funcionar perfeitamente na atual tecnologia de IA, apesar de seus muitos – e aparentemente inevitáveis – erros ("alucinações"). Pessoas que usam IA para gerar ilustrações de seus personagens de D&D envolvidos em aventuras épicas de sua sessão de jogo anterior não se importam com um dedo extra. Se o chatbot que alimenta a ferramenta automática de conversão de texto em voz para um turista errar algumas palavras, ainda será muito melhor do que a alternativa de falar lenta e alto em seu próprio idioma enquanto faz gestos enfáticos com as mãos.

Existem muitos desses aplicativos, e muitas das pessoas que se beneficiam deles sem dúvida pagariam algo por eles. O problema – do ponto de vista de uma empresa de IA – é que estes não são apenas *riscos baixos* , mas também de baixo *valor* . Seus usuários pagariam *algo* por eles, mas não muito.

Para que a IA mantenha os seus servidores ligados durante o período de desilusão que se aproxima, terá também de localizar aplicações de *elevado valor.* Economicamente falando, a função das aplicações de baixo valor é absorver o excesso de capacidade e produzir valor nas margens depois que as aplicações de alto valor pagam as contas. As aplicações de baixo valor são um acompanhamento, como os assentos de um avião cujas despesas operacionais totais são pagas antecipadamente pelos passageiros da

1 https://news.ycombinator.com/item?id=39883571
2 https://www.semianalysis.com/p/the-inference-cost-of-search-disruption

classe executiva. Sem a principal receita dos aplicativos de alto valor, os servidores são desligados e os aplicativos de baixo valor desaparecem.[3]

Agora, existem muitas aplicações de alto valor que a indústria de IA identificou para seus produtos. Em termos gerais, estas aplicações de alto valor partilham o mesmo problema: são todas de alto risco, o que significa que são muito sensíveis a erros. Erros cometidos por aplicativos que produzem códigos, dirigem carros ou identificam massas cancerígenas em radiografias de tórax têm consequências extremamente graves.

Algumas empresas podem ser insensíveis a essas consequências. A Air Canada substituiu sua equipe humana de atendimento ao cliente por chatbots que simplesmente mentiam para os passageiros, roubando-lhes centenas de dólares no processo. Mas o processo para recuperar seu dinheiro depois de ser fraudado pelo chatbot da Air Canada é tão oneroso que apenas um passageiro se preocupou em passar por isso, passando dez semanas esgotando todos os mecanismos de revisão interna da Air Canada antes de lutar pelo seu caso por mais semanas no regulador.[4]

Nunca há apenas uma formiga. Se esse usuário foi fraudado por um chatbot da Air Canada, o mesmo aconteceu com centenas ou milhares de outros clientes. A Air Canada não precisa reembolsá-los. A empresa afirma tacitamente que, sendo a principal companhia aérea do país e quase monopolista, é demasiado grande para falir e demasiado grande para ser detida, o que significa que é demasiado grande para *se importar*.

A empresa mostra que, para alguns clientes empresariais, a IA não precisa ser capaz de realizar o trabalho de um trabalhador para ser uma compra inteligente: um chatbot pode substituir um trabalhador, falhar no trabalho do trabalhador e ainda assim economizar dinheiro para a empresa.

Não posso prever se os monopolistas sociopatas do mundo são numerosos e poderosos o suficiente para manter as luzes acesas para as empresas de IA através de contratos de sistemas de automação que lhes permitem cometer fraudes livres de consequências, substituindo trabalhadores por chatbots que servem como áreas de desmoralização para clientes furiosos.

Mas mesmo estipulando que isto é suficiente, é intrinsecamente instável. Qualquer coisa que não possa durar para sempre eventualmente para, e a substituição em massa de humanos por software fraudulento de alta velocidade parece provavelmente alimentar a já ardente fornalha do antitruste moderno.[5]

É claro que as empresas de IA têm a sua própria resposta para este enigma. Um cliente de alto risco/alto valor ainda pode demitir trabalhadores e substituí-los por IA – eles

3 https://locusmag.com/2023/12/commentary-cory-doctorow-what-kind-of-bubble-is-ai/

4 https://bc.ctvnews.ca/air-canada-s-chatbot-gave-a-b-c-man-the-wrong-information-now-the-airline-has-to-pay-for-the-mistake-1.6769454

5 https://www.eff.org/de/deeplinks/2021/08/party-its-1979-og-antitrust-back-baby

só precisam contratar menos trabalhadores mais baratos para supervisionar a IA e monitorá-la em busca de "alucinações". Essa é a solução "humano no circuito".

O humano nessa história tem algumas lacunas gritantes. Do ponto de vista do trabalhador, servir como humano num esquema que reduz as folhas salariais através da IA é um pesadelo – o pior tipo possível de automatização.

Vamos fazer uma pequena pausa na teoria da automação aqui. A automação pode *aumentar* um trabalhador. Podemos chamar isto de "centauro" – o trabalhador descarrega uma tarefa repetitiva, ou uma que requer um elevado grau de vigilância, ou (o pior de tudo) ambas. Eles são uma cabeça humana em um corpo de robô (daí "centauro"). Pense no sistema de sensor/visão do seu carro que emite um sinal sonoro se você ativar a seta enquanto um carro está no seu ponto cego. Você está no comando, mas recebe uma segunda opinião do robô.

Da mesma forma, considere uma ferramenta de IA que verifique novamente o diagnóstico de uma radiografia de tórax feita por um radiologista e sugira uma segunda análise quando sua avaliação não corresponder à do radiologista. Mais uma vez, o humano está no comando, mas o robô está servindo como proteção e auxílio, usando sua inesgotável vigilância robótica para aumentar a habilidade humana.

Isso são centauros. Eles são a boa automação. Depois, há a *má* automação: o centauro *reverso* , quando o humano é usado para ampliar a capacidade do robô.

Os selecionadores de armazém da Amazon ficam em um só lugar enquanto as estantes robóticas chegam até eles em alta velocidade; então, as pulseiras hápticas presas em seus pulsos zumbem para eles, orientando-os a pegar itens específicos e movê-los para uma cesta, enquanto um terceiro sistema de automação os penaliza por fazerem pausas para ir ao banheiro ou mesmo apenas andarem e sacudirem os membros para evitar uma lesão por esforço repetitivo. Esta é uma cabeça robótica usando um corpo humano – e destruindo-o no processo.

Um radiologista assistido por IA processa *menos* radiografias de tórax todos os dias, custando *mais ao seu empregador* , além do custo da IA. Não é isso que as empresas de IA estão vendendo. Eles estão oferecendo aos hospitais o poder de criar centauros reversos: IAs assistidas por radiologistas. Isso é o que significa "humano no circuito".

Isto é um problema para os trabalhadores, mas também é um problema para os seus chefes (assumindo que esses chefes realmente se preocupam em corrigir as alucinações da IA, em vez de fornecerem uma folha de parreira que lhes permita cometer fraudes ou matar pessoas e transferir a culpa para uma IA impunível).

Os humanos são bons em muitas coisas, mas não são bons em *vigilância eterna e perfeita* . Escrever código é difícil, mas realizar a revisão do código (onde você verifica se há erros no código de outra pessoa) é muito mais difícil - e fica *ainda mais difícil* se o

código que você está revisando *geralmente estiver* bom, porque isso exige que você mantenha sua vigilância para algo que só ocorre em intervalos raros e imprevisíveis.[6]

Mas para que uma loja de codificação reduza o custo de um lápis de IA, o ser humano envolvido precisa ser capaz de processar uma *grande quantidade* de código gerado por IA. Substituir um humano por uma IA não produz nenhuma economia se você precisar contratar mais dois humanos para se revezar na leitura detalhada do código da IA.

Esta é a falha fatal nos esquemas de táxis robóticos. O "humano no circuito" que deveria impedir o robô assassino de colidir com outros carros, entrar no trânsito em sentido contrário ou atropelar pedestres não é um motorista, é um *instrutor de direção* . Este é um trabalho *muito* mais difícil do que ser motorista, mesmo quando o aluno motorista que você está monitorando é um ser humano, cometendo erros humanos em velocidade humana. É ainda mais difícil quando o aluno motorista é um robô, cometendo erros na velocidade do computador.[7]

É por isso que a falida empresa de táxi-robô Cruise teve que implantar três monitores humanos qualificados e bem pagos para supervisionar cada dois de seus robôs assassinos, enquanto os táxis tradicionais operam por uma fração do custo com um único motorista humano, precarizado e mal pago.[8]

O problema da vigilância ja é fatal para a estratégia do "humano no circuito", mas há outro problema que é *ainda mais* fatal: os *tipos* de erros que as IAs cometem.

Fundamentalmente, a IA é estatística aplicada. Uma empresa de IA treina sua IA alimentando-a com muitos dados sobre o mundo real. O programa processa esses dados, procurando correlações estatísticas nesses dados, e cria um modelo do mundo com base nessas correlações. Um chatbot é um programa de adivinhação da próxima palavra, e um gerador de “arte” de IA é um programa de adivinhação do próximo pixel. Eles estão recorrendo a bilhões de documentos para encontrar a maneira estatisticamente mais provável de terminar uma frase ou uma linha de pixels em um gráfico.[9]

Isto significa que a IA não comete apenas erros – ela comete erros sutis, os tipos de erros que são mais difíceis de serem detectados por um ser humano no circuito, porque são as formas estatisticamente mais prováveis de dar errado. Claro, notamos erros grosseiros nos resultados da IA, como afirmar com segurança que um ser humano vivo está morto.[10]

Mas os erros mais comuns que as IAs cometem são aqueles que não percebemos, porque estão perfeitamente camuflados como verdade. Pense no erro recorrente de programação de IA que insere uma chamada para uma biblioteca inexistente chamada “huggingface-cli”, que é como a biblioteca seria chamada se os desenvolvedores

6 https://twitter.com/qntm/status/1773779967521780169
7 https://pluralistic.net/2024/04/01/human-in-the-loop/#monkey-in-the-middle
8 https://pluralistic.net/2024/01/11/robots-stole-my-jerb/#computer-says-no
9 https://dl.acm.org/doi/10.1145/3442188.3445922
10 https://www.tomsguide.com/opinion/according-to-chatgpt-im-dead

seguissem de forma confiável as convenções de nomenclatura. Mas devido a uma inconsistência humana, a biblioteca real tem um nome ligeiramente diferente. O fato de as IAs inserirem repetidamente referências à biblioteca inexistente abriu uma vulnerabilidade – um pesquisador de segurança criou uma biblioteca maliciosa (inerte) com esse nome e enganou inúmeras empresas para compilá-la em seu código porque seus revisores humanos não perceberam a mentira do chatbot (estatisticamente indistinguível da verdade).[11]

Para um instrutor de direção ou um revisor de código supervisionando um sujeito humano, a maioria dos erros é comparativamente fácil de detectar, porque são os tipos de erros que levam a nomenclatura inconsistente de bibliotecas – onde um humano se comportou de forma errática ou irregular. Mas quando *a realidade* é irregular ou errática, a IA cometerá erros ao presumir que as coisas são estatisticamente normais.

Esses são os tipos de erros mais difíceis de detectar. Eles não poderiam ser mais difíceis de serem detectados por um humano se fossem *especificamente projetados* para passar despercebidos. O ser humano envolvido não está apenas sendo solicitado a identificar erros – ele está sendo ativamente enganado. A IA não está apenas errada, ela está construindo um quebra-cabeça sutil no estilo “o que há de errado com esta imagem”. Não apenas um desses quebra-cabeças: milhões deles, em alta velocidade, que devem ser resolvidos pelo "humano no circuito", que deve permanecer perfeitamente vigilante para coisas que são, por definição, quase totalmente imperceptíveis.

Este é um novo tormento especial para os centauros invertidos – e um problema significativo para as empresas de IA que esperam acumular e manter um número suficiente de clientes de alto valor e de alto risco nas suas contas para resistir ao próximo período de desilusão.

Isso é muito sombrio, mas fica ainda mais sombrio. As empresas de IA argumentam que têm uma terceira linha de negócios, uma forma de ganhar dinheiro para os seus clientes, para além dos presentes da automação para as suas folhas de pagamento: alegam que podem realizar tarefas científicas difíceis a uma velocidade sobre-humana, produzindo resultados de milhares de milhões de dólares (novos materiais, novos medicamentos, novas proteínas) a uma velocidade inimaginável.

No entanto, estas afirmações – amplificadas com credulidade pela imprensa não técnica – continuam a desmoronar quando são testadas por especialistas que compreendem os domínios esotéricos nos quais se diz que a IA tem uma vantagem imbatível. Por exemplo, o Google afirmou que sua IA Deepmind descobriu “milhões de novos materiais”, “equivalente a quase 800 anos de conhecimento”, constituindo “uma expansão de ordem de grandeza em materiais estáveis conhecidos pela humanidade”.[12]

11 https://www.theregister.com/2024/03/28/ai_bots_hallucinate_software_packages/
12 https://deepmind.google/discover/blog/millions-of-new-materials-discovered-with-deep-learning/

Foi uma farsa. Quando cientistas de materiais independentes analisaram amostras representativas destes "novos materiais", concluíram que "nenhum material novo foi descoberto" e que nenhum destes materiais era "credível, útil e novo".[13]

Como escreve Brian Merchant, as afirmações sobre IA são assustadoramente semelhantes a "fumaça e espelhos" – o deslumbrante campo de distorção da realidade criado pela tecnologia de lanterna mágica do século XVII, ao qual milhões de pessoas atribuíram capacidades esotéricas, graças às afirmações bizarras dos promotores da tecnologia.[14]

O fato de termos um nome de quatrocentos anos para esse fenômeno, e ainda assim estarmos sendo vítimas dele, é francamente um pouco deprimente. E, para nosso azar, acontece que os bots de terapia de IA não podem nos ajudar com isso – em vez disso, eles são capazes de literalmente nos convencer a nos matar.[15]

(*) Cory Doctorow, escritor, ativista, jornalista e blogueiro, coeditor do portal Boing Boing, ex-diretor da Electronic Frontier Foundation e cofundador do Open Rights Group da Inglaterra.

13 https://www.404media.co/google-says-it-discovered-millions-of-new-materials-with-ai-human-researchers/
14 https://www.bloodinthemachine.com/p/ai-really-is-smoke-and-mirrors
15 https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says

https://www.techpolicy.press/why-do-we-need-to-discuss-sovocation-information-integrity/

Por que precisamos discutir a chamada “integridade da informação”?

Nina Santos* 03/04/2024

O termo “integridade da informação” é cada vez mais empregado, especialmente por organizações internacionais[1] e organizações que desenvolvem planos para combater a desinformação e promover a produção e divulgação de informações factuais.[2] Agora, o termo também foi adotado pelo governo brasileiro. Em 2023, foram assinados pelo menos quatro instrumentos de cooperação entre o Brasil e outros países que utilizam este termo.[3] No contexto da presidência brasileira do G20, essa ideia ganhou ainda mais destaque e norteou ações relacionadas ao combate à desinformação, ao discurso de ódio, à defesa da regulamentação das plataformas digitais e à construção de um espaço digital democrático ou saudável.

No Brasil, o uso do termo parece representar uma tentativa de deslocar o debate sobre o atual ecossistema comunicacional de uma perspectiva negativa de combate a fenômenos negativos – como a desinformação, o discurso de ódio ou as teorias da conspiração – para uma estratégia positiva e objetiva. Além disso, o governo brasileiro tem argumentado que o termo é uma oportunidade para superar conceitos politicamente sensíveis ou insuficientes para lidar com o problema da disseminação de falsidades nas mídias sociais e digitais.

Deve-se dizer também que a noção de integridade da informação transmite duas ideias importantes: em primeiro lugar, que é um debate central para as democracias contemporâneas; e segundo, que a normatividade da integridade da informação tem um viés coletivo, segundo o qual o conceito precisa ser abordado.

Tudo isto é certamente muito positivo, mas o fato é que não há – ou há muito pouca – literatura acadêmica não americana sobre a ideia de integridade da informação, o que levanta preocupações sobre preconceitos culturais e dificulta a construção teórica e política do termo. Afinal, o que significa o termo “integridade da informação”? O que é pressuposto nesta ideia e como ela se traduz em outras línguas? Quais são os parâmetros para avaliar se ele se adapta a diferentes contextos? E, acima de tudo, serve os interesses da maioria da população global? A que realmente precisamos estar atentos quando discutimos o cenário atual das comunicações sob uma perspectiva que interessa ao Sul Global?

Reconstituindo a história do termo “integridade da informação”

A expressão “integridade da informação” ganhou recentemente notoriedade global, especialmente desde o *Policy Brief 8*, publicado pelas Nações Unidas em junho de

1 https://www.ndi.org/infotegrity
2 https://www.undp.org/policy-centre/oslo/information-integrity
3 https://www.mercosur.int/pt-br/declaracao-especial-dos-presidentes-do-mercosul-sobre-democracia-e-integridade-da-informacao-em-ambientes-digitais/

2023.[4] Neste documento, “integridade da informação” refere-se à “acuidade, consistência e fiabilidade da informação, que é ameaçada pela desinformação e pelo discurso de ódio" (p.5). Também neste documento é apresentada a ideia de “integridade da informação” em oposição à “poluição da informação”.

O *Policy Brief 8*, que indica a construção de um código de conduta para integridade da informação em plataformas digitais, propõe um “compromisso com a integridade da informação”. De acordo com a ONU, isto implica que "todas as partes interessadas devem abster-se de utilizar, apoiar ou amplificar a desinformação e o discurso de ódio para qualquer fim, incluindo para alcançar objetivos políticos, militares ou outros objetivos estratégicos, incitar à violência, minar processos democráticos ou atingir populações civis, grupos, comunidades ou indivíduos vulneráveis" (p.21).

Pouco mais de um ano antes, em fevereiro de 2022, o Programa das Nações Unidas para o Desenvolvimento (PNUD) publicou o documento *Integridade da Informação: Forjando um Caminho para a Verdade, Resiliência e Confiança*.[5] O objetivo do texto é justamente tentar dar uma base para o uso do termo, e nele há, portanto, um esforço maior para conceituá-lo.

Para o PNUD, “o conceito de integridade da informação é emprestado dos sistemas corporativos, onde se refere à segurança da informação e à proteção de dados dentro das empresas. Aplicada de forma mais ampla, a integridade da informação é determinada pela precisão, consistência e confiabilidade do conteúdo da informação, processos e sistemas para manter um ecossistema de informação saudável'. Requer acesso dos cidadãos a informações confiáveis, equilibradas e completas sobre assuntos atuais, ações governamentais, atores políticos e outros elementos relevantes para suas percepções políticas e tomada de decisões" (p.4). . As referências utilizadas pelo PNUD para definir o termo são de organizações do Norte Global, incluindo citações a um documento de uma empresa privada, Yonder , que não está mais disponível na Internet;[6] e outro do Club de Madrid, uma organização global com sede na Espanha.[7]

É importante destacar que, a partir de 2021, há uma literatura crescente sobre integridade da informação, principalmente – e eu diria quase exclusivamente – de instituições e pesquisadores dos EUA e da Europa. Isto não é necessariamente um problema *em si* , mas exige que reconheçamos a história do termo – e o que está embutido nele. Isto permitir-nos-á discutir o seu significado para diferentes realidades, exigências e prioridades em contextos de Maioria Global.

Problemas com o termo tal como está

1. É preciso enfatizar o foco no espaço e no fluxo, não na unidade

A ideia de “integridade da informação”, especialmente em português, pode dar a impressão de que o foco está na unidade de informação, que precisa estar intacta. Ou seja, haveria um remetente, um produtor da informação, que publicaria uma unidade

4 https://indonesia.un.org/en/236014-our-common-agenda-policy-brief-8-information-integrity-digital-platforms
5 https://www.undp.org/publications/information-integrity-forging-pathway-truth-resilience-and-trust
6 https://www.prnewswire.com/news-releases/primer-acquires-yonder-adds-disinformation-analysis-to-ai-portfolio-for-information-operations-301562840.html
7 https://clubmadrid.org/wp-content/uploads/2019/03/Protecting-Information-Integrity-WEB.pdf

de informação que deveria ser protegida, mantida na íntegra até ser recebida. Essa ideia não corresponde ao cenário comunicacional que temos hoje.

Em primeiro lugar, ser capaz de avaliar a integridade da informação pressupõe a capacidade de definir quem produz informação com integridade e como o faz. Vejamos um exemplo simples: um governo lança uma campanha de vacinação para combater a COVID-19. Este processo de vacinação é apoiado por organizações internacionais, pesquisas revisadas por pares e uma série de mecanismos de validação. Há, no entanto, campanhas de desinformação que deturpam o uso da vacina e acabam prejudicando a cobertura vacinal. Seriam necessários mecanismos governamentais de proteção da informação para mantê-la intacta e garantir que chega ao destinatário de uma forma consistente e fiável. Neste caso, a ideia de integridade da informação é justificada pelo interesse público, especialmente no que diz respeito à legitimidade dos mecanismos governamentais para lidar com questões de saúde pública – uma ação que representa uma pequena parte do combate à desinformação.

No entanto, vamos considerar outro exemplo. Um governo decide proibir uma manifestação pública num determinado contexto, com base na interpretação de que criaria riscos para a estabilidade democrática. Já um movimento social defende o direito de manifestação e entende que, na realidade, trata-se de uma tentativa do governo de limitar as críticas ao sistema. A situação é fictícia, mas temos visto experiências como esta em várias partes do mundo, como na França, com as recentes tentativas de protesto em apoio aos palestinos. Neste caso, em que sentido a integridade da informação é preservada? É impossível responder a esta questão simplesmente porque o problema não está na informação em si ou na sua integridade, mas na compreensão de todo o sistema social que envolve este processo e que precisa de ser compreendido e interpretado para além da unidade de informação. Em outras palavras, a integridade da informação não deve ser considerada fora do contexto político e social em que toma forma.

Além disso, a ideia de integridade da informação poderia implicar que o problema reside principalmente em fornecer aos cidadãos informações consideradas verdadeiras, completas e confiáveis. Ou seja, ao proteger a integridade das informações, os cidadãos poderiam exercer plenamente a sua cidadania. No entanto, precisamos considerar que a recepção da informação pode ser problemática – como muitas vezes é – e este é também um problema de comunicação crucial.

Voltemos ao exemplo das vacinas: vamos supor que a informação completa de um governo chegue às pessoas. Mesmo assim, muitas vezes decidem não se vacinar. Não porque a informação completa não tenha chegado até elas, mas porque não faz sentido dentro da visão de mundo que adotaram. Essa formação de cosmovisão é resultado de diversos fatores e fluxos de comunicação, que podem incluir teorias conspiratórias, operações de desinformação e posicionamentos políticos extremos. São processos de comunicação, mas não é apenas a integridade das informações que poderá contê-los.

Um terceiro ponto que precisa ser discutido decorre do fato de que é preciso considerar que grande parte dos problemas do atual cenário comunicacional reside nos fluxos. Os caminhos digitais que a informação percorre para chegar aos cidadãos (especialmente aqueles através de plataformas digitais) têm intermediários que não

existiam no modelo tradicional de comunicação entre remetente e destinatário. Portanto, há uma série de problemas que não residem na informação em si, mas no ambiente por onde ela circula, o que impacta diretamente nos seus efeitos sociais. Para traçar um paralelo, quando falamos de integridade eleitoral, estamos falando de "integridade da eleição" e não de "integridade da votação". Pensamos no sistema, no funcionamento social de uma soma de mecanismos, e não na unidade da decisão do eleitor.

Este foco na unidade e no papel de um transmissor forte e centralizado não é à toa. Ela vem do contexto por trás do uso do termo, que é um contexto de luta contra interferências externas e de proteção de um sistema hegemônico de disseminação de informações.

2. Falta de consenso na tradução para o português

O termo "integridade da informação" foi cunhado em inglês e não existe uma forma única de traduzi-lo para o português. A versão portuguesa do *Policy Brief 8* e os acordos internacionais assinados pelo governo brasileiro falam em "integridade da informação", mas também há menções a "integridade informacional", por exemplo, o que não é exatamente a mesma coisa. Uma ideia menos difundida é a de "integridade do ambiente/espaço/ecossistema comunicacional". O problema básico é que, mais uma vez, estamos importando um conceito externo sem muita discussão. Isto dificulta a escolha de uma tradução – e, portanto, de um significado social – uma vez que não há acumulação do que ela realmente significa.

3. Importações sucessivas de conceitos do Norte Global e uma guerra que não resolve os nossos problemas

Grande parte da discussão sobre o novo cenário comunicacional tem sido baseada em termos estrangeiros que simplesmente não possuem tradução precisa para o português. Foi o caso das "fake news", que, como vários autores salientaram, não são a mesma coisa que "notícias falsas".[8] Foi também o caso da diferença entre "misinformation" e "disinformation", que é impossível de traduzir com precisão em português, fazendo com que muitos coloquem ambos os fenómenos no mesmo balde de "desinformação".

Agora estamos mais uma vez adotando um termo – e uma imagem – estrangeiros simplesmente por tentar encontrar uma tradução linguística, sem pensar no seu real significado. As referências ao termo utilizado até agora mostram que está claramente relacionado com tentativas de proteger o ambiente de comunicações dos EUA contra ameaças externas, especialmente de países não ocidentais. É uma ideia que implica um posicionamento geopolítico que não dá conta dos nossos problemas. É verdade que as ameaças externas ao ambiente de comunicações brasileiro – e ao Sul Global – são reais e precisam ser estudadas e combatidas, mas isso não me parece ser o cerne do problema do ambiente de comunicações que temos hoje.

Lembro-me de uma história contada pelo presidente Luiz Inácio Lula da Silva sobre sua primeira viagem à reunião do G8 em 2003. Ele conta que foi abordado pelo então presidente dos EUA, George W. Bush, que lhe perguntou como o Brasil se envolveria na

8 https://periodicos.ufsc.br/index.php/jornalismo/article/view/1984-6924.2019v16n2p33

Guerra do Iraque. Lula então respondeu: “Presidente Bush, o Iraque não é problema do Brasil. Tenho outra guerra para travar no meu país, que é combater a miséria e a fome dos 50 milhões de brasileiros que vivem na linha da pobreza”.[9] Em 2003, Lula sublinhou que usaria a sua proeminência internacional para se concentrar na luta contra a fome – e não na guerra no Iraque, como gostariam os atores hegemônicos. Em 2024, momento de novo destaque internacional para o Brasil em que os debates sobre informação estão no centro da agenda, qual a proposta do Brasil? Considerando a realidade do Brasil, da América Latina, dos BRICS e do Sul Global, o que é realmente relevante para nós no debate sobre um ambiente de comunicação digital?

A oportunidade de construir uma agenda informacional a partir do Sul

Hoje, o Brasil ocupa posição central na discussão da transformação digital. Assim como em 2014, quando o Brasil aprovou o Marco Civil da Internet, o país tem uma nova oportunidade de trazer as preocupações do Sul Global para o primeiro plano das discussões sobre como construir padrões digitais para nossas sociedades contemporâneas. Isto se deve em grande parte aos árduos esforços de diferentes setores do governo, da sociedade civil e da academia que, de forma muito perspicaz e articulada, viram esta questão como uma prioridade. Essa conquista não é trivial e precisa ser comemorada.

Para aproveitar esta oportunidade, precisamos urgentemente de desenvolver a nossa própria interpretação do problema. Não se trata de paroquialismo, de construir algo brasileiro para o Brasil; pelo contrário, trata-se de aproveitar a possibilidade de desempenhar um papel de liderança nas estruturas internacionais para questionar a ordem estabelecida e mostrar que algo produzido a partir do Sul pode lidar com os problemas globais.

Entendo que o termo “integridade da informação” tenta criar um arcabouço para construir um imaginário do espaço digital que queremos, o que considero mais que necessário. Bem, o que queremos que signifique um espaço de comunicação saudável, justo e democrático? Quais devem ser os parâmetros para acessá-lo?

Quando falamos de comunicação e informação no Brasil e nos países do Sul Global, estamos frequentemente falando de realidades que são amplamente dominadas por organizações noticiosas comerciais, hegemônicas e extremamente concentradas; estamos falando de muitos países onde a comunicação através de aplicações de mensagens é absolutamente central; estamos lidando com democracias jovens e muitas vezes instáveis; estamos nos referindo a sociedades com níveis abismais de desigualdade social, o que impacta a forma como as pessoas consomem informação; estamos falando de países onde não só circula o discurso de ódio, mas serve para reforçar a opressão histórica, como o racismo; estamos lidando com países fortemente impactados por problemas socioambientais; e, com toda a ênfase necessária, estamos falando de países que estão física e imaginativamente distantes das sedes das grandes

9 https://www.facebook.com/Lula/posts/1377452145657115/?paipv=0&eav=AfbFgZ3p6AmJRFKZF3ucN8fFyBHbZoJO-viUEcFjSirLZtuw7qTc_eZWXRD_rbLRzgU&_rdr

empresas tecnológicas, que tratam estes países e os seus cidadãos como menos importantes.

Precisamos contestar a ideia de “integridade da informação” e trazer estes elementos, que são centrais para a maioria da população mundial, para o centro do debate. Nosso desafio é combinar a força da sociedade civil, dos governos e dos seus intelectuais para trazer ao mundo uma visão inovadora, criativa e proativa de que espaço de comunicação democrático queremos.

(*) Nina Santos é diretora do Aláfia Lab, coordenadora geral do *desinformante e investigadora do Instituto Nacional de Ciência e Tecnologia para a Democracia Digital (INCT.DD) e do Centre d'Analyse et de Recherche Interdisciplinaires sur les Médias (Universidade Panthéon-Assas). É membro do Painel Internacional sobre Ambiente de Informação, do comitê diretor da Global Coalition for Tech Justice e do Comitê de Integridade e Transparência Digital em Plataformas de Internet do Tribunal Superior Eleitoral brasileiro. Também é professora da pós-graduação em Estratégias de Comunicação Digital da FGV e do mestrado em comunicação, sistemas de informação e mídias da Université Sorbonne-Nouvelle e autora do livro "Lógicas das mídias sociais: visibilidade e mediação nos protestos brasileiros de 2013" (Palgrave Macmillan, 2022). Nina integrou ainda o grupo de trabalho para a regulamentação da Procuradoria Nacional de Defesa da Democracia (2023), da Procuradoria-Geral da República, e do grupo de trabalho para estratégias de combate ao discurso de ódio e ao extremismo (2023), da o Ministério dos Direitos Humanos e Cidadania.

https://epic.org/safeguarding-democratic-elections-in-the-age-of-ai/

Protegendo as eleições democráticas na era da IA

4 de abril de 2024 | Sophie Nyombi Nantanda, estagiária de primavera da EPIC

À medida que as eleições nacionais se aproximam nos Estados Unidos, as preocupações sobre chamadas de robôs geradas por inteligência artificial (IA) e meios de comunicação manipulados tornaram-se mais urgentes.[1] As eleições de 2024 marcam um momento crucial na história americana, visto que é o primeiro ano de eleições presidenciais em um momento de crescentes avanços no conteúdo gerado por IA. Serão necessárias intervenções legais e tecnológicas para mitigar estas ameaças antes de novembro.

Há razões para acreditar que os EUA não estão adequadamente equipados para resolver este problema. Os funcionários eleitorais estaduais e locais precisam de financiamento e recursos adicionais para enfrentar ameaças que vão da IA à violência pessoal.[2] Sublinhando ainda mais a urgência de abordar estas questões, a confiança no sistema eleitoral está diminuindo, deixando os americanos mais suscetíveis à informação duvidosa e às campanhas de desinformação.[3]

Na era digital, o combate à desinformação gerada pela IA, às chamadas automáticas e aos *deepfakes* exige atenção. Tecnologias como ferramentas de verificação de conteúdo alimentadas por IA e algoritmos de detecção de *deepfakes* (marca d'água digital) podem ajudar a enfrentar essa ameaça, embora com imperfeições. Embora proporcionem um ponto de partida, o papel do envolvimento público na denúncia de desinformação também é essencial. Plataformas dedicadas surgiram para esse fim. Estas plataformas servem como ferramentas vitais para capacitar os indivíduos a sinalizar conteúdos enganosos, promovendo um esforço colaborativo no combate à desinformação e salvaguardando a integridade do nosso território digital. No entanto, para fortalecer verdadeiramente o nosso ecossistema de informação, uma legislação abrangente para a governança da IA pode surgir como a solução mais eficaz. A exploração destas abordagens multifacetadas abre o caminho para a salvaguarda da democracia nas eleições.

Chamadas e Mídia Manipulada

Há alguns meses,[4] a tecnologia de clonagem de voz por IA foi empregada para imitar a voz do presidente Joe Biden com a intenção de dissuadir os eleitores de participarem das primárias presidenciais democratas do estado. Esta desinformação representa uma ameaça à democracia e, em alguns aspectos, o sistema jurídico respondeu rapidamente. Os suspeitos foram notificados;[5] a Comissão Federal de Comunicações esclareceu que as chamadas usando voz alimentada por IA poderiam violar a Lei de Proteção ao Consumidor

Telefônico (TCPA) se feitas sem o consentimento da parte chamada;[6] a Comissão Federal de Comércio propôs responsabilização assumir identidade falsa usando IA;[7] e litigantes privados processaram as partes responsáveis pelas ligações.[8] Mas todas estas medidas surgiram muito depois da mensagem ofensiva já ter sido transmitida. Mais precisa ser feito para evitar esses tipos de chamadas geradas por IA antes que elas aconteçam.

O aumento de chamadas geradas por IA representa um desafio significativo à integridade dos processos eleitorais. Embora os sistemas de chamadas automatizadas tenham sido tradicionalmente utilizados por campanhas políticas para chegar aos eleitores e disseminar mensagens de campanha, a utilização maliciosa de chamadas automáticas nas eleições não é em si nova. As chamadas automáticas já foram utilizadas para disseminar informações falsas, manipular a opinião pública e suprimir a participação eleitoral, minando o processo democrático.

Por exemplo, o Supremo Tribunal do Michigan ouviu recentemente um caso em que dois agentes políticos teriam utilizado 85.000 chamadas automáticas para divulgar informações falsas e dissuadir as pessoas de participarem nas eleições presidenciais de 2020.[9] Mas a IA ameaça intensificar radicalmente o problema das chamadas automáticas eleitorais enganosas. Os avanços na tecnologia de IA permitem a geração automatizada de simulações de voz hiper-realistas, facilitando a criação de chamadas telefônicas e mídias persuasivas e enganosas em grande escala. Para dar uma ideia da escala: o YouTube excluiu recentemente mil vídeos de golpes de IA usando celebridades.[10]

Imagens, vídeos e clipes de áudio gerados por IA também agravam as preocupações sobre a desinformação e a desinformação na publicidade política. A tecnologia *deepfake*, em particular, permite a criação de meios de comunicação sintéticos altamente realistas e difíceis de detectar, como vídeos de candidatos políticos dizendo ou fazendo coisas que nunca disseram ou fizeram. Modelos de IA como Midjourney estão sendo usados para gerar imagens enganosas a serem usadas como desinformação política.[11] Midjourney anunciou algumas medidas proativas para evitar que seus usuários fabriquem imagens falsificadas do presidente Joe Biden e do ex-presidente Donald Trump,[12] mas estas por si só não chegarão perto de resolver o problema da informação duvidosa e da desinformação.

Num caso notável, os apoiadores de Donald Trump criaram e compartilharam imagens falsas de eleitores negros geradas por IA para os encorajar a votar nos republicanos.[13] Este tipo de desinformação representa uma ameaça significativa à integridade das campanhas eleitorais, uma vez que os meios de comunicação falsos gerados pela IA podem influenciar a opinião pública, prejudicar a reputação dos candidatos e minar a confiança nas instituições democráticas.

Deepfakes também podem semear confusão entre o público, confundindo os limites entre o que é autêntico e o que é fabricado. Isto pode fazer com que os indivíduos rotulem involuntariamente informações genuínas como falsas – ou permitir que outros o façam intencionalmente. Os professores Danielle Citron e Bobby Chesney rotularam isso como o dividendo do mentiroso:[14] ou seja, o fenômeno pelo qual a existência de *deepfakes* verossímeis permite e incentiva maus atores e líderes autoritários a rotularem conteúdo verdadeiro como falso.

A circulação de informações duvidosas e desinformação geradas pela IA através de plataformas de redes sociais agrava ainda mais o desafio, como se viu nas recentes eleições presidenciais.[15] Organizações como o Instituto Knight[16] e o Centro de Política Cibernética de Stanford[17] mostraram como narrativas falsas podem espalhar-se rapidamente e sem controle através de sistemas de recomendação algorítmica, amplificando a polarização e minando o discurso democrático. Os mecanismos de verificação de fatos podem ser úteis para combater a propagação de desinformação.[18] Por exemplo, a Meta implementou uma série de ferramentas para lidar com a desinformação das suas plataformas.[19] No entanto, estas ferramentas e mecanismos podem ser sobrecarregados pelo grande volume e sofisticação do conteúdo falso gerado pela IA e podem ter as suas próprias desvantagens (discutidas abaixo). Além disso, a Meta está se preparando para descontinuar o CrowdTangle , uma ferramenta de análise de dados utilizada para identificar informações incorretas no Facebook e no Instagram.[20] A ferramenta será desativada apenas três meses antes das eleições presidenciais dos EUA, restringindo a capacidade de investigadores, jornalistas e outros identificarem tendências perigosas de desinformação num momento chave.

Medidas Legislativas e Tecnológicas

Em resposta a estes desafios, os decisores políticos, as empresas tecnológicas e as organizações da sociedade civil devem trabalhar para desenvolver estratégias abrangentes para enfrentar as ameaças à integridade eleitoral relacionadas com a IA. As legislaturas estaduais estão avaliando e promulgando leis que regulamentariam os *deepfakes* nos processos eleitorais, muitas vezes obtendo apoio bipartidário.[21] A implementação de tais leis ajudaria a combater a propagação da informação duvidosa e da desinformação geradas pela IA, aumentaria a transparência e a responsabilização na publicidade política e promoveria a formação midiática para capacitar os eleitores a discernir os fatos da ficção. A FCC também tem autoridade para reprimir chamadas automáticas que disseminam informações falsas, como fez no passado (embora a violação deva estar relacionada ao TCPA e ao consentimento,[22] e não ao conteúdo das chamadas).[23] Sanções mais rigorosas podem ajudar a Comissão a dissuadir melhor essas práticas enganosas.

Soluções tecnológicas, como ferramentas de verificação de conteúdo alimentadas por IA e algoritmos de detecção de *deepfakes*, podem ajudar a detectar e mitigar o impacto de mídias falsas geradas por IA, mas não podem resolver o problema por si só e podem introduzir problemas adicionais.[24] Por exemplo, embora os detectores de *deepfakes* possam procurar indicadores biométricos distintos num vídeo,[25] como os batimentos cardíacos de um indivíduo ou uma voz produzida por órgãos vocais humanos naturais em vez de sintetizados, a sua eficácia não é garantida.[26]

Os algoritmos atualmente disponíveis lutam para detectar consistentemente *deepfakes* de alta qualidade produzidos por meio de tecnologias avançadas de IA. Além disso, estes detectores também representam riscos potenciais para a privacidade e a equidade. As proteções regulamentares, como a legislação sobre marcas de água da IA,[27] poderiam servir como uma medida eficaz – embora imperfeita – para conter a propagação de informações equivocadas e desinformação. O ceticismo decorre de preocupações em torno da padronização e adoção generalizada de práticas de marcas d'água digitais. Sem um quadro universalmente acordado para a implementação e reconhecimento de marcas d'água, a sua eficácia poderá ser limitada.

No entanto, apesar destes desafios, a adoção de legislação sobre marcas d'água de IA significaria um passo à frente em relação ao atual panorama regulatório. Representaria uma grande melhoria na luta contra a infiormação duvidosa e a desinformação, oferecendo às autoridades uma ferramenta adicional para salvaguardar a integridade dos conteúdos digitais e proteger o discurso público.

Existem também fontes que incentivam o público em geral a denunciar a desinformação para impedir a sua disseminação.[28] Capacitar os indivíduos para sinalizar conteúdos enganosos ajuda as autoridades e as plataformas a resolverem tais casos rapidamente. Juntamente com os mecanismos de denúncia, a promoção do aprendizado midiático dota as pessoas de ferramentas para discernir informações credíveis. Aproveitar a tecnologia para uma comunicação eficiente aumenta a participação e a defesa contra a desinformação, promovendo uma abordagem colaborativa para um ecossistema de informação mais resiliente.

De acordo com a Agência de Cibersegurança e Segurança de Infraestruturas (CISA), as estratégias mais eficazes para enfrentar ameaças generativas melhoradas pela IA nas eleições alinham-se com as melhores práticas de segurança cibernética de longa data, que podem já ter sido implementadas. Estas medidas incluem o reforço na segurança das contas nas redes sociais, a implementação de protocolos robustos de segurança de e-mail para combater ataques de *phishing*, o reforço das defesas do perímetro da rede para detectar e prevenir o acesso não autorizado, a realização regular de auditorias de segurança e avaliações de vulnerabilidade, e a promoção de uma maior colaboração e

partilha de informações entre agências governamentais, funcionários eleitorais e especialistas em segurança cibernética.

Mais medidas de segurança cibernética podem ser encontradas aqui.[29] No entanto, embora estas estratégias possam já ter sido implementadas em vários graus, é importante reconhecer que a segurança cibernética é um campo em evolução e que surgem continuamente novas ameaças. Portanto, embora possam ter sido feitos progressos em determinadas áreas, ainda há muito espaço para melhorias e para a implementação de medidas de cibersegurança mais abrangentes e proativas para salvaguardar eficazmente a integridade eleitoral.

A convergência da IA e das eleições apresenta desafios para as sociedades democráticas. À medida que enfrentamos as ameaças representadas pelas chamadas automáticas geradas pela IA e pelos meios de comunicação manipulados, é imperativo que o governo assuma a liderança no trabalho no sentido de defender os princípios de transparência, integridade e responsabilização nos processos eleitorais.

O Fórum Económico Mundial, no Relatório de Riscos Globais de 2024, previu que a desinformação e a informação enganosa poderiam perturbar significativamente os processos eleitorais em várias economias, incluindo Bangladesh, Índia, Indonésia, México, Paquistão e o Reino Unido, nos próximos dois anos.[30] Preocupações semelhantes ajudaram a motivar a União Europeia a adotar a Lei dos Serviços Digitais, que exige que as plataformas de redes sociais combatam campanhas com motivação política e a propagação de informações falsas.[31] Entretanto, a Lei da IA da União Europeia (aprovada pelo Parlamento Europeu e pelo Conselho da União Europeia e que aguarda uma adoção formal) representa um esforço inovador para regular todo o ciclo de vida do desenvolvimento, implementação e utilização da IA.

Uma das disposições mais críticas da Lei da IA diz respeito à classificação dos sistemas de IA que visam influenciar os processos eleitorais como IA de alto risco, com rigorosas obrigações de transparência, responsabilização e utilização responsável. Os Estados estão promulgando rapidamente legislação destinada a abordar a produção de *deepfakes* gerados por IA, em antecipação às eleições presidenciais de 2024. Mais de 100 projetos de lei foram apresentados ou aprovados em 40 legislaturas estaduais somente neste ano.[32] No entanto, ao adotarem um modelo regulatório abrangente semelhante ao da União Europeia, os Estados Unidos podem fortalecer a sua integridade eleitoral, garantindo justiça e fidelidade às escolhas dos eleitores.

Notas

1 https://epic.org/generative-ai-and-elections-the-approaching-train-wreck/

2 https://cyberscoop.com/deepfakes-dollars-deep-state-fears-election-officials-concerns-2024/

3 https://cyberscoop.com/warner-election-interference-disinformation/

4 https://www.cbsnews.com/news/fake-biden-robocall-new-hampshire-primary/

5 https://ncdoj.gov/wp-content/uploads/2024/02/State-AG-Task-Force-NOTICE-Letter-to-LIFE-CORP-Feb.-2024-1.pdf

6 https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal

7 https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals

8 https://www.washingtonpost.com/politics/2024/03/16/biden-deepfake-robocall-lawsuit-new-hampshire/

9 https://michiganadvance.com/2023/11/09/michigan-supreme-court-hears-2020-election-robocall-misinformation-case/

10 https://www.404media.co/youtube-deletes-1-000-videos-of-celebrity-ai-scam-ads/?utm_source=substack&utm_medium=email

11 https://apnews.com/article/fact-check-trump-nypd-stormy-daniels-539393517762

12 https://apnews.com/article/midjourney-ai-imagegenerator-biden-trump-deepfakes-bc6c254ddb20e36c5e750b4570889ce1

13 https://www.bbc.com/news/world-us-canada-68440150

14 https://scholarship.law.bu.edu/faculty_scholarship/640/

15 https://www.npr.org/2021/10/22/1048543513/facebook-groups-jan-6-

16 https://knightcolumbia.org/content/the-algorithmic-management-of-polarization-and-violence-on-social-media

17 https://www.journaloffreespeechlaw.org/keller.pdf

18 https://www.wired.com/story/fact-checkers-ai-chatgpt-misinformation/

19 https://ai.meta.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/

20 https://arstechnica.com/tech-policy/2024/03/really-bad-timing-meta-is-killing-misinformation-analysis-tool-on-august-14/

21 https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/

22 https://docs.fcc.gov/public/attachments/FCC-21-97A1.pdf

23 https://www.nbcnews.com/politics/elections/robocalls-voters-2020-election-result-5-million-fine-rcna88391

24 https://www.neilsahota.com/technological-solutions-for-deepfake-detection-during-elections/

25 https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html#gs.5kvo5c

26 https://www.neilsahota.com/technological-solutions-for-deepfake-detection-during-elections/

27 https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/

28 https://reportdisinfo.org/

29 https://www.cisa.gov/sites/default/files/2024-01/Consolidated_Risk_in_Focus_Gen_AI_ElectionsV2_508c.pdf

30 https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf?utm_source=substack&utm_medium=email

31 https://www.euractiv.com/section/disinformation/news/eu-citizens-see-ai-and-deepfakes-as-a-threat-for-next-elections-survey-finds

32 https://statescoop.com/deepfakes-presidential-election-ai-2024/#:~:text=New laws penalizing unlabeled AI,are attempting to fight deepfakes

https://www.noemamag.com/we-need-to-rewild-the-Internet/

Precisamos regenerar a Internet

A Internet tornou-se uma monocultura extrativa e frágil. Mas podemos revitalizá-lo usando as lições aprendidas pelos ecologistas.

POR MARIA FARRELL E ROBIN BERJON* – 16 DE ABRIL DE 2024

"A palavra para mundo é floresta" - Ursula K. Le Guin

No final do século XVIII, as autoridades da Prússia e da Saxônia começaram a reorganizar as suas florestas complexas e diversas em filas retas de árvores de uma única espécie. As florestas tinham sido fontes de alimento, pasto, abrigo, medicamentos, camas e muito mais para as pessoas que viviam nelas e ao seu redor, mas para o início do estado moderno, eram simplesmente uma fonte de madeira.

A chamada "silvicultura científica" foi o *growth hacking* daquele século. Tornou a produção de madeira mais fácil de contar, prever e colher, e significou que os proprietários já não dependiam de silvicultores locais qualificados para gerir as florestas. Eles foram substituídos por trabalhadores menos qualificados, seguindo instruções algorítmicas básicas para manter a monocultura arrumada e o sub-bosque vazio.

A informação e o poder de tomada de decisão fluíam agora diretamente para o topo. Décadas mais tarde, quando a primeira colheita foi derrubada, grandes fortunas foram feitas, árvore por árvore padronizada. As florestas derrubadas foram replantadas, na esperança de prolongar o boom. Os leitores do antropólogo político americano da anarquia e da ordem,[1] James C. Scott, sabem o que aconteceu a seguir.[2]

Foi um desastre tão grave que uma nova palavra, *Waldsterben* , ou "morte na floresta", foi cunhada para descrever o resultado. Todas da mesma espécie e idade, as árvores eram derrubadas pelas tempestades, devastadas por insetos e doenças — até mesmo as que sobreviviam eram magras e fracas. As florestas estavam agora tão limpas e nuas que estavam quase mortas. A primeira recompensa magnífica não foi o início de riquezas infinitas, mas uma colheita única de milênios de riqueza do solo construída pela biodiversidade e pela simbiose. A complexidade foi a galinha dos ovos de ouro e ela foi massacrada.

A história da silvicultura científica alemã transmite uma verdade intemporal: quando simplificamos sistemas complexos, destruímo-los e as consequências devastadoras por vezes só são óbvias quando é tarde demais.

Esse impulso de eliminar a confusão que torna a vida resiliente é o que muitos biólogos conservacionistas chamam de "patologia do comando e controle".[3] Hoje, o mesmo impulso para centralizar, controlar e extrair levou a Internet ao mesmo destino que as florestas devastadas.[4]

1 https://theanarchistlibrary.org/library/james-c-scott-two-cheers-for-anarchism
2 https://files.libcom.org/files/Seeing Like a State - James C. Scott.pdf
3 https://www.jstor.org/stable/2386849
4 https://www.nytimes.com/2023/12/21/opinion/Internet-aging-gen-z.html

A década de 2010 da Internet, os seus anos de expansão, podem ter sido a primeira colheita gloriosa que esgotou uma bonança de diversidade. A complexa rede de interações humanas que prosperou com a diversidade tecnológica inicial da Internet está agora encurralada em mecanismos de extração de dados que abrangem todo o mundo, gerando enormes fortunas para poucos.[5]

Nossos espaços online não são ecossistemas, embora as empresas de tecnologia adorem essa palavra.[6] São plantações; ambientes altamente concentrados e controlados, mais próximos da agricultura industrial dos confinamentos de gado ou das granjas em gaiolas que enlouquecem as criaturas presas neles.

Todos nós sabemos disso. Vemos isso cada vez que pegamos nossos telefones. Mas o que a maioria das pessoas não percebeu é como esta concentração atinge profundamente a infraestrutura da Internet – os canais e os protocolos, os cabos e as redes, os motores de busca e os navegadores. Estas estruturas determinam a forma como construímos e utilizamos a Internet, agora e no futuro.

Eles se concentraram em uma série de duopólios quase planetários. Por exemplo, em abril de 2024, os navegadores de Internet do Google e da Apple capturaram quase 85% da participação no mercado mundial,[7] e os dois sistemas operacionais de desktop da Microsoft e da Apple, mais de 80%.[8] O Google administra 84% da pesquisa global e a Microsoft 3%.[9] Pouco mais da metade de todos os telefones vêm da Apple e da Samsung,[10] enquanto mais de 99% dos sistemas operacionais móveis são softwares do Google ou da Apple.[11] Dois provedores de computação em nuvem, Amazon Web Services e Azure da Microsoft, representam mais de 50% do mercado global.[12] Os clientes de e-mail da Apple e do Google gerenciam quase 90% do e-mail global.[13] Google e Cloudflare respondem cerca de 50% das solicitações globais de sistemas de nomes de domínio.

Dois tipos de tudo podem ser suficientes para encher uma arca fictícia e repovoar um mundo em ruínas, mas não podem gerir uma "rede de redes" global e aberta onde todos tenham a mesma oportunidade de inovar e competir. Não é surpresa que o engenheiro da Internet Leslie Daigle tenha denominado a concentração e consolidação da arquitectura técnica da Internet como "'mudanças climáticas' do ecossistema da Internet".[14]

Jardins murados têm raízes profundas

A Internet tornou os gigantes da tecnologia possíveis. Os seus serviços foram dimensionados globalmente, através do seu núcleo aberto e interoperável. Mas, na última década, eles também trabalharam para incluir em seus domínios proprietários os serviços variados, concorrentes e muitas vezes de código aberto ou fornecidos coletivamente nos quais a Internet é construída. Embora isto melhore a sua eficiência operacional, também garante que as condições florescentes do seu próprio surgimento

5 https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/
6 https://crookedtimber.org/2022/12/08/your-platform-is-not-an-ecosystem/
7 https://gs.statcounter.com/browser-market-share/
8 https://www.statista.com/statistics/268237/global-market-share-held-by-operating-systems-since-2009/
9 https://gs.statcounter.com/search-engine-host-market-share
10 https://gs.statcounter.com/vendor-market-share/mobile
11 https://gs.statcounter.com/os-market-share/mobile/worldwide
12 https://www.hava.io/blog/2024-cloud-market-share-analysis-decoding-industry-leaders-and-trends
13 https://www.litmus.com/email-client-market-share
14 https://www.thinkingcat.com/wordpress/wp-content/uploads/2020/08/2019-InvariantsUpdated.pdf

não sejam repetidas por potenciais concorrentes. Para os gigantes da tecnologia, o longo período de evolução da Internet aberta acabou. A Internet deles não é um ecossistema. É um zoológico.

Google, Amazon, Microsoft e Meta estão consolidando o seu controle profundamente na infraestrutura subjacente através de aquisições, integração vertical, construção de redes proprietárias, criação de pontos de estrangulamento e concentração de funções de diferentes camadas técnicas num único silo de controle de cima para baixo. Eles podem dar-se ao luxo de fazê-lo, utilizando a vasta riqueza obtida na sua colheita única de riqueza coletiva e global.

> "Esse impulso de eliminar a confusão que torna a vida resiliente é o que muitos biólogos conservacionistas chamam de 'patologia do comando e controle'".

Tomados em conjunto, o confinamento das infraestruturas e a imposição da monocultura tecnológica bloqueiam o nosso futuro. Os internautas gostam de falar sobre "a pilha", ou a arquitetura em camadas de protocolos, software e hardware, operada por diferentes provedores de serviços que, coletivamente, proporcionam o milagre diário da conexão. É um sistema complicado e dinâmico com um valor básico incorporado à concepção central: as principais funções são mantidas separadas para garantir resiliência, generalidade e criar espaço para inovação.

Inicialmente financiada pelos militares dos EUA e concebida por investigadores acadêmicos para funcionar em tempos de guerra,[15] a Internet evoluiu para funcionar em qualquer lugar, em qualquer condição, operada por qualquer pessoa que quisesse conectar-se. Mas o que era um jogo de Tetris dinâmico e em constante evolução, com "jogadores" e "camadas" distintos, está hoje se consolidando em um sistema de placas tectônicas compactadas que abrange um continente.[16] A infraestrutura não é apenas o que vemos na superfície; são as forças abaixo que criam montanhas e provocam tsunamis. Quem controla a infraestrutura determina o futuro. Se você duvida disso, considere que na Europa ainda usamos estradas e vivemos em vilas e cidades que o Império Romano mapeou há 2.000 anos.

Em 2019, alguns engenheiros de Internet do órgão global de definição de padrões, a Força-Tarefa de Engenharia da Internet, deram o alarme. Daigle, um engenheiro respeitado que já havia presidido seu comitê de supervisão e conselho de arquitetura da Internet, escreveu em um resumo de política que a consolidação significava que as estruturas de rede estavam ossificando em toda a pilha, tornando mais difícil desalojar os titulares e violando um princípio fundamental da Internet: que ela não não crie "favoritos permanentes".[17] A consolidação não apenas elimina a concorrência. Restringe os tipos de relações possíveis entre operadores de diferentes serviços.

Como disse Daigle: "Quanto mais soluções proprietárias são construídas e implantadas em vez de soluções colaborativas baseadas em padrões abertos, menos a Internet sobrevive como plataforma para inovação futura". A consolidação mata a colaboração entre provedores de serviços através da pilha, reorganizando uma série de

15 https://cs.stanford.edu/people/eroberts/courses/soco/projects/distributed-computing/html/history.html
16 https://datatracker.ietf.org/doc/draft-mcfadden-cnsldtn-effects/
17 https://www.thinkingcat.com/wordpress/wp-content/uploads/2020/08/2019-InvariantsUpdated.pdf

relacionamentos diferentes – competitivos, colaborativos – em um único relacionamento predatório.

Desde então, as organizações de desenvolvimento de normas tentaram diversas iniciativas para identificar e abordar a consolidação da infraestrutura, mas estas fracassaram. Atolados em minúcias técnicas, incapazes de se separarem dos interesses dos seus empregadores e dos valores profissionais profundamente arraigados de simplificação e controle,[18] a maioria dos engenheiros da Internet simplesmente não conseguia ver a floresta nas árvores.

De perto, a concentração na Internet parece complexa demais para ser desvendada; de longe, parece muito difícil de lidar. Mas e se pensássemos na Internet não como um "hiperobjeto" do Juízo Final,[19] mas como um ecossistema danificado e em dificuldades, enfrentando a destruição? E se olhássemos para isso não com horror impotente face à invasão sobrenatural dos seus atuais controladores, mas com compaixão, construtividade e esperança?

Os tecnólogos são excelentes em soluções incrementais, mas para regenerar habitats inteiros, precisamos de aprender com os ecologistas que têm uma visão de todo o sistema. Os ecologistas também sabem como continuar quando os outros primeiro os ignoram e depois dizem que é tarde demais, como se mobilizar e trabalhar coletivamente, e como construir bolsões de diversidade e resiliência que irão durar mais que eles, criando possibilidades para um futuro abundante que eles podem imaginar, mas nunca controlar. Não precisamos reparar a infraestrutura da Internet. Precisamos regenerá-la.

O que é *rewilding*?

A regeneração (*rewilding*) "visa restaurar ecossistemas saudáveis através da criação de espaços selvagens e biodiversos", de acordo com a União Internacional para a Conservação da Natureza.[20] Mais ambiciosa e tolerante ao risco do que a conservação tradicional, visa ecossistemas inteiros para abrir espaço para redes alimentares complexas e para o surgimento de relações interespécies inesperadas. Está menos interessada em salvar espécies específicas ameaçadas. As espécies individuais são apenas componentes do ecossistema, e focar nos componentes perde de vista o todo. Os ecossistemas florescem através de múltiplos pontos de contacto entre os seus vários elementos, tal como as redes de computadores. E, tal como nas redes de computadores, as interações dos ecossistemas são multifacetadas e produtivas.

Rewilding tem muito a oferecer às pessoas que se preocupam com a Internet. Como Paul Jepson e Cain Blythe escreveram no seu livro *Rewilding: The Radical New Science of Ecological Recovery*,[21] a regeneração presta atenção "às propriedades emergentes das interações entre 'coisas' nos ecossistemas… uma mudança do pensamento linear para o pensamento sistêmico".

É uma abordagem fundamentalmente alegre e profissional para o que pode parecer insolúvel. Não microgerencia. Cria espaço para "processos ecológicos promoverem ecossistemas complexos e auto-organizados". A regeneração coloca em prática o que

18 https://archive.org/details/whatengineerskno0000vinc
19 https://books.google.com/books/about/Hyperobjects.html?id=qu5zDwAAQBAJ&source=kp_book_description
20 https://www.iucn.org/resources/issues-brief/benefits-and-risks-rewilding
21 https://mitpress.mit.edu/9780262046763/rewilding/

todo bom gestor sabe: contrate as melhores pessoas que puder, forneça o que elas precisam para prosperar e depois saia do caminho. É o oposto de comando e controle.

> "A complexa rede de interações humanas que prosperou com a diversidade tecnológica inicial da Internet está agora encurralada em mecanismos de extração de dados que abrangem todo o mundo, gerando enormes fortunas para poucos."

Regenerar a Internet é mais do que uma metáfora. É uma estrutura e um plano. Dá-nos novos olhos para o perverso problema da extração e do controle, e novos meios e aliados para o resolvê-lo. Reconhece que acabar com os monopólios da Internet não é apenas um problema intelectual. É emocional. Responde a questões como estas: como podemos continuar quando os monopólios têm mais dinheiro e poder? Como agimos coletivamente quando eles subornam os nossos espaços comunitários, financiamento e redes? E como comunicamos aos nossos aliados como será a solução?

Rewilding é uma visão positiva para as redes em que queremos viver e uma história partilhada sobre como chegaremos lá. Ela enxerta uma nova árvore no velho e cansado estoque da tecnologia.

O que a ecologia sabe

A ecologia sabe muito sobre sistemas complexos dos quais os tecnólogos podem se beneficiar. Primeiro, sabe que as mudanças de referenciais são reais.[22]

Se você nasceu por volta da década de 1970, provavelmente se lembra de muito mais insetos mortos no para-brisa do carro de seus pais do que no seu. As populações globais de insetos terrestres estão diminuindo ao ritmo de 9% por década.[23] Se você é um *geek*, provavelmente programou seu próprio computador para fazer jogos básicos. Você certamente se lembra de uma Web com mais para ler do que os mesmos cinco sites.[24] Você pode até ter escrito seu próprio blog.

Mas muitas pessoas nascidas depois de 2000 provavelmente pensam que um mundo com poucos insetos, pouco ruído ambiente de cantos de pássaros, onde você usa regularmente apenas algumas mídias sociais e aplicativos de mensagens (em vez de uma *Web inteira*) é normal. Como escreveram Jepson e Blythe, a mudança de referenciais é "onde cada geração assume que a natureza que experimentaram na sua juventude é normal e involuntariamente aceita o declínio e os danos das gerações anteriores". O dano já está presente. Até parece natural.

A ecologia sabe que a mudança de referenciais diminui a urgência coletiva e aprofunda as divisões geracionais. As pessoas que se preocupam com a monocultura e o controle da Internet costumam ser etiquetadas como nostálgicas que remontam a uma era pioneira. É terrivelmente difícil regenerar uma infraestrutura aberta e competitiva para as gerações mais jovens, que foram criadas para assumir que duas ou três plataformas, duas lojas de aplicativos, dois sistemas operacionais, dois navegadores, uma nuvem/megaloja e um único mecanismo de busca para o mundo compreende *a Internet* . Se a Internet para você é o enorme silo de arranha-céus em que você mora e a

22 https://esajournals.onlinelibrary.wiley.com/doi/10.1002/fee.1794
23 https://www.science.org/doi/10.1126/science.aax9931
24 https://theconversation.com/we-spent-six-years-scouring-billions-of-links-and-found-the-web-is-both-expanding-and-shrinking-159215

única coisa que você pode ver do lado de fora é o outro enorme silo de arranha-céus, então como você pode imaginar outra coisa?

O poder digital concentrado produz os mesmos sintomas que o comando e o controle produzem nos ecossistemas biológicos; angústia aguda pontuada por colapsos repentinos quando os pontos de inflexão são alcançados. Que escala é necessária para que o *rewilding* tenha sucesso? Uma coisa é reintroduzir lobos nos 8.992 km2 de Yellowstone,[25] e outra bem diferente é isolar cerca de 52 km2 de um *polder* (terra recuperada de um corpo de água) conhecido como Oostvaardersplassen, perto de Amsterdã. O grande e diversificado Yellowstone é provavelmente complexo o suficiente para se adaptar às mudanças, mas Oostvaardersplassen tem enfrentado dificuldades .[26]

> "Nossos espaços online não são ecossistemas, embora as empresas de tecnologia adorem essa palavra. São plantações; ambientes altamente concentrados e controlados… que enlouquecem as criaturas presas neles."

Na década de 1980, o governo holandês tentou regenerar uma seção do Oostvaardersplassen coberto de vegetação. Um ecologista independente do governo, Frans Vera, disse que os juncos e os arbustos dominariam a menos que herbívoros agora extintos os pastassem. No lugar dos antigos auroques, a agência estatal de gestão florestal introduziu o gado alemão Heck, conhecido pelo mal humor, e no lugar de um extinto pônei das estepes, uma raça semi-selvagem polonesa.[27]

Cerca de 30 anos depois, sem predadores naturais, e depois que os planos para um corredor de vida selvagem para outra reserva fracasaram, havia muito mais animais do que a limitada vegetação de inverno conseguia sustentar. As pessoas ficaram horrorizadas com vacas e pôneis famintos e, a partir de 2018, as agências governamentais instituíram verificações e abates de bem-estar animal.[28]

Apenas voltar o relógio era insuficiente. O segmento de Oostvaardersplassen era muito pequeno e desconectado para ser reestruturado. Como os animais não tinham para onde ir, o pastoreio excessivo e o colapso eram inevitáveis, uma lição embaraçosa mas necessária. A regeneração é um trabalho em andamento. Não se trata de tentar reverter os ecossistemas a um Éden mítico. Em vez disso, os regeneradores procuram reconstruir a resiliência, restaurando processos naturais autônomos e deixando-os operar em escala para gerar complexidade. Mas a regeneração, em si uma intervenção humana, pode precisar de várias tentativas para acertar.

Por mais que tentemos, a Internet não está retornando às interfaces antigas e comuns, como FTP e Gopher, ou às organizações que operam seus próprios servidores de e-mail novamente, em vez de soluções prontas para uso, como o G-Suite. Mas parte do que precisamos já está aqui, principalmente na Web. Observe o ressurgimento de feeds RSS, boletins informativos por e-mail e blogs, à medida que descobrimos (mais uma vez) que depender de um aplicativo para hospedar conversas globais cria um ponto

25 https://www.yellowstonepark.com/things-to-do/wildlife/wolf-reintroduction-changes-ecosystem/
26 https://www.theguardian.com/environment/2022/jun/21/pioneering-dutch-rewilding-project-oostvaardersplassen-works-to-rebuild-controversial-reputation-aoe
27 https://www.popsci.com/nazi-bred-cows-are-too-ferocious-farm/
28 https://www.theguardian.com/environment/2018/apr/27/dutch-rewilding-experiment-backfires-as-thousands-of-animals-starve

único de falha e controle. Novos sistemas estão crescendo, como o Fediverse com suas ilhas federadas,[29] ou o Bluesky com escolha algorítmica[30] e moderação combinável.[31]

Não sabemos o que o futuro reserva. Nosso trabalho é manter abertas o máximo de oportunidades que pudermos, confiando que aqueles que vierem depois as aproveitarão. Em vez de definir testes de pureza para qual tipo de Internet é mais parecido com o original, podemos testar as alterações em relação aos valores do design original. Será que os novos padrões protegem a "generalidade" da rede, ou seja, a sua capacidade de suportar múltiplas utilizações, ou a funcionalidade é limitada para otimizar a eficiência das maiores empresas tecnológicas?

Já em 1985, os ecologistas especialistas em plantas Steward T.A. Pickett e Peter S. White escreveram em *The Ecology of Natural Disturbance and Patch Dynamics*, que um "paradoxo essencial da conservação da natureza selvagem é que procuramos preservar o que deve mudar".[32] Alguns engenheiros de rede sabem disso. David Clark, professor do Instituto de Tecnologia de Massachusetts que trabalhou em alguns dos primeiros protocolos da Internet, escreveu um livro inteiro sobre outras arquiteturas de rede que poderiam ter sido construídas se valores diferentes, como segurança ou gerenciamento centralizado, tivessem sido priorizados pelos criadores da Internet.[33]

Mas a nossa Internet decolou porque foi projetada como uma rede de uso geral, construída para conectar qualquer pessoa. Nossa Internet foi construída para ser complexa e imbatível, para fazer coisas que ainda não podemos imaginar. Quando entrevistamos Clark, ele nos disse que "'complexo' implica um sistema no qual você tem um comportamento emergente, um sistema no qual você não pode modelar os resultados. Suas intuições podem estar erradas. Mas um sistema demasiado simples significa oportunidades perdidas." Tudo o que fazemos coletivamente e que vale a pena é complexo e, portanto, um pouco mais confuso. É nas brechas que novas pessoas e ideias entram.

A infraestrutura da Internet é um ecossistema degradado, mas também é um ambiente construído, como uma cidade. A sua imprevisibilidade torna-o produtivo, valioso e profundamente humano. Em 1961, Jane Jacobs, uma ativista americano-canadense e autora de *The Death and Life of Great American Cities*, argumentou que os bairros de uso misto eram mais seguros, mais felizes,[34] mais prósperos e mais habitáveis[35] do que os projetos estéreis e altamente controladores de planejadores urbanos como Robert Moses, de Nova York.[36]

> "Como um ambiente construído de cima para baixo, a Internet tornou-se algo que é feito para nós, e não algo que refazemos coletivamente todos os dias."

Tal como as torres dominadas pelo crime, ao estilo de Corbusier, onde Moses amontoou as pessoas quando demoliu bairros de uso misto e construiu estradas

29 https://www.theverge.com/24063290/fediverse-explained-activitypub-social-media-open-protocol
30 https://bsky.social/about/blog/3-30-2023-algorithmic-choice
31 https://bsky.social/about/blog/4-13-2023-moderation
32 https://www.sciencedirect.com/book/9780125545204/the-ecology-of-natural-disturbance-and-patch-dynamics#book-info
33 https://mitpress.mit.edu/9780262547703/designing-an-internet/
34 https://www.penguinrandomhouse.com/books/86058/the-death-and-life-of-great-american-cities-by-jane-jacobs/
35 https://savingplaces.org/stories/a-tale-of-two-planners-jane-jacobs-and-robert-moses
36 https://www.bloomberg.com/news/articles/2017-07-09/robert-moses-and-his-racist-parkway-explained

através deles, a Internet concentrada de cima para baixo de hoje é, para muitos, um lugar desagradável e prejudicial. Os seus proprietários são difíceis de remover e os seus interesses não se alinham com os nossos.

Como escreveu Jacobs: "Como em todas as utopias, o direito de ter planos de qualquer importância pertencia apenas aos planejadores responsáveis". Como um ambiente construído de cima para baixo, a Internet tornou-se algo que é feito para nós, e não algo que refazemos coletivamente todos os dias.

Os ecossistemas perduram porque as espécies servem como freios e contrapesos entre si. Eles têm diferentes modos de interação, não apenas extração, mas mutualismo, comensalismo, competição e predação. Em ecossistemas prósperos, os predadores estão sujeitos a limites.[37] Eles são apenas uma parte de uma teia complexa que transmite calorias, e não uma passagem só de ida para o fim da evolução.

Os ecologistas sabem que diversidade é resiliência.

Em 18 de julho de 2001, 11 vagões de um trem de carga de 60 vagões descarrilaram no túnel Howard Street sob o bairro de Mid-Town Belvedere, ao norte do centro de Baltimore.[38] Em poucos minutos, um tanque contendo um produto químico altamente inflamável foi perfurado. O produto químico que escapou incendiou-se e logo os vagões adjacentes pegaram fogo em um incêndio que levou cerca de cinco dias para ser apagado. O desastre se multiplicou e se espalhou. As grossas paredes do túnel de tijolos funcionavam como um forno e as temperaturas subiram para quase 1.000 graus Celsius.[39] Uma adutora de mais de um metro de diâmetro acima dos túneis estourou, inundando o túnel com milhões de galões em poucas horas. Mas isso só esfriou um pouco. Três semanas depois, uma explosão ligada ao combustível químico[40] explodiu tampas de bueiros localizadas a até três quilômetros de distância.[41]

A WorldCom, então a segunda maior empresa de telefonia de longa distância dos EUA, tinha cabos de fibra óptica no túnel transportando grandes volumes de tráfego telefônico e de Internet. Contudo, de acordo com Clark, professor do MIT, o planejamento de resiliência da WorldCom fez com que o tráfego fosse distribuído por diferentes redes de fibra em antecipação a este tipo de evento.

No papel, a WorldCom tinha redundância de rede. Mas quase imediatamente, o tráfego da Internet nos EUA desacelerou e as linhas telefônicas transatlânticas e da Costa Leste da WorldCom caíram.[42] A estreita topografia física da região concentrou todas essas diferentes redes de fibra em um único ponto de estrangulamento, o túnel Howard Street. A resiliência da WorldCom foi, literalmente, incinerada. Tinha redundância tecnológica, mas não diversidade. Às vezes não percebemos a concentração até que seja tarde demais.

Clark conta a história do incêndio no túnel Howard Street para mostrar que os gargalos nem sempre são óbvios, especialmente no nível operacional, e sistemas enormes que parecem seguros, devido ao seu tamanho e recursos, podem desmoronar inesperadamente.

---

37 https://www.livingwithwolves.org/wolf-science-weekly/
38 https://www.ntsb.gov/investigations/AccidentReports/Reports/RAB0408.pdf
39 https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=900095
40 https://mde.maryland.gov/programs/pressroom/Pages/436.aspx
41 https://www.bullsheet.com/bullsheet.com/tunnelfire.html
42 https://www.nytimes.com/2001/07/20/us/fire-in-baltimore-snarls-internet-traffic-too.html

Na Internet de hoje, grande parte do tráfego passa pelas redes privadas das empresas de tecnologia, por exemplo, os próprios cabos submarinos do Google e da Meta.[43] Grande parte do tráfego da Internet é servido por algumas redes dominantes de distribuição de conteúdo, como Cloudflare e Akamai, que administram suas próprias redes de servidores proxy e datacenters. Da mesma forma, esse tráfego passa por um número cada vez menor de resolvedores de sistemas de nomes de domínio (DNS), que funcionam como listas telefônicas para a Internet, vinculando nomes de sites a seus endereços numéricos.

Tudo isso melhora a velocidade e a eficiência da rede, mas cria gargalos novos e não óbvios, como o túnel Howard Street. Os provedores de serviços centralizados dizem que têm mais recursos e são mais qualificados em ataques e falhas, mas também são alvos grandes e atraentes para invasores e possíveis pontos únicos de falha do sistema.[44]

Em 21 de outubro de 2016, dezenas de grandes sites dos EUA pararam de funcionar repentinamente. Os nomes de domínio pertencentes ao Airbnb, Amazon, PayPal, CNN e The New York Times simplesmente não foram resolvidos. Todos eram clientes do provedor comercial de serviços DNS, Dyn, que foi atingido por um ataque cibernético. Os hackers infectaram dezenas de milhares de dispositivos conectados à Internet com software malicioso, criando uma rede de dispositivos sequestrados, ou uma botnet, que usaram para bombardear Dyn com consultas até que ela entrasse em colapso.[45] As maiores marcas de Internet dos Estados Unidos foram derrubadas por nada mais do que uma rede de babás eletrônicas, webcams de segurança e outros dispositivos de consumo.[46] Embora todos provavelmente tivessem planejamento de resiliência e redundâncias, eles falharam porque um único ponto de estrangulamento – em uma camada crucial da infraestrutura – falhou.

> “Acidentes, incêndios e inundações podem ser simplesmente entropia em ação, mas infraestruturas sistemicamente concentradas e arriscadas são escolhas manifestadas – e podemos fazer escolhas melhores.”

Interrupções generalizadas devido a pontos de estrangulamento centralizados tornaram-se tão comuns que os investidores até as utilizam para identificar oportunidades. Quando uma falha do provedor de nuvem Fastly tirou sites de alta visibilidade do ar em 2021, o preço de suas ações disparou.[47] Os investidores ficaram encantados com as manchetes que os informavam sobre um obscuro fornecedor de serviços técnicos com aparente bloqueio num serviço essencial. Para os investidores, esta falha infraestrutural crítica não parece uma fragilidade, mas sim uma oportunidade de lucro.

O resultado da estreiteza infraestrutural é uma fragilidade embutida que só notamos após um colapso. Mas a monocultura também é altamente visível em nossas ferramentas de busca e navegação. Pesquisa, navegação e mídias sociais são a forma como encontramos e compartilhamos conhecimento e como nos comunicamos.

---

43 https://www.wsj.com/articles/google-amazon-meta-and-microsoft-weave-a-fiber-optic-web-of-power-11642222824
44 https://www.tandfonline.com/doi/full/10.1080/23738871.2020.1728355
45 https://coverlink.com/case-study/mirai-ddos-attack-on-dyn
46 https://www.washingtonpost.com/opinions/the-day-of-the-zombie-baby-monitors-when-hackers-weaponized-the-internet-of-things/2016/10/25/167fdf42-9a1b-11e6-b3c9-f662adaa0048_story.html
47 https://www.ft.com/content/4c68df91-98d1-4942-87cf-f734ea2cdd73

Constituem uma infraestrutura epistémica e democrática global e crítica, controlada por apenas algumas empresas norte-americanas. Colisões, incêndios e inundações podem ser simplesmente entropia em acção, mas infraestruturas sistemicamente concentradas e arriscadas são escolhas manifestadas – e podemos fazer escolhas melhores.

A aparência de uma Internet renovada

Uma Internet renovada terá muito mais opções de serviços. Alguns serviços, como busca e mídia social, serão desmembrados, como aconteceu com a AT&T.[48] Em vez de as empresas tecnológicas extrairem e venderem dados pessoais, diferentes modelos de pagamento financiarão a infraestrutura de que necessitamos. Neste momento, há pouca provisão explícita para bens públicos como protocolos de Internet e navegadores, essenciais para fazer a Internet funcionar. As maiores empresas de tecnologia os subsidiam e influenciam profundamente.

Parte do *rewilding* significa tomar de volta o que foi puxado para o acervo das *Big Techs* e pagar pelos custos reais da conectividade. Continuaremos a pagar diretamente algumas coisas, como a conectividade básica, e outras, como os navegadores, apoiaremos indiretamente, mas de forma transparente, conforme descrito abaixo. A Internet redesenhada terá inúmeras maneiras de se conectar e se relacionar. Não haverá apenas um ou dois números para ligar se os líderes de um golpe político decidirem desligar a Internet a meio da noite, como aconteceu em lugares como o Egito[49] e Myanmar.[50] Nenhuma entidade estará permanentemente no topo. Uma Internet renovada será um lugar mais interessante, utilizável, estável e agradável para se estar.

Através de extensa pesquisa, a economista ganhadora do Nobel Elinor Ostrom descobriu que "quando os indivíduos estão bem informados sobre o problema que enfrentam e sobre quem mais está envolvido, e podem construir ambientes onde a confiança e a reciprocidade possam emergir, crescer e ser sustentadas ao longo do tempo, custosos e ações positivas são frequentemente tomadas sem esperar que uma autoridade externa imponha regras, monitore o cumprimento e avalie penalidades."[51] Ostrom encontrou pessoas que se organizavam espontaneamente para gerir os recursos naturais – desde a cooperação entre empresas de água na Califórnia até pescadores de lagosta do Maine que se organizavam para evitar a sobrepesca.[52]

A auto-organização também existe como parte de uma função fundamental da Internet: a coordenação do tráfego. Os pontos de troca de Internet (PTTs) são um exemplo de gerenciamento de recursos comuns, onde os provedores de serviços de Internet (ISPs) concordam coletivamente em transportar os dados uns dos outros por baixo ou nenhum custo. Operadores de rede de todos os tipos – empresas de telecomunicações, grandes empresas de tecnologia, universidades, governos e emissoras – precisam enviar grandes quantidades de dados através de redes de outros ISPs para que cheguem ao seu destino.

48 https://www.eff.org/deeplinks/2021/02/what-att-breakup-teaches-us-about-big-tech-breakup
49 https://www.nytimes.com/2011/01/29/technology/internet/29cutoff.html
50 https://www.reuters.com/graphics/MYANMAR-POLITICS/INTERNET-RESTRICTION/rlgpdbreepo/
51 https://www.sciencedirect.com/science/article/abs/pii/S0959378010000634
52 https://www.yesmagazine.org/economy/2021/08/11/the-commons-lobster-maine-elinor-ostrom

Se conseguissem isso separadamente por meio de contratos individuais, gastariam muito mais tempo e dinheiro. Em vez disso, muitas vezes formam PTTs, normalmente como associações independentes e sem fins lucrativos. Para além de administrarem o tráfego, os PTTs formaram, em muitos países — e especialmente os em desenvolvimento — a espinha dorsal de uma comunidade técnica florescente que impulsiona ainda mais o desenvolvimento econômico.

Tanto entre as pessoas quanto na Internet, as conexões são generativas. Desde normas técnicas à gestão de recursos comuns e até mesmo a redes de banda larga mais localizadas, conhecidas como "altnets", a reconstituição da Internet já dispõe de uma profunda caixa de ferramentas de ação colectiva pronta a ser implementada.

O novo impulso para antitruste e concorrência

A lista de infraestruturas a diversificar é longa. Além de pipes e protocolos, existem sistemas operacionais, navegadores, mecanismos de busca, Sistema de Nomes de Domínio, mídias sociais, publicidade, provedores de nuvem, lojas de aplicativos, empresas de IA e muito mais. E essas tecnologias também estão interligadas.

Mas mostrar o que pode ser feito numa área cria oportunidades em outras. Primeiro, vamos começar com a regulamentação.

Nem sempre é necessária uma grande ideia nova, como a regeneração, para enquadrar e motivar grandes mudanças estruturais. Às vezes, reviver uma ideia antiga é suficiente. A "Ordem Executiva sobre a Promoção da Concorrência na Economia Americana" de 2021 do presidente Biden reviveu o escopo original,[53] pró-trabalhador e anti-trustes e a urgência do ativista jurídico do início do século 20 e juiz da Suprema Corte Louis D. Brandeis, juntamente com regras e enquadramentos que datam de antes do New Deal da década de 1930.

> "Regenerar um ambiente já construído não é apenas sentar e ver que coisa tenra e viva pode abrir caminho através do concreto. É demolir as estruturas que bloqueiam a luz para todos que não são ricos o suficiente para viver no último andar."

A lei antitruste dos EUA foi criada para quebrar o poder dos oligarcas do petróleo, do aço e dos caminhos-de-ferro que ameaçavam a jovem democracia dos Estados Unidos.[54] Deu aos trabalhadores proteções básicas e considerou a igualdade de oportunidades econômicas como essencial para a liberdade. Esta visão da concorrência como essencial foi eliminada pelas políticas econômicas da Escola de Chicago na década de 1970 e pelas decisões judiciais dos juízes da era Reagan ao longo das décadas.[55] Eles acreditavam que a intervenção só deveria ser permitida quando o poder do monopólio provocasse o aumento dos preços ao consumidor.[56] A monocultura intelectual desse limiar de danos ao consumidor espalhou-se desde então por todo o mundo.

É por isso que os governos simplesmente ficaram de lado enquanto as empresas de tecnologia do século XXI migravam para o oligopólio. Se o único critério de ação de um

53 https://www.whitehouse.gov/briefing-room/presidential-actions/2021/07/09/executive-order-on-promoting-competition-in-the-american-economy/

54 https://www.motherjones.com/politics/2023/11/how-gilded-age-lawmakers-saved-america-from-plutocracy/

55 https://bfi.uchicago.edu/wp-content/uploads/2022/08/BFI_WP_2022-104.pdf

56 https://lawreview.uchicago.edu/print-archive/chicago-school-and-forgotten-political-dimension-antitrust-law

regulador é garantir que os consumidores não paguem um dentavo a mais, então os serviços gratuitos ou subsidiados por dados das plataformas tecnológicas nem sequer são registados. (É claro que os consumidores pagam de outras formas, uma vez que estes gigantes da tecnologia exploram as suas informações pessoais para obter lucro.) Esta abordagem laissez-faire permitiu que as maiores empresas sufocassem a concorrência, adquirindo os seus concorrentes e integrando verticalmente os prestadores de serviços, criando os problemas que temos hoje.[57]

Os reguladores e responsáveis pela aplicação da lei em Washington e Bruxelas dizem agora que aprenderam essa lição e não permitirão que o domínio da IA aconteça como aconteceu com a concentração na Internet. A presidente da Comissão Federal de Comércio, Lina Khan, e o responsável pela aplicação da lei antitruste do Departamento de Justiça dos EUA, Jonathan Kanter,[58] estão identificando pontos de estrangulamento[59] na "pilha" de IA – concentração no controle de chips de processamento, conjuntos de dados, capacidade de computação, inovação de algoritmos, plataformas de distribuição e interfaces de usuário[60] – e analisando para ver se afetam a concorrência sistêmica. Esta é uma notícia potencialmente boa para as pessoas que desejam evitar que o atual domínio dos gigantes da tecnologia se estenda ao nosso futuro da IA.

Na assinatura da ordem executiva sobre a concorrência em 2021, o Presidente Biden disse: "Capitalismo sem competição não é capitalismo; é exploração."[61] Os responsáveis pela aplicação da lei de Biden estão mudando os tipos de casos que assumem e alargando as teorias jurídicas aplicáveis sobre os danos que causam aos juízes. Em vez do enfoque tradicionalmente estreito nos preços no consumidor, os casos atuais argumentam que os danos econômicos perpetrados pelas empresas dominantes incluem os sofridos pelos seus trabalhadores, pelas pequenas empresas e pelo mercado como um todo.

Khan e Kanter abandonaram modelos estreitos e obscuros de comportamento de mercado em favor de experiências do mundo real de profissionais de saúde, agricultores e escritores. Eles *entendem* que o bloqueio de oportunidades econômicas alimenta as ações da extrema direita. Eles tornaram a aplicação das leis antitruste e das políticas de proteção à concorrência explicitamente em coerção versus escolha, poder versus democracia. Kanter disse numa conferência recente em Bruxelas que "a concentração excessiva de poder é uma ameaça… não se trata apenas de preços ou produção, mas se trata de liberdade e oportunidades".[62]

As autoridades em Washington e Bruxelas estão começando a impedir preventivamente que as empresas tecnológicas utilizem seu domínio em um campo para apropriar-se de outro. Após análise da FTC dos EUA e da Comissão Europeia, a Amazon abandonou recentemente o seu plano de adquirir o fabricante de eletrodomésticos, iRobot.[63] Os reguladores de ambos os lados do Atlântico também

57 https://www.bloomberg.com/news/articles/2020-07-27/big-tech-goes-on-shopping-spree-brushing-off-antitrust-scrutiny
58 https://www.nytimes.com/2024/03/22/technology/jonathan-kanter-apple-antitrust.html
59 https://www.theverge.com/2023/3/28/23660101/ai-competition-ftc-doj-lina-khan-jonathan-kanter-antitrust-summit
60 https://www.justice.gov/opa/speech/assistant-attorney-general-jonathan-kanter-delivers-remarks-22nd-international
61 https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/07/09/remarks-by-president-biden-at-signing-of-an-executive-order-promoting-competition-in-the-american-economy/
62 https://bruxconference2024.clevercast.com/webcast/w-qodbzp/
63 https://www.bbc.com/news/business-68131819

tomaram medidas para impedir a Apple de usar o domínio da plataforma iPhone para restringir a concorrência nas lojas de aplicativos e dominar os mercados futuros, por exemplo, incentivando o uso do CarPlay nas montadoras e limitando o acesso ao seu sistema de carteira digital *tap-to-pay* no setor de serviços financeiros.

Ainda assim, até agora, as ações de fiscalização concentraram-se nas partes altamente visíveis e voltadas para o consumidor da Internet exploradora e proprietária dos gigantes da tecnologia.[64] As poucas e estreitas medidas da ordem executiva de 2021 que visam reduzir os monopólios baseados em infraestruturas, apenas evitam abusos *futuros* , como a apropriação do espectro radioelétrico, e não aqueles já bloqueados.[65] É evidente que a melhor maneira de lidar com os monopólios é, em primeiro lugar, impedir que eles aconteçam. Mas, a menos que os reguladores e os responsáveis pela aplicação da lei erradiquem agora o domínio existente destes gigantes, viveremos no atual monopólio de infraestruturas durante décadas, talvez até um século.

Mesmo os reguladores activistas têm evitado aplicar as soluções mais duras para a concentração em mercados há muito consolidados, tais como requisitos de não discriminação, interoperabilidade funcional e separações estruturais, ou seja, o desmembramento de empresas. E declarar que os monopólios de busca e de redes sociais[66] são, na verdade, serviços públicos[67] – e forçá-los a agir como operadores comuns abertos a todos – ainda é demasiado extremo para a maioria.

Mas regenerar um ambiente construído não é apenas sentar e ver que coisa tenra e viva pode abrir caminho através do concreto. É destruir as estruturas que bloqueiam a luz para todos que não são ricos o suficiente para viver no último andar.

> “Os ecologistas reorientaram o seu campo como uma 'disciplina de crise', um campo de estudo em que não se trata apenas de aprender sobre as coisas, mas de salvá-las. Nós, tecnólogos, precisamos fazer o mesmo.”

Quando o escritor e ativista Cory Doctorow escreveu sobre como nos libertar das garras da Big Tech,[68] ele disse que, embora o desmembramento de grandes empresas provavelmente leve décadas, fornecer uma interoperabilidade forte e obrigatória abriria espaço inovador e retardaria o fluxo de dinheiro para as maiores empresas – dinheiro que de outra forma usariam para aprofundar os seus fossos.

Doctorow descreve “comcom”, ou compatibilidade competitiva, como uma espécie de “interoperabilidade de guerrilha, alcançada através de engenharia reversa, bots, descartes e outras táticas sem permissão”. Antes que um emaranhado de leis invasivas surgisse para estrangulá-lo, o comcom era o meio pelo qual as pessoas descobriam como consertar carros e tratores ou reescrever software. A comcom impulsiona o comportamento de tentar todas as táticas até que uma funcione que você vê em um ecossistema em desenvolvimento.

Em um ecossistema, a diversidade de espécies é outra forma de dizer “diversidade de táticas”, uma vez que cada nova tática bem sucedida cria um novo nicho a ocupar. Quer se trate de um polvo se camuflando como uma cobra marinha, de um cuco

64 https://www.nytimes.com/2024/03/04/technology/europe-apple-meta-google-microsoft.html
65 https://www.vice.com/en/article/mbjpjb/a-short-history-of-wireless-spectrum-the-most-complicated-puzzle-youve-ever-seen
66 https://www.bnnbloomberg.ca/big-tech-s-natural-monopoly-tough-to-self-regulate-malone-says-1.1679411
67 https://www.wired.com/story/no-facebook-google-not-public-utilities/
68 https://www.noemamag.com/freeing-ourselves-from-the-clutches-of-big-tech/

contrabandeando seus filhotes para o ninho de outro pássaro, de orquídeas produzindo flores que se parecem com uma abelha fêmea ou de parasitas influenciando hospedeiros roedores a correrem riscos fatais, cada micronicho evolutivo é criado por uma tática de sucesso. Comcom é simplesmente diversidade tática; é como os organismos interagem em sistemas complexos e dinâmicos. E os humanos demonstraram o epítome do pensamento de curto prazo ao capacitar os oligarcas que estão tentando acabar com ele.

Esforços estão em andamento. A UE já tem vários anos de experiência com mandatos de interoperabilidade e conhecimentos preciosos sobre a forma como empresas determinadas trabalham para contornar tais leis. Os EUA, no entanto, ainda estão nos primeiros dias de garantia da interoperabilidade de software, por exemplo, para videoconferências.[69]

Talvez uma forma de motivar e encorajar os reguladores e os responsáveis pela aplicação da lei em todo o mundo seja explicar que a arquitetura subterrânea da Internet se tornou uma terra sombria onde a evolução praticamente parou. Os esforços dos reguladores para tornar competitiva a Internet visível terão poucos resultados, a menos que também enfrentem a devastação que está subjacente.

Próximos passos

Muito do que precisamos já está aqui. Além dos reguladores que insistem em em visão e novas estratégias de litígio ousadas e corajosas, precisamos de políticas governamentais vigorosas e pró-competitivas em matéria de aquisições, investimentos e infraestruturas físicas. As universidades devem rejeitar ofertas de financiamento para pesquisas de empresas de tecnologia porque isso sempre vem com condições,[70] tanto explícitas como implícitas.[71]

Em vez disso, precisamos de mais pesquisa tecnológica com financiamento público e com resultados divulgados publicamente. Essa pesquisa deverá investigar a concentração de poder no ecossistema da Internet e alternativas práticas a ela. Precisamos reconhecer que grande parte da infraestrutura da Internet é um serviço de utilidade pública de fato sobre o qual temos de recuperar o controle.

Temos de garantir incentivos regulamentares e financeiros e apoio a alternativas, incluindo a gestão comum de recursos, redes comunitárias e uma miríade de outros mecanismos de colaboração que as pessoas têm utilizado para viabilizar bens públicos essenciais, como estradas, defesa e água potável.

Tudo isso exige dinheiro. Os governos estão sedentos de receitas fiscais derivadas do ingresso inédito inesperado dos gigantes tecnológicos de hoje, por isso é claro onde está o dinheiro. Precisamos recuperá-lo.

Sabemos de tudo isso, mas ainda achamos muito difícil agir coletivamente. Por que? Agrupados em plantações tecnológicas rígidas, em vez de ecossistemas diversificados e funcionais, é difícil imaginar alternativas. Mesmo aqueles que conseguem ver claramente podem sentir-se desamparados e sozinhos. A *rewilding* une tudo o que

69 https://www.theverge.com/2024/4/8/24119268/wyden-secure-interoperable-goverment-collaboration-technology-act-encryption

70 https://www.washingtonpost.com/technology/2023/12/06/academic-research-meta-google-university-influence/

71 https://edition.cnn.com/2023/12/04/tech/facebook-disinformation-whistleblower/index.html e https://www.newstatesman.com/science-tech/2019/06/how-big-tech-funds-debate-ai-ethics

sabemos que precisamos fazer e traz consigo uma caixa de ferramentas e uma visão totalmente novas.

Os ecologistas enfrentam os mesmos sistemas de exploração e estão se organizando com um sentido de urgência, em escala e em vários domínios. Eles vêem claramente que as questões não são isoladas,[72] mas são exemplos da mesma patologia de comando e controle, extração e dominação que o antropólogo político James C. Scott notou pela primeira vez na silvicultura científica. As soluções são as mesmas na ecologia e na tecnologia: usar agressivamente o estado de direito para nivelar o capital e o poder desiguais, e depois apressar-se a preencher as lacunas com melhores formas de fazer as coisas.

Mantenha a Internet como a Internet

Susan Leigh Star, socióloga e teórica de infraestruturas e redes, escreveu no seu influente artigo de 1999, "The Ethnography of Infrastructure":

> "Estudar uma cidade e negligenciar os seus esgotos e fontes de energia (como muitos fizeram), perderá aspectos essenciais da justiça distributiva e do poder de planeamento. Estude um sistema de informação e negligencie seus padrões, fios e configurações, e você perderá aspectos igualmente essenciais de estética, justiça e mudança."[73]

Os protocolos e padrões técnicos que fundamentam a infraestrutura da Internet são ostensivamente desenvolvidos em organizações que definem padrõe abertas e colaborativas, mas também estão cada vez mais sob o controle de algumas empresas. O que parecem ser normas "voluntárias" são muitas vezes as escolhas comerciais das maiores empresas.

O domínio dos organismos de normalização por parte das grandes empresas também molda o que *não* é padronizado – por exemplo, a busca na Web, que é efetivamente um monopólio global. Embora os esforços para abordar diretamente a consolidação da Internet[74] tenham sido repetidamente levantados nos organismos de padronização,[75] pouco progresso foi feito. Isto está a prejudicar a credibilidade desses organismos, especialmente fora dos EUA.[76] Os As entidades de padronização devem mudar radicalmente ou perderão o seu mandato global implícito de administrar o futuro da Internet.

Precisamos que os padrões da Internet sejam globais, abertos e produtivos. São as amarras que dão à Internet a sua forma planetária, os fios finos, mas fortes como aço, que mantêm unida a sua interoperabilidade contra a fragmentação e o domínio permanente.

Faça com que leis e padrões funcionem juntos

Em 2018, um pequeno grupo de californianos conseguiu que o Legislativo aprovasse[77] a Lei de Privacidade do Consumidor da Califórnia.[78] Aninhada no estatuto estava uma disposição despretensiosa, o "direito de optar por não vender ou compartilhar" suas

72 https://www.wbur.org/onpoint/2023/12/08/inside-the-rewilding-movement
73 https://ics.uci.edu/~wscacchi/GameLab/Recommended%20Readings/ethnography-infrastructure-Star-1999.pdf
74 https://www.ietf.org/archive/id/draft-nottingham-avoiding-internet-centralization-02.html
75 https://datatracker.ietf.org/doc/pdf/draft-mcfadden-cnsldtn-effects-01
76 https://www.gov.uk/cma-cases/investigation-into-googles-privacy-sandbox-browser-changes
77 https://www.nytimes.com/2018/05/13/business/california-data-privacy-ballot-measure.html
78 https://oag.ca.gov/privacy/ccpa#:~:text=The California Consumer Privacy Act,how to implement the law.

informações pessoais por meio de um "controle de privacidade global habilitado pelo usuário" ou sinal CPG, que requeria um método automatizado para fazer isso. A lei não definia como o CPG funcionaria. Como era necessário um padrão técnico para que navegadores, empresas e provedores falassem a mesma língua, os detalhes do sinal foram delegados a um grupo de especialistas.

Em julho de 2021, o procurador-geral da Califórnia determinou que todas as empresas usassem o recém-criado CPG para consumidores residentes na Califórnia que visitassem seus sites.[79] O grupo de especialistas está agora orientando a especificação técnica através do desenvolvimento de padrões globais da Web no Consórcio WWW.[80] Para residentes da Califórnia, o CPG automatiza a solicitação para "aceitar" ou "rejeitar" vendas de seus dados, como rastreamento baseado em cookies, em seus sites. No entanto, ainda não é compatível com os principais navegadores padrão, como Chrome e Safari. A ampla adoção levará tempo, mas é um pequeno passo na mudança dos resultados do mundo real, ao inserir as práticas antimonopólio profundamente no acervo de padrões – e já está sendo adotado em outros lugares.[81]

O GPC não é o primeiro padrão aberto legalmente obrigatório, mas foi deliberadamente projetado desde o primeiro dia para unir a formulação de políticas e o estabelecimento de padrões. A ideia está ganhando terreno. Um relatório recente do Conselho dos Direitos Humanos das Nações Unidas recomenda que os estados deleguem "funções reguladoras a organizações que estabelecem normas".[82]

Torne os provedores de serviços - não os usuários - transparentes

A Internet de hoje oferece transparência mínima dos principais provedores de infraestrutura de Internet. Por exemplo, os navegadores são peças de infraestrutura altamente complexas que determinam como bilhões de pessoas usam a Web, mas são fornecidos gratuitamente. Os mecanismos de busca mais usados fazem acordos financeiros opacos com as empresas dos navegadores, pagando-as para serem definidos como padrão. Como poucas pessoas mudam seu mecanismo de busca preferido, navegadores como Safari e Firefox ganham dinheiro padronizando a barra de pesquisa para o Google,[83] garantindo seu domínio mesmo quando piora a qualidade dos resultados da busca.[84]

Isso cria um dilema. Se as autoridades antitruste impusessem concorrência, os navegadores perderiam a sua principal fonte de rendimento. As infraestruturas exigem dinheiro, mas a natureza planetária da Internet desafia o nosso modelo de financiamento público, deixando a porta aberta à captação privada. Contudo, se encararmos o atual sistema opaco como aquilo que é, uma espécie de tributação não estatal, então poderemos criar uma alternativa.

Os motores de busca são um local lógico para os governos exigirem a cobrança de uma taxa que apoia navegadores e outras infraestruturas essenciais da Internet, que

79 https://www.huntonprivacyblog.com/2021/07/15/california-attorney-general-updates-ccpa-faqs-indicating-mandatory-compliance-with-global-privacy-control/
80 https://w3cping.github.io/administrivia/2023/charter.html
81 https://usercentrics.com/knowledge-hub/what-is-global-privacy-control/
82 https://documents.un.org/doc/undoc/gen/g23/117/05/pdf/g2311705.pdf?token=2yMgO0WoPF6QcYggQX&fe=true
83 https://www.forbes.com/sites/johanmoreno/2021/08/27/google-estimated-to-be-paying-15-billion-to-remain-default-search-engine-on-safari/
84 https://www.theatlantic.com/technology/archive/2023/09/google-search-size-usefulness-decline/675409/

poderia ser financiada de forma transparente sob supervisão aberta, transnacional e multilateral.

Abra espaço para crescer

Precisamos parar de pensar que a infraestrutura da Internet é muito difícil de consertar. É o sistema subjacente que usamos para quase tudo o que fazemos. O antigo primeiro-ministro da Suécia, Carl Bildt, e o antigo vice-ministro dos Negócios Estrangeiros canadense, Gordon Smith, escreveram em 2016 que a Internet estava se tornando "a infraestrutura de todas as infraestruturas".[85] É assim que organizamos, conectamos e construímos conhecimento, até mesmo — talvez — inteligência planetária. Neste momento, está concentrada, frágil e totalmente tóxica.

Os ecologistas reorientaram o seu campo como uma "disciplina de crise", um campo de estudo em que não se trata apenas de aprender coisas, mas de salvá-las.[86] Nós, tecnólogos, precisamos fazer o mesmo. Regenerar a Internet significa melhorar o que as pessoas estão fazendo através da regulamentação, do estabelecimento de normas e de novas formas de organizar e construir infraestruturas, para contar uma história partilhada sobre onde queremos ir. É uma visão compartilhada com muitas estratégias. Os instrumentos de que necessitamos para nos afastarmos das monoculturas tecnológicas extrativas estão disponíveis ou prontos para serem construídos.

(*) Maria Farrell é escritora e palestrante sobre tecnologia e o futuro. Ela trabalhou em política tecnológica na Câmara de Comércio Internacional, na Internet Corporation for Assigned Names and Numbers e no Banco Mundial. Robin Berjon é especialista em governança digital e contribuiu para vários padrões da Web, incluindo o Global Privacy Control. Ele trabalha em novos protocolos da Web, como o InterPlanetary File System, e faz parte do Conselho de Administração do World Wide Web Consortium e do Painel Consultivo de Tecnologia do Gabinete do Comissário de Informação do Reino Unido.

85 https://www.tandfonline.com/doi/full/10.1080/23738871.2016.1235908
86 https://onlinelibrary.wiley.com/doi/abs/10.1111/1600-0498.12149

NETmundial+10: Evoluindo a Governança Multissetorial da Internet e Processos de Políticas Digitais[1]

Everton T. Rodrigues[2]
Flávio R. Wagner[3]
Vinicius W. O. Santos[4]

A crescente complexidade das discussões relacionadas à Internet e seus mecanismos de governança tem nos anos de 2024 e 2025 um momento crítico. Diferentes espaços competem pelo protagonismo da discussão de temas caros ao cotidiano de uso da Internet, sem necessariamente considerar seus aspectos de governança de maneira mais dedicada. De forma simplificada, iniciativas decorrentes dos processos WSIS[5] e do Pacto Digital Global (GDC) proposto pelo Secretário Geral da ONU são alguns exemplos de esforços que têm movimentado as discussões no campo[6].

O uso intercambiável de termos como "governança da Internet", "governança digital", "políticas digitais" e outros para tratar de assuntos muito similares dá pistas sobre a fragmentação dos debates e dos temas discutidos, indicando uma demanda crescente por coordenação nesse ambiente. Para enfrentar esses e outros desafios, principalmente a partir da promoção e expansão das práticas multissetoriais, o Comitê Gestor da Internet no Brasil (CGI.br), com o apoio de interlocutores relevantes de diferentes países e setores, aprovou a realização do evento NETmundial+10[7], como um espaço para reforçar a relevância da governança multissetorial da Internet e dos processos de políticas digitais nos mais diversos níveis.

[1] Este texto foi escrito por integrantes do Secretariado do NETmundial+10.
[2] Mestre em Divulgação Científica e Cultural pela Unicamp e Assessor especialista do CGI.br.
[3] Doutor em Engenharia da Computação pela Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU) e Professor da Universidade Federal do Rio Grande do Sul.
[4] Doutor em Política Científica e Tecnológica pela Unicamp e Coordenador de Governança e Políticas de Internet no CGI.br.
[5] World Summit on the Information Society, ou Cúpula Mundial da Sociedade da Informação - CMSI, no acrônimo em português: https://www.itu.int/net/wsis/
[6] Ver https://www.un.org/techenvoy/global-digital-compact
[7] Diversas informações e detalhes sobre o processo serão tratados neste texto. Ainda assim, muitos outros detalhes, materiais e referências diversas estarão acessíveis no site oficial do evento em https://netmundial.br/

Antecedentes do NETmundial+10

As disputas entre diferentes grupos de atores em torno dos mecanismos institucionais que pretendem ter poder decisivo sobre a Internet não são novidade. Há como rastreá-las até os antecedentes da WSIS e das discussões em torno da criação da ICANN[8], entre inúmeros outros processos do campo. O avanço da chamada Agenda Digital é algo que se materializa nos mais diversos processos e arenas, em contextos nacionais, regionais e globais, seja do ponto de vista individual dos Estados, seja em processos globais multilaterais e multissetoriais. Este avanço é também bastante visível em agências das Nações Unidas tais como a UIT, que consolida sua atuação em torno das tecnologias de informação e comunicação (TICs) de maneira ampla, a Unesco e sua forte atuação no campo da inteligência artificial (IA) e outros assuntos como a governança de plataformas digitais, assim como o próprio Fórum de Governança da Internet (IGF), que desde 2006 já debate temas bem mais abrangentes sobre o ambiente digital como um todo.

A Agenda de Túnis[9] criou o Fórum de Governança da Internet (IGF)[10], uma plataforma global multissetorial que facilita a discussão de questões de políticas públicas relativas à Internet, com um mandato inicial de 5 anos. Em 2010 esse mandato foi renovado por outros 5 anos; a segunda renovação de mandato foi feita durante o processo WSIS+10, com a aprovação de 10 anos adicionais, a serem encerrados em 2025[11], quando acontecerá a Revisão WSIS+20. O sucesso da Internet decorre precisamente do esforço de incontáveis atores, tanto na sua implementação como nos seus mecanismos de governança. É por isso que a Agenda de Túnis reconhece que a Governança da Internet se dá "com base na plena participação de todas as partes interessadas, tanto dos países desenvolvidos como dos países em desenvolvimento, dentro dos seus respectivos papéis e responsabilidades". Esse espírito de colaboração não deveria ser restrito somente às camadas de operação técnica da Internet, mas também deve permear todos os seus pontos de coordenação e governança, inclusive na camada de aplicações e serviços.

Tensões políticas, com potencial de impacto para instituições-chave do funcionamento da Internet, que não haviam sido resolvidas desde a CMSI, chegaram ao seu ápice durante a primeira metade da década de 2010. O vínculo da ICANN com

[8] Corporação da Internet para a Atribuição de Nomes e Números, https://www.icann.org/

[9] Documento resultante da segunda fase da Cúpula Mundial da Sociedade da Informação, realizada em 2005. Ver Cadernos CGI.br | Documentos da Cúpula Mundial sobre a Sociedade da Informação: Genebra 2003 e Túnis 2005, disponível em https://cgi.br/publicacao/cadernos-cgi-br-documentos-cmsi/

[10] Ver https://www.intgovforum.org/

[11] Ver https://www.intgovforum.org/en/content/the-igf-and-un-processes

o governo dos Estados Unidos da América[12], por meio de um contrato que supervisionava as funções da IANA[13], foi tópico permanente de discussão até meados dessa década, já que alguns atores entendiam que esse contrato daria preponderância ao governo daquele país no controle do sistema de nomes de domínios (DNS) e da Internet.

Tal percepção e outras preocupações ganharam força quando Edward Snowden revelou, em meados de junho de 2013, um esquema em escala global de vigilância em massa e espionagem pelo governo dos EUA. Snowden mostrou que a espionagem era realizada contra cidadãos e autoridades de todo o planeta, incluindo a presidenta brasileira à época, Dilma Rousseff, que se reuniu em setembro[14] com o CGI.br para debater esse assunto e a tramitação do *Marco Civil da Internet*[15]. Posteriormente, Dilma utilizou o *Decálogo de Princípios* [16] do CGI.br em seu discurso na abertura da Assembleia Geral da ONU em setembro de 2013, quando abordou os casos de espionagem e chamou a comunidade internacional para debater, no Brasil, a governança global da Internet. Em outubro, uma carta conjunta de atores-chave da comunidade técnica da Internet, reunidos em Montevidéu (Uruguai), apontou a clara necessidade de reforçar e desenvolver continuamente os mecanismos para a cooperação global multissetorial na Internet[17]. Mesmo com a mobilização de atores em escala global para discutir esse e outros assuntos, faltava um espaço de debate com as condições apropriadas, inexistentes nos fóruns estabelecidos à época, o que culminou no anúncio do evento NETmundial para 2014.

O Encontro NETmundial ocorreu nos dias 23 e 24 de abril de 2014, em São Paulo, reunindo 1.480 participantes presencialmente e remotamente, de uma diversidade de 97 países. Ele foi baseado em um modelo multissetorial com real igualdade de participação para todos ("equal footing"), organizado em comitês, que objetivou uma participação ampla e diversa. Um Comitê de Alto Nível e um Comitê Executivo Multissetorial, com atores variados, incluindo membros governamentais de alto escalão de diversos países, cuidaram da dinâmica e da programação do evento, que

[12] A ICANN está sob as leis da Califórnia e do governo dos EUA.

[13] A Internet Assigned Numbers Authority (IANA) tem como função principal a atribuição e alocação de identificadores únicos que dão base ao funcionamento da Internet globalmente. De maneira geral, a IANA é responsável pelo gerenciamento do que se conhece por recursos críticos da Internet: nomes de domínio, recursos de numeração e a base de protocolos para a operação da rede globalmente.

[14] Ver https://cgi.br/noticia/releases/cgi-br-reune-se-com-a-presidenta-da-republica/

[15] Ver https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm

[16] Ver https://principios.cgi.br/

[17] Declaração de Montevidéu sobre o futuro da cooperação na Internet
https://www.icann.org/en/announcements/details/montevideo-statement-on-the-future-of-internet-cooperation-7-10-2013-en

foi realizado em uma parceria entre o CGI.br e a rede /1Net, um fórum que abrigava diferentes organizações técnicas globais envolvidas com a governança da Internet. Como resultado, o evento produziu a "Declaração Multissetorial NETmundial"[18], também conhecida como "Declaração Multissetorial de São Paulo", que contém duas partes: um conjunto de princípios para o desenvolvimento, o uso e a governança da Internet global; e um roteiro para a evolução futura da Internet.

Nos anos subsequentes ao NETmundial, diversos atores consideraram a possibilidade de se fazer um evento de seguimento, para dar continuidade ao sucesso do primeiro. Duas sessões foram organizadas nos IGFs de 2018 e 2019, para dar continuidade ao legado do NETmundial. No IGF 2018[19], em Paris, um workshop focou em avaliar a evolução dos debates de Governança da Internet, a implementação dos princípios de 2014 e a realização de um evento de revisão do NETmundial no ano seguinte. No IGF de Berlim, em 2019, uma atividade do "dia zero" foi realizada, além de um pré-evento chamado "NETmundial+5"[20], que debateu a sequência e implementação dos princípios adotados em 2014. Embora essa ação não tenha angariado articulações sólidas o suficiente para a continuidade das discussões, ela serviu para que diferentes atores resguardassem o legado do NETmundial para eventualmente realizar outra rodada de discussão nos mesmos moldes.

Em 2018, em uma espécie de ponto de inflexão no campo, o Secretário-Geral da ONU constituiu um painel de alto nível para tratar de temas relacionados com o tema da cooperação digital. O painel foi constituído por profissionais de diversos países, e entregou, em 2019, um relatório sobre seu trabalho, intitulado "A Era da Interdependência Digital: Relatório do Painel de Alto Nível sobre Cooperação Digital"[21]. O relatório lida com uma diversidade de temas e gerou uma série de respostas, contribuições e processos posteriores, engajando *stakeholders* globais os mais variados.

Mais recentemente, em 2020, o Secretário-Geral da ONU, como parte das comemorações de 75 anos da entidade, publicou o relatório "Nossa Agenda Comum"

[18] Declaração Multissetorial de São Paulo (NETmundial). Disponível em https://www.cgi.br/media/docs/publicacoes/4/Documento_NETmundial_pt.pdf

[19] IGF 2018 WS #178 Towards NetMundial +5: Disponível em https://www.intgovforum.org/en/content/igf-2018-ws-178-towards-netmundial-5

[20] IGF 2019 Pre-Event #32 NETmundial+5: The Legacy and Implications for Future Internet Governance https://www.intgovforum.org/en/content/igf-2019-pre-event-32-netmundial5-the-legacy-and-implications-for-future-internet-governance

[21] "The Age of Digital Interdependence: Report of the High-level Panel on Digital Cooperation", no original em inglês. Ver https://cgi.br/publicacao/cadernos-cgi-br-a-era-da-interdependencia-digital/

(*Our Common Agenda*), no qual propôs um Pacto Digital Global (GDC)[22], a ser aprovado durante a Cúpula do Futuro[23], planejada para setembro de 2024. Junto a isso, a comunidade global prepara-se, ainda, para a revisão WSIS+20, a ser realizada em 2025, em que se discutirá, entre outros, os caminhos e o futuro do IGF.

Todo esse arcabouço de debates e processos decisórios, muitos concorrentes e em paralelo, dão o tom da complexidade do ecossistema, que precisa lidar com diversas sobreposições e pouca coordenação, elementos principais das discussões relacionadas à fragmentação da governança da Internet e processos de políticas digitais. A essa fragmentação acrescenta-se uma multiplicidade de outros espaços globais de discussão de temas específicos, tais como inteligência artificial e segurança cibernética, em grande parte em um contexto intergovernamental, com baixa oportunidade de participação significativa dos demais setores.

A condução das negociações do GDC[24] tem levantado diversos questionamentos pela comunidade global, principalmente do ponto de vista da transparência e participação de atores não governamentais. Tais preocupações não se relacionam unicamente com a atuação de governos, mas também com a forma como esse processo enxerga os demais atores relevantes. Em junho de 2023, o Enviado de Tecnologia do Secretário-Geral das Nações Unidas, Embaixador Amandeep Singh Gill (*UN Tech Envoy*), declarou que um modelo tripartite seria o necessário para compor um novo Fórum de Cooperação Digital, e dele participariam o setor privado, os governos e a "sociedade civil" (entendida aqui de maneira ampliada, agregando também a comunidade técnica e a academia). A declaração provocou uma reação de atores relevantes da comunidade técnica: ICANN, ARIN[25] e APNIC[26] publicaram um comunicado conjunto[27] com fortes críticas às declarações do Tech Envoy e em defesa do modelo multissetorial da Internet que trate a comunidade técnica como um ator com interesses distintos dos demais setores.

Desde a publicação do relatório Nossa Agenda Comum, pelo Secretário-Geral da ONU, a realização da Cúpula do Futuro apresentou-se como um risco concreto para as discussões abertas e multissetoriais, voltadas à obtenção de consenso entre diferentes grupos de atores interessados no desenvolvimento e uso da Internet,

[22] Ver https://www.un.org/techenvoy/global-digital-compact
[23] Ver https://www.un.org/en/summit-of-the-future
[24] A negociação e aprovação do GDC será efetivada entre os estados-membros da ONU.
[25] Um dos cinco Registros Regionais da Internet, responsável principalmente pela América do Norte.
[26] Um dos cinco Registros Regionais da Internet, responsável pela região da Ásia-Pacífico.
[27] O comunicado está publicado no site da ICANN: https://www.icann.org/en/blogs/details/the-global-digital-compact-a-top-down-attempt-to-minimize-the-role-of-the-technical-community-21-08-2023-en

levando-se em conta que o condão decisório desta cúpula é restrito ao âmbito governamental. No mesmo sentido, o reforço de espaços multilaterais trazido pelas discussões atualmente em andamento a respeito do GDC traz riscos ao IGF como um esforço global multissetorial capaz de ser aprimorado para entregar resultados ainda mais relevantes. Dessa forma, dada a preponderância atual de iniciativas multilaterais que pretendem ser responsáveis pelo desenvolvimento e uso da Internet, o modelo de governança multissetorial precisaria ser reafirmado antes que esses processos decisórios avançassem ainda mais.

Entendendo o momento como crítico para a definição dos rumos da governança do ecossistema digital, diversos atores da comunidade global interessada, principalmente aqueles já tradicionalmente envolvidos com a governança da Internet, identificaram a necessidade de um espaço de discussão essencialmente multissetorial que pudesse proporcionar um ambiente adequado para os atores dos diferentes setores formarem consenso em torno de mensagens firmes, com recomendações concretas sobre a preservação e aplicação do multissetorialismo nos mais diversos espaços de governança, inclusive em espaços multilaterais. Assim, após diversos diálogos e articulações no interior dessa comunidade, o NETmundial+10 foi proposto enquanto uma possível ferramenta para tratar todas essas questões.

A concepção do NETmundial+10

Em setembro de 2023, a partir de uma análise do cenário descrito acima, o CGI.br aprovou a realização do evento NETmundial+10 para debater os desafios globais relacionados à governança da Internet e do espaço digital, com a condição de que o evento somente seria viável do ponto de vista político a partir de uma articulação com parceiros internacionais. Dessa forma, o CGI.br passou a dialogar com atores e potenciais parceiros nos principais eventos de governança da Internet, como o IGF 2023, realizado em Quioto, Japão; a ICANN78, realizada em Hamburgo, Alemanha; e a Conferência Mundial da Internet, realizada em Wuzhen, China, entre outros. Esses primeiros contatos proporcionaram as linhas gerais do que um NETmundial+10 deveria abordar, atendendo tanto ao que foi aprovado pelo CGI.br como aos anseios dos diferentes atores consultados. Durante essas reuniões, buscou-se também eventual apoio institucional para a realização do evento. Dentre os tópicos mais frequentes nas reuniões, pode-se destacar:

- a defesa do modelo multissetorial para a governança da Internet;

- o NETmundial+10 não deveria sobrepor atividades ou replicar estruturas do IGF;
- o reforço do papel do IGF; e
- a necessidade de uma metodologia leve de trabalho, dado o curto prazo para a organização do evento.

Linha do tempo resumida

- 22 de setembro de 2023: Aprovação da realização do NETmundial+10 pelo CGI.br: https://cgi.br/reunioes/ata/2023/09/22/
- Outubro e novembro de 2023: reuniões com potenciais apoiadores e parceiros, durante o IGF em Quioto, no Japão; a ICANN78, em Hamburgo, na Alemanha; e a Conferência Mundial da Internet, em Wuzhen, China.
- 23 de novembro de 2023: publicação da nota "NETmundial+10 – Global challenges for the governance of the digital world": https://cgi.br/noticia/notas/netmundial-10-global-challenges-for-the-governance-of-the-digital-world/.
- Dezembro de 2023: reuniões do grupo de escopo do NETmundial+10, que definiu as linhas principais do evento.
- 22 de dezembro de 2023: Declaração conjunta do NETmundial+10: https://netmundial.br/statement/joint-statement-of-the-netmundial10
- Janeiro de 2024: coleta de assinaturas para a Declaração Conjunta.
- Fevereiro de 2024: composição e início dos trabalhos do Comitê Executivo de Alto Nível (HLEC).
- Março de 2024: processo para as manifestações de interesse em participar do evento.
- 22 de março de 2024: lançamento da consulta online do NETmundial+10 e da programação preliminar.
- 10 de abril de 2024: encerramento da consulta online do NETmundial+10.
- 25 de abril de 2024: publicação da versão preliminar da declaração final, com base no conteúdo recebido por meio da consulta online.
- 29 e 30 de abril de 2024: realização do evento e coleta de contribuições nas Sessões de Trabalho.

- 30 de abril de 2024: reunião final do HLEC e publicação do Documento Final do NETmundial+10, ao final do evento.

Processo antes e durante o evento

O NETmundial+10 foi fortemente baseado no evento de 2014, com necessárias adaptações de conteúdos e dinâmicas. Tal como em 2014, o evento foi construído de forma a propiciar diálogos multissetoriais significativos que pudessem gerar resultados tangíveis sobre os temas em discussão. Diferentemente de 2014, quando o evento se focou em discussões temáticas do campo da governança da Internet e gerou princípios em diversas áreas, tais como inovação, segurança e direitos humanos, o evento de 2024 atuou sobre um recorte mais restrito da agenda, buscando oferecer propostas concretas de mecanismos e diretrizes para o avanço das práticas multissetoriais em todas as esferas relevantes.

No início, foram realizados diálogos informais com potenciais parceiros da iniciativa, oriundos de diferentes setores e regiões geográficas, de forma a identificar forças, fraquezas, riscos e oportunidades, bem como estabelecer a rede de apoiadores-chave para viabilizar o evento politicamente. Após diversos diálogos e devolutivas positivas sobre a ideia e os objetivos, as linhas gerais de um evento foram aprimoradas e o CGI.br passou a de fato construir uma rede para o NETmundial+10. O primeiro passo foi estabelecer um "grupo de escopo", que trabalhou basicamente durante o mês de dezembro de 2023 e definiu as bases e as linhas mestras do evento, materializadas na declaração conjunta que foi amplamente disseminada no início de janeiro de 2024, para coleta de assinaturas.

O grupo foi montado a partir, principalmente, do grupo de pessoas e organizações com quem o CGI.br havia conversado durante os eventos de governança no segundo semestre de 2023. Foi um grupo formado por cerca de 15 pessoas, com diversidade de gênero, setores e regiões. Com reuniões remotas no mês de dezembro de 2023, o grupo debateu e construiu uma declaração conjunta, por meio da qual o CGI.br e seus parceiros divulgaram para a comunidade global o processo em andamento para organizar o evento, além de especificar o escopo das discussões no avanço do multissetorialismo para a governança do mundo digital, especialmente no que se refere aos mecanismos e práticas que devem estar na base de seu funcionamento.

Com a declaração publicada no início de janeiro, um processo de coleta de assinaturas foi iniciado e durou até o fim do mês, tendo sido recebidos cerca de 300 apoios de indivíduos e organizações de todo o mundo. Todos os apoios foram registrados e documentados no site do evento. Dali em diante foi iniciado o processo

de composição do que seria o Comitê Executivo do evento. Em 2014, o evento havia contado com uma estrutura em torno de quatro comitês diferentes, com missões bem estabelecidas. Para 2024, houve uma redução dessa estrutura, de modo a adequar as expectativas ao tempo e recursos disponíveis para a realização do evento. A estrutura do Comitê Executivo de 2014 foi tomada como base e acoplada às demais, dando origem ao que ficou conhecido como Comitê Executivo de Alto Nível, ou High Level Executive Committee, em inglês (HLEC). O HLEC foi composto e iniciou seus trabalhos em meados de fevereiro de 2024, quando o processo de produção do evento foi acelerado.

Com pouco mais de dois meses de trabalho intenso, o HLEC, apoiado por um Secretariado reduzido, fez diversas reuniões online de modo a debater e tomar as diversas decisões relacionadas com a organização do evento. Pela estrutura mais compacta, algumas funções foram separadas. Enquanto o HLEC ficou responsável pelo conteúdo do evento, o anfitrião, o CGI.br, ficou totalmente responsável pelas decisões de infraestrutura e logística para o evento. O CGI.br, por meio do Núcleo de Informação e Coordenação do .br (NIC.br), organizou toda a infraestrutura física necessária. Também o CGI.br/NIC.br, por meio da alocação de recursos e pessoal, proveu o secretariado necessário às atividades do HLEC desde outubro de 2023 até a finalização de todos os processos relativos ao evento.

Para viabilizar e dinamizar a organização do evento, o HLEC foi dividido em três subgrupos: Participação, Programação e Consulta. O Subgrupo de Participação ficou responsável pela definição de critérios para a seleção de participantes, regras e orientações sobre a participação no processo, além de outras questões relacionadas com informação, participação e engajamento. O Subgrupo de Programação ficou responsável por toda a construção da programação do evento, debatendo os critérios, linhas gerais, formatos dominantes, dinâmicas de interação e composição de sessões. O Subgrupo de Consulta foi responsável por projetar a consulta online que preparou as bases para o evento, definindo os temas e conteúdo da consulta, bem como outros aspectos de implementação necessários, além de uma apreciação posterior dos resultados.

Ao longo do processo, a dinâmica foi, basicamente, a de reuniões de subgrupos seguidas de reuniões gerais do HLEC. Cada reunião individual de subgrupo tratava de aspectos específicos das pautas daquele subgrupo, com posterior relato na reunião geral do HLEC. As reuniões do HLEC também tinham o papel de pautar tópicos em aberto, além de encaminhar questões para os subgrupos, ou mesmo tratar de outras questões não previstas e/ou pendentes. No total, foram sete

reuniões online do HLEC, antes do evento. Na sequência, o HLEC realizou sua primeira reunião presencial no dia anterior ao NETmundial+10, 28 de abril de 2024. O comitê se reuniu ainda nos dias 29 e 30, durante o evento, para continuar processando contribuições da comunidade e preparar o documento final. No dia 30, o HLEC permaneceu em reunião durante toda a tarde, para conseguir processar todas as contribuições recebidas e finalizar o documento que foi apresentado na plenária final. Os subgrupos, por sua vez, realizaram diversas reuniões ao longo do processo de organização do evento, de maneira não necessariamente uniforme, sempre dependendo das demandas de cada grupo.

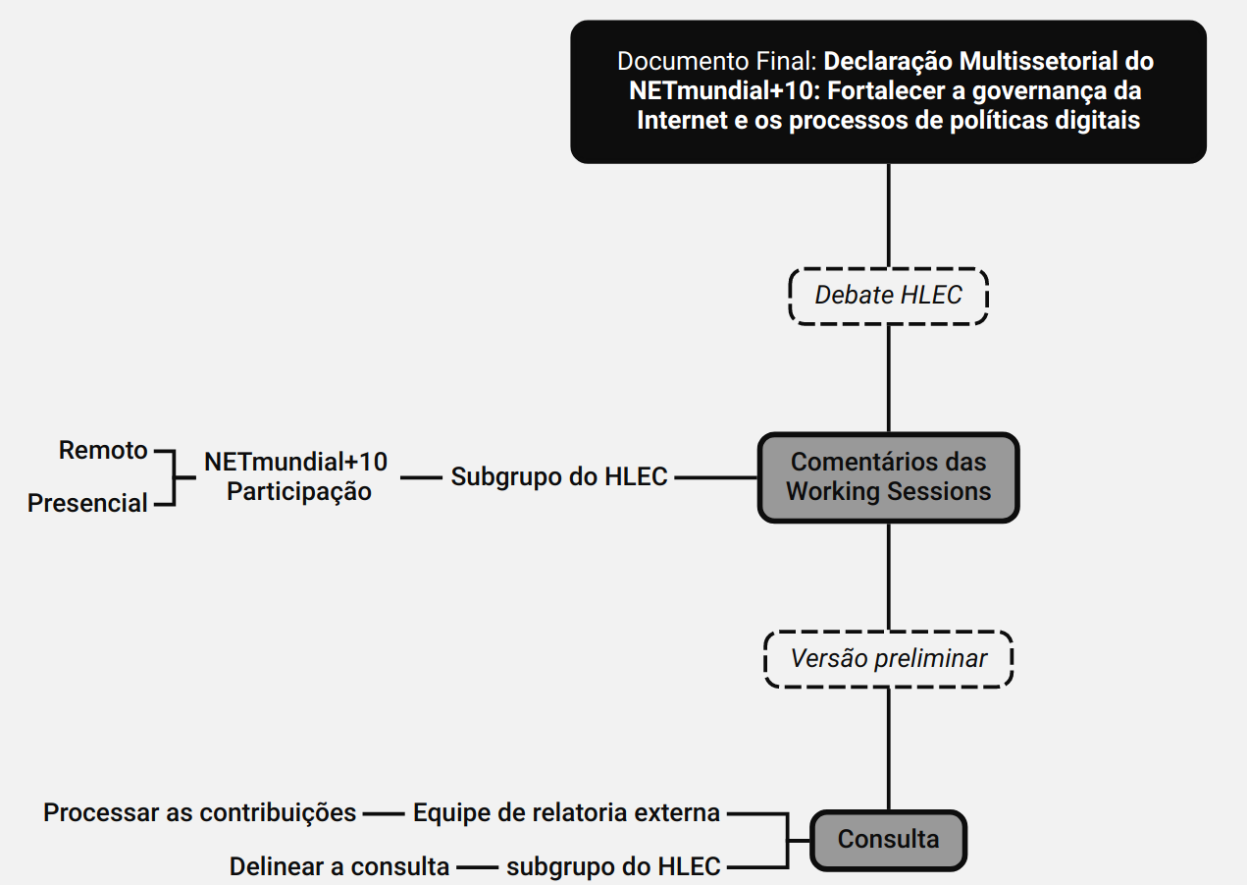
A consulta online foi a espinha dorsal de todo o processo até a declaração final do evento. O subgrupo do HLEC responsável por delinear o conteúdo e o processo da consulta fez diversas reuniões para chegar a uma versão final da consulta que foi publicada na plataforma. O formato final envolveu uma mescla de perguntas abertas e fechadas, sendo estas últimas baseadas em estrutura de escala Likert, em que os respondentes precisavam indicar o nível de acordo ou desacordo com determinadas afirmações. O conjunto de questões ainda contou com uma questão do tipo ranqueamento e outra de priorização. A consulta foi lançada no dia 22 de março de 2024, sendo encerrada no dia 10 de abril seguinte, recebendo contribuições diversas de um total de 154 respondentes. Uma equipe dedicada analisou e categorizou os conteúdos recebidos, produzindo um relatório que foi utilizado pelo secretariado e pelo HLEC nas discussões do documento final do evento. A partir do relatório da consulta, o secretariado preparou um conjunto de conteúdos estruturados pelas contribuições, de modo que este foi o insumo inicial de trabalho do HLEC para a elaboração da declaração final.

Passada a primeira fase de processamento da consulta, passou-se à segunda fase, em que o HLEC escolheu membros responsáveis pela redação de partes específicas do documento final (*pen holders*). Os *pen holders* trabalharam por um espaço de tempo muito curto e entregaram propostas de consolidação dos trechos com base nos insumos a eles providos. A estrutura do documento final foi quase que totalmente baseada na estrutura de tópicos da consulta online, de modo a manter a coerência do processo e garantir a adição de contribuições da comunidade de maneira mais direta. Após esse primeiro trabalho de consolidação dos *pen holders*, o HLEC se reuniu e finalizou a versão preliminar do documento, que foi publicada em 25 de abril de 2024, poucos dias antes do evento, de modo a dar conhecimento prévio dos conteúdos à comunidade, para proporcionar tempo e espaço para debates dentro dos setores.

Com o documento preliminar publicado, o HLEC continuou os debates e a preparação para o evento, já debatendo e definindo questões metodológicas e de processo. Membros do HLEC continuaram responsáveis por partes específicas do documento, tendo a missão de processar contribuições recebidas dos participantes durante o evento para propor versões finais das seções do documento. O evento contou com três sessões de trabalho, focadas em cada uma das principais seções temáticas do documento final. Após cada sessão, uma equipe de relatoria enviava sumários das contribuições dos participantes para que os membros do HLEC pudessem processar os conteúdos. Esse trabalho alimentou uma versão final do documento, que foi debatida pelo HLEC durante toda a tarde do segundo e último dia de evento. A versão final resultante dessa reunião foi apresentada na plenária final, com leitura completa pelos membros do HLEC que estavam na sessão. O resultado foi muito bem recebido, tanto pelos participantes do evento como por muitos outros atores da comunidade de governança da Internet.

Na sequência do evento, diversos diálogos foram iniciados para a divulgação do documento final e aumento de seu impacto. Atores da comunidade global têm utilizado os resultados do NETmundial+10 em diálogos dentro de suas próprias comunidades, além de mencioná-los em outros processos relevantes. Muito se tem conversado sobre a necessidade de expandir tais diálogos e aumentar a conscientização sobre a declaração NETmundial+10, de forma a qualificar os debates e levar consensos multissetoriais para outros processos decisórios.

Diversos membros da comunidade já têm feito referência explícita, inclusive, ao que passou a ser identificado como “Diretrizes de São Paulo”, que correspondem a uma parte importante do documento, com recomendações essenciais para a implementação de mecanismos multissetoriais de governança. A comunidade também já começa a pensar na possível implementação de outra recomendação expressa incluída no documento final do NETmundial+10, que indica o IGF como o guardião dessas diretrizes.



Esquema simplificado de como se deu o processo de construção da declaração final.

Declaração final e as Diretrizes Multissetoriais de São Paulo

Assim como em 2014, o principal produto do NETmundial+10 foi sua declaração final[28]. O documento seguiu quase que totalmente a estrutura apresentada durante a consulta online, preservando suas três principais seções temáticas: "Princípios para a Governança da Internet"; "Diretrizes para a Implementação de Mecanismos Multissetoriais"; e "Contribuições a Processos em Curso". Dado que o evento durou apenas dois dias, e a fim de adiantar o trabalho de redação, uma versão preliminar do documento foi elaborada pelo HLEC com base nas contribuições enviadas em resposta à consulta online. Diversas alterações de redação foram efetuadas durante os dois dias do evento, como resultado do processo de engajamento da comunidade, e materializadas no texto do documento final.

A declaração final contém, já em sua seção inicial, os seus principais objetivos. Dessa forma, o NETmundial+10:

[28] Ver https://netmundial.br/pdf/NETmundial10-MultistakeholderStatement-2024.pdf

- ***Ratifica a declaração do NETmundial de 2014,*** *que afirma que a Internet é um recurso global que deve ser gerido no interesse público, em conformidade com o direito internacional e a legislação internacional de direitos humanos;*
- ***Reconhece a relevância da transparência e da prestação de contas*** *para melhorar a governança da Internet e os processos de políticas digitais;*
- ***Reafirma que os 10 princípios para os processos de governança da Internet adotados em 2014 permanecem relevantes*** *e recomenda a sua aplicação no tratamento dos desafios atuais e futuros das políticas digitais;*
- ***Propõe diretrizes operacionais*** *para a implementação desses princípios em diferentes situações;*
- ***Faz contribuições para diversos processos em andamento*** *relacionados à evolução da arquitetura de governança para as políticas digitais; e*
- ***Recomenda que os princípios e as diretrizes apresentados neste documento sejam implementados por todas as partes interessadas, em todos os níveis.***

Dessa forma, a declaração multissetorial do NETmundial+10 tem basicamente três pilares: reafirmação dos princípios de processos de governança adotados em 2014, recomendação de diretrizes operacionais para a implementação dos princípios, e consolidação de mensagens para outros processos em curso que podem se beneficiar dos conteúdos da declaração.

O documento apresenta diversos conteúdos relevantes que merecem análise detida, como as recomendações de implementação do princípio multissetorial e as mensagens sobre coordenação dos espaços de governança. O coração da declaração, contudo, é o conjunto de diretrizes e etapas de processo para a colaboração e tomada de decisão multissetorial, as "Diretrizes Multissetoriais de São Paulo". No total, são 13 diretrizes, passando por aspectos específicos de como deve se dar a participação multissetorial, a necessidade de deliberação, transparência e prestação de contas, entre outros. As etapas de processo, por sua vez, envolvem orientações que passam pela identificação e engajamento de atores, compartilhamento de informações, participação equitativa, poderes da comunidade, entre outros. Espera-se que as diretrizes e etapas de processo possam ser integradas aos mais diversos processos multissetoriais e/ou multilaterais, em todos os níveis.

Um outro aspecto forte no documento é a seção dedicada a enviar mensagens a outros processos em curso. Dentre as mensagens expressas, pode-se destacar a

recomendação para que o Pacto Digital Global não crie espaços novos que possam se sobrepor com outros já existentes, indicando que os espaços atuais sejam reforçados para atingirem os objetivos esperados. Outro aspecto relevante foi a defesa do Fórum de Governança da Internet (IGF) como espaço preferencial para a coordenação de processos e monitoramento de decisões relevantes. Para além disso, o documento do NETmundial+10 também recomenda que o IGF seja o guardião das Diretrizes Multissetoriais de São Paulo, acompanhando sua implementação e evolução no ecossistema.

Como parte do próprio processo de debate multissetorial, e mesmo como parte do processo de escutar a comunidade, é também resultado do NETmundial+10 uma expressão "nova" para se referir ao campo: "Governança da Internet e Processos de Políticas Digitais". A expressão foi resultado do consenso multissetorial do Comitê Executivo, quando o mesmo debatia as contribuições da comunidade e tentava consolidar as diferentes visões. A adoção dessa expressão foi uma tentativa de resolver a cacofonia apontada em diversas contribuições, sobre o uso de diferentes expressões para se referir às mesmas coisas, de maneira intercambiável (e.g.: governança digital, governança da Internet, políticas digitais, etc).

O sucesso do NETmundial+10 se conjuga no futuro

O NETmundial+10 foi um esforço multissetorial de sucesso incontestável do ponto de vista logístico e político. A decisão de realizá-lo no primeiro semestre de 2024 não atendeu somente a um anseio por repetir ou comemorar o evento de 2014, mas também de proporcionar uma oportunidade para que o documento final pudesse ser utilizado nos eventos-chave na agenda de 2024-2025. Dessa forma, todos os interessados em defender o modelo multissetorial de governança da Internet e processos de políticas digitais têm a oportunidade de se apropriarem do formato de realização do evento e dos seus resultados, como o seu documento final e as Diretrizes de São Paulo, nele contidas, aplicando-as em todos os espaços de governança multissetoriais e multilaterais. O impacto real do NETmundial+10 não se constitui na realização do evento, mas no aprimoramento efetivo da arquitetura e dos mecanismos de governança da Internet e processos de políticas digitais.

Apesar da urgência permanente que permeou toda a organização e realização do evento, a comunidade interessada em defender o modelo multissetorial saiu fortalecida com a mobilização que foi proporcionada. O NETmundial+10 possibilitou que os interessados em continuar a construir e evoluir os mecanismos de Governança da Internet e Processos de Políticas Digitais dessem vários passos

concretos na direção da construção de consensos para o futuro da Internet. Esse feito, em si, já constitui um forte recado para as negociações crescentemente realizadas em silos, conduzidos por diferentes atores, especialmente intergovernamentais, em espaços cada vez mais fragmentados e sobrepostos, sem que a comunidade tenha uma ferramenta forte e adequada para avançar em ações e posições realmente concretas para o aprimoramento da governança do ecossistema digital global.

Governança da inteligência artificial (IA): a interação entre o local, o regional e o global em direção a uma solidariedade transnacional[1]

Adeboye Adegoke[2], Bruno Bioni[3], Fernanda K. Martins[4], Júlia Mendonça[5], Paula Guedes[6], Tainá Junquilho[7]

Introdução: geopolítica da IA

Para escrever sobre a governança global da IA é necessário refletir sobre como combinar o contexto nacional com o internacional[8]. Milton Santos, intelectual brasileiro e um dos principais acadêmicos do mundo no tema da globalização, enfatiza a interdependência entre o local e o global, argumentando que um não existe sem o outro de forma que ambos se influenciam em um processo de realimentação[9]. Segundo Santos, existem de fato espaços de globalização onde a geopolítica favorece uns e exclui outros com base nas virtualidades-potencialidades de certos grupos em detrimento de outros. Esta parece ser a questão subjacente à governança da IA e a questão mais premente de todos os fóruns nacionais e internacionais de elaboração de políticas. Então, essa tecnologia poderia evitar tornar aqueles que já são periféricos ainda mais periféricos devido à dinâmica desigual tecnologia-poder, como ocorreu em outros momentos da nossa história e que ainda hoje persiste?

Neste contexto, o debate atual deve centrar-se numa relação dinâmica emancipatória entre políticas regulatórias locais, regionais e globais no contexto da IA. Por esta razão, é preciso reconhecer que as iniciativas de regulação locais e regionais desempenham um papel fundamental na definição do desenvolvimento tecnológico e na proteção dos direitos humanos, enquanto o panorama global influencia, tanto negativa como positivamente, estes movimentos de governança através dos chamados processos regulatórios, normativos, e de interoperabilidade técnica. Caso contrário, o progresso tecnológico da IA reproduziria ou ampliaria as divisões técnico-econômicas já existentes.

1 Este ensaio é uma versão textual estendida da participação dos coautores do T20 no evento paralelo do G20 sobre "Aproveitando a Inteligência Artificial para a Equidade Social e o Desenvolvimento Sustentável": https://www.g20.org/en/calendar/side-events /aproveitando-a-inteligência-artificial-para-a-equidade-social-e-o-desenvolvimento-sustentável

2 Gerente Sênior da Paradigm Initiative e membro da força-tarefa de transformação digital inclusiva da T2O.

3 Fundador-diretor executivo da Data Privacy Brasil, doutor em Direito e co-presidente líder da Força-Tarefa de transformação digital inclusiva da T20.

4 Diretora de pesquisa e desenvolvimento na InternetLab, doutora em Ciências Sociais e membro da força-tarefa de transformação digital inclusiva da T2O.

5 Pesquisadora na Data Privacy Brasil e mestranda em Direito.

6 Ex-pesquisadora na Data Privacy Brasil e doutoranda em Direito.

7 Professora de Direito, Tecnologia e Inovação no Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP), doutora em Direito e membro da força-tarefa de transformação digital inclusiva da T2O.

8 Bruno Bioni, Marina Garrote, Paula Guedes, Temas centrais na regulação de IA: O local, o regional e o global na busca por interoperabilidade regulatória. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

9 Oxford Analytica, O aumento global da IA cria um paradoxo ambiental. Emerald Expert Briefings, n. oxan-db, 2024.

Principais áreas de políticas para a equidade social e o desenvolvimento sustentável

O G20 deveria promover um programa global de desenvolvimento de capacidades que responda às necessidades dos países desfavorecidos do sul global, capacitando-os para avaliar os potenciais impactos sociais, econômicos e ambientais das tecnologias para mitigar os riscos e maximizar o potencial da transformação digital. As estratégias locais e globais devem promover e dar prioridade à IA que promova a equidade social e, portanto, uma abordagem centrada no ser humano para enfrentar "os maiores desafios do mundo, nomeadamente, mas não limitados, à crise climática, à saúde global e à educação"[10]. Neste sentido, as Nações Unidas têm pressionado pela priorização da IA orientada à aceleração dos Objetivos de Desenvolvimento Sustentável (ODS)[11].

Os países do G20 devem reforçar a cooperação internacional em matéria de IA, baseada em um equilíbrio entre a proteção dos direitos das comunidades vulneráveis e a dignidade do trabalho que sustenta os sistemas de IA. Os países devem apoiar a investigação sobre a qualidade e dignidade do trabalho envolvendo dados, a liberdade de associação dos trabalhadores digitais e políticas sobre programas de requalificação centrados nas mulheres e nos grupos minoritários. A promoção dos ODS deve preocupar-se com a infraestrutura humana da IA com trabalho digno e qualidade de vida para a cidadania. Isto envolve diferentes formas de trabalho no ciclo de vida da IA, incluindo trabalho de identificação de dados e outras formas de trabalho terceirizadas no Sul Global.

A cooperação global também é necessária para resolver o chamado paradoxo ambiental da IA[12]. Se, por um lado, a IA pode ser poderosa para automatizar com alta precisão o desmatamento e prever desastres climáticos, por outro lado, alguns modelos específicos consomem enormes volumes de recursos naturais. Cálculos computacionais complexos, especialmente de grandes modelos de linguagem (LLM)[13], têm um enorme impacto nas emissões de gases de efeito estufa. Isso porque funcionam com hardware de alto desempenho e grande infraestrutura de aglomerados de computação, consumindo muita eletricidade e água para refrigeração. Além de considerar a compensação fiscal ambiental para grandes modelos de linguagem de IA, o G20 deveria apoiar o conhecimento científico-interdisciplinar em escala global,

10 G7 Japão, Leaders' Annex: Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems. 2023. Available at: https://g7g20-documents.org/database/document/2023-g7-japan-leaders-leaders-annex-hiroshima-process-international-code-of-conduct-for-organizations-developing-advanced-ai-systems. Acesso em 20 de maio de 2024.

11 Nações Unidas, Governing AI for Humanity: Interim Report. 2023, p.18. Disponível em: https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf. Acesso em: 20 de maio de 2024.

12 Oxford Analytica, AI global surge creates environmental paradox. Emerald Expert Briefings, n. oxan-db, 2024.

13 Xia Fan, Serge Stinckwich, On the Unsustainability of ChatGPT: Impact of Large Language Models on the Sustainable Development Goals. Disponível em: https://unu.edu/macau/blog-post/unsustainability-chatgpt-impact-large-language-models-sustainable-development-goals. Acesso em: 20 de maio de 2024.

semelhante ao Painel Intergovernamental Sobre Mudanças Climáticas (IPCC), como propôs o órgão consultivo de IA da ONU[14].

Além disso, os países do G20 deveriam considerar a proibição do gasto de recursos públicos a sistemas de IA que perpetuam a injustiça social. Por exemplo, em países como o Brasil, houve um aumento significativo na implementação de tecnologia de reconhecimento facial pelas agências de aplicação da lei[15]. Estes sistemas de IA exacerbam frequentemente a insegurança em vez de aumentarem a segurança, levando a detenções injustas e reforçando o racismo sistêmico que afeta particularmente a população afrodescendente. Entretanto, outros sistemas de IA com potencial para mitigar a letalidade policial e aumentar a transparência nas operações policiais, como os integrados com dados das câmaras corporais da polícia, têm sido largamente ignorados. Idealmente, a IA deveria servir como uma ferramenta de contra-vigilância para aumentar a segurança daqueles que foram historicamente marginalizados e vigiados, concretizando o que ficou conhecido como AI4SG (Inteligência Artificial para Bens Sociais)[16].

Por último, a cooperação internacional deve encorajar o desenvolvimento e a utilização da IA em vários idiomas, ecoando as recentes discussões entre o presidente brasileiro Lula e o primeiro-ministro espanhol[17]. Governar a IA vai além da garantia de segurança e confiança; envolve também prevenir uma divisão epistemológica entre o Sul e o Norte globais. Isto requer uma dedicação coletiva para alavancar a IA para o bem comum, evitando ao mesmo tempo a perpetuação das desigualdades e divisões existentes. A verdadeira transformação global através da IA só pode ser alcançada através de uma governança inclusiva e da igualdade de acesso.

14 Nações Unidas, Governing AI for Humanity: Interim Report. 2023, p.18. Disponível em: https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf. Acesso em: 20 de maio de 2024.

15 Júlia Maria Pereira Dias, Tainá Aguiar Junquilho, Racismo algorítmico: uma análise sobre os riscos do uso do reconhecimento facial pelos órgãos de segurança pública. In: TIC, Governança da Internet, gênero e diversidade: tendências e desafios. Organização Bia Barbosa et al. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2024, p. 125-150.

16 L. Floridi et al. How to Design AI for Social Good : Seven Essential Factors. Science and Engineering Ethics, v. 26, n. 3, p. 1771–1796, 2020.

17 Governo Federal, Brasil, "Conseguimos provar a afinidade entre nossos governos," disse Lula do Brasil sobre a visita do Presidente da Espanha. 2024. Disponível em: https://www.gov.br/planalto/en/latest-news/2024/03/201cwe-were-able-to-prove-the-affinity-between-our-governments-201d-said-brazil2019s-lula-about-visit-of-the-president-of-spain. Acesso em: 20 maio 2024.

A necessária interação entre *hard* e *soft law*[18]: dos princípios éticos a uma governança eficaz e democrática[19]

Os desenvolvimentos recentes na governança da IA sublinham a necessidade de medidas regulamentares que vão além da mera confiança em valores éticos e no cumprimento voluntário. A convergência das leis nacionais com os quadros internacionais deve não só reconhecer, mas também desencadear uma governança eficaz dos riscos globais associados à IA.

Conforme descrito nos princípios de IA do G20 de 2019, na Declaração de Hiroshima e na Declaração de Bletchley do G7 em 2023[20], o reconhecimento dos riscos inerentes à IA que afetam todos a nível internacional sublinha a necessidade de uma ação coordenada além-fronteiras. Neste sentido, o recente alerta do secretário dos EUA enfatiza a necessidade urgente de regulamentar a IA para evitar que ela nos governe[21]. Isto está alinhado com o recente lançamento da Resolução da ONU: "Aproveitar as oportunidades de sistemas de inteligência artificial seguros, protegidos e confiáveis para o desenvolvimento sustentável"[22]. A Resolução sublinha a importância de respeitar os direitos humanos ao longo de todo o ciclo de vida da IA, exortando as nações e as partes interessadas a evitarem sistemas de IA que violem os direitos humanos ou ponham em perigo grupos vulneráveis, tanto online como offline. Pode o G20 impor pelo menos uma moratória contra o reconhecimento facial para efeitos de aplicação da lei, uma vez que esses sistemas estão amplificando preconceitos e causando ainda mais insegurança para a sociedade?

No atual panorama da governança da IA, deve haver uma mudança notável no sentido da convergência de medidas nacionais de *hard law* com quadros internacionais de *soft law*, a serem orientadas numa lógica afirmativa baseada nos direitos humanos. Caso contrário, não haverá um esboço de medidas concretas para evitar uma onda tecnológica opressiva.

18 *Soft Law* refere-se a instrumentos normativos de direito internacional que não possuem força de Lei, não geram sanções e não possuem caráter vinculativo, apesar de possuírem o potencial de, por exemplo, regular comportamentos sociais e gerar efeitos de outras ordens. Por sua vez, *hard law* refere-se ao conjunto das normas de direito internacional que estabelecem regras vinculativas no direito interno, por exemplo, através de tratados e acordos, os quais permitem a aplicação de sanções concretas, seja em órgãos do país signatário ou em tribunais internacionais. Ver: NASSER, Salem Hikmat. Soft law: um estudo sobre as normas e as fontes do Direito Internacional. 2004. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2004. . Acesso em: 13 jun. 2024.

19 Governo do Reino Unido. A Declaração de Bletchley pelos países participantes da Cúpula de Segurança da IA. 1-2 de novembro de 2023. Disponível em: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. Acesso em: 20 de maio de 2024.

20 CNN. "Artificial intelligence could contribute to humanity's extinction, warns new report." CNN, 12 Mar 2024. Disponível em https://edition.cnn.com/2024/03/12/business/artificial-intelligence-ai-report-extinction/index.html. Acesso em: 20 maio 2024.

21 Nações Unidas, Biblioteca Digital,"Improving the effectiveness of the United Nations: role of youth in peace and security." Nova York, 2023. Disponível em https://digitallibrary.un.org/record/4040897?ln=en&v=pdf. Acesso em: 20 maio 2024.

22 Senado Federal do Brasil, Comissões. "Documentos da Comissão de Constituição, Justiça e Cidadania." Disponível em: https://legis.senado.leg.br/comissoes/arquivos?ap=8139&codcol=2629. Acesso em: 20 de maio de 2024.

Há quase quatro anos, a Unesco estabeleceu a centralidade das ferramentas de governança, como as avaliações de impacto algorítmicas. Hoje, o Grupo de Trabalho sobre Economia Digital do G20 está discutindo um conjunto de ferramentas para avaliar e mapear a IA para melhorar os serviços públicos. Localmente, a nível nacional, a Ordem Executiva de Biden, a Lei da União Europeia sobre IA e as propostas regulamentares do Canadá e do Brasil tornaram ou procuraram tornar esse tipo de avaliação obrigatória para riscos elevados de IA com escrutínio público. Um movimento progressista que procura um tipo de governança democrática local e global com governança em rede multissetorial e multidisciplinar.

Existem várias respostas de convergência locais e transnacionais que visam reduzir a assimetria de informação em jogo. Por exemplo, a recente resolução da ONU sobre IA que propõe a criação de um Comitê de IA para a Lei de Serviços Digitais (DSA) da UE, e o projeto de lei brasileiro de IA[23] que buscam requerer que pesquisas adquiridas permitam acesso a dados para uma melhor compreensão dos algoritmos. Neste sentido, deve haver interoperabilidade regulamentar com ênfase no escrutínio público e, em última análise, na deliberação social para determinar quais são os riscos aceitáveis e como maximizar os benefícios reais e não especulativos das tecnologias de IA.

A abordagem acima mencionada deve basear-se na tradição jurídica enraizada na regulação ambiental e na justiça territorial e, portanto, defendendo a inclusão das vozes afetadas e dos grupos vulneráveis nesse diálogo sobre governança. Caso contrário, não haverá uma verdadeira responsabilização devido à falta de um fórum público para avaliar se os sistemas de IA representam riscos globais e/ou locais toleráveis, bem como benefícios reais e substantivos para a sociedade.

Em conclusão, o desafio global, tanto a nível nacional como internacional, depende da escassez de evidências e de colaboração para governar a IA e não ser governado por ela. Enfrentar este desafio requer abordagens multifacetadas e ação coletiva. A proposta apresentada pelo Painel de Alto Nível sobre IA oferece um passo concreto para o estabelecimento de um órgão multilateral semelhante ao IPCC, reunindo cientistas de diversas origens para produzir conhecimento como um bem comum global no domínio da IA.

Interoperabilidade regulatória: governança de dados e uma abordagem normativa adaptativa

No contexto do Pacto Digital Global (GDC), o G20 e a ONU devem estabelecer uma posição comum sobre a governança de dados e adotar um quadrode referência para avaliar os potenciais benefícios e danos da utilização de dados, incluindo a IA. Para permitir que a transformação digital obtenha o máximo valor público possível, os

23 Barbara Prainsack et al. Data solidarity: a blueprint for governing health futures. The Lancet Digital Health, v. 4, n. 11, p. e773-e774, 2022.

instrumentos de governança devem ser desenvolvidos, implementados e monitorados através de processos inclusivos e participativos. A implementação em etapas de novas abordagens de governança – começando por setores como a saúde – ajudará a testar os seus benefícios em contextos específicos, a minimizar potenciais danos e a construir a confiança do público. O G20 tem a responsabilidade coletiva de garantir que as práticas digitais melhoram a vida de todas as pessoas e que os danos são prevenidos de forma mais eficaz. A solidariedade de dados fornece um modelo de como fazer isto acontecer e oferece um quadro para alinhar diversas abordagens de governança com um objetivo comum[24]. Um conjunto de instrumentos políticos propostos para concretizar a solidariedade de dados, bem como uma ferramenta para avaliar o valor público da utilização de dados, foi desenvolvido e poderia ser prontamente implementado em todos os países do G20.

O G20 também deve fornecer um quadro comum e recursos financeiros para a governança participativa e a conceção conjunta dessas infraestruturas, que devem ser transparentes, responsáveis, interoperáveis e, de preferência, de acesso aberto. Deve haver um entendimento comum e uma cooperação transfronteiriça entre o Norte e o Sul globais para uma maturidade eficaz da governança de dados. Uma forte colaboração entre as partes interessadas ao longo de todo o ciclo de vida da informação, através de políticas de dados abertas e de uma abordagem centrada no cidadão, é essencial para garantir que o interesse público impulsione os dados (justiça de dados) e não o contrário .

Além disso, o G20 deve propor abordagens regulamentares que sejam flexíveis, adaptáveis e holísticas relativamente a todos os componentes da governança da IA, permitindo testes rápidos e ajustamentos em resposta a efeitos inibidores, riscos emergentes e novos desafios. Ao apoiar quadros regulamentares adaptativos, o G20 pode criar um ambiente propício à inovação, assegurando ao mesmo tempo que as tecnologias de IA sejam desenvolvidas e implementadas de uma forma ética, responsável e respeitadora dos direitos humanos, avançando assim os ODS e satisfazendo as necessidades da maioria global.

O G20 deve adotar uma abordagem colaborativa e multilateral nos esforços de governança da IA, incluindo o desenvolvimento de normas para ferramentas de avaliação de riscos da indústria. O G20 deveria recomendar auditorias algorítmicas obrigatórias para sistemas de IA de alto risco.

O G20 deveria:

- apelar a uma avaliação de impacto da IA em vários níveis, abrangendo questões jurídicas e sócio-éticas[25];

24 Alessandro Mantelero, AI and Big Data: A blueprint for a human rights, social and ethical impact assessment. Computer Law & Security Review, v. 34, n. 4, p. 754-772, 2018.

25 Bruno Bioni, Ecologia: Uma Narrativa Inteligente para Proteção de Dados Pessoais em Cidades Inteligentes. TIC eGOV 2017: livro eletrônico. 2017. Disponível em: https://brunobioni.com.br/wp-content/uploads/2019/05/TIC_eGOV_2017_livro_eletronico-55-62.pdf. Acesso em: 20 maio 2024.

- definir as melhores práticas para este exercício, incluindo o papel da participação das partes interessadas na conceção conjunta de sistemas de IA;
- promover a transparência na gestão de riscos;
- aprofundar a componente jurídica e sócio-ética da avaliação, baseando-se em soluções operacionais universais e quantificação para avaliações de impacto nos direitos humanos, articulando ao mesmo tempo o papel das diferentes partes interessadas para colmatar a lacuna entre as necessidades regulamentares e a promoção da inovação.

Isto deve incluir a implementação de uma estratégia inovadora de risco-oportunidade para gerir o impacto da IA nos mercados de trabalho do Sul Global.

Além disso, é importante sublinhar que as avaliações de impacto são um meio para atingir um fim e não um fim em si mesmas. A este respeito, os acadêmicos e a sociedade civil na maior parte do mundo necessitam de recursos não só para fazer cumprir estes mecanismos, mas também para realizar investigação empírica com base nas evidências que fornecem. Consequentemente, tanto a concepção de tecnologias como a investigação conduzida para responsabilizar os criadores de tecnologias precisam de ser mais representativas e orientadas para a comunidade. Por exemplo, a maior parte do mundo precisa realmente de ter os seus próprios painéis de referência e implementar métodos qualitativos mais inclusivos que não estão sendo abrangidos pelo trabalho atual no terreno.

Conclusão

Embora a interoperabilidade técnica e regulamentar seja essencial, a sua realização deve ser abordada com cautela para evitar um ressurgimento da colonização, onde as normas e padrões técnicos globais dominantes sufocam as abordagens e requisitos locais e regionais. Em vez disso, deveríamos esforçar-nos por estabelecer mecanismos de emancipação e autodeterminação em que todas as partes interessadas contribuam equitativamente para a criação de um ecossistema que promova tanto a inovação como os direitos humanos.

Esta abordagem reconhece a diversidade de valores, culturas e interesses envolvidos na governança da IA, com o objetivo de encontrar um equilíbrio que incentive o avanço tecnológico responsável, salvaguardando simultaneamente os direitos fundamentais. As principais considerações incluem:

- a interligação entre as leis nacionais e a legislação não vinculativa internacional enraizadas nos direitos humanos e, mais especificamente, explorando a relação entre a justiça de dados e a transformação digital inclusiva. Além disso, enfatizando a natureza do conceito centrada na cidadania, tendo em conta as assimetrias históricas de poder relativas à exclusão digital já existente e às epistemologias em desfavor do Sul Global;

- avaliação de impacto algorítmica obrigatória que abranja questões jurídicas e sócio-éticas para a IA de alto risco e a implementação de outras ferramentas de governança para reduzir a assimetria de informação, a fim de estabelecer uma governança democrática com escrutínio público;
- a interação entre a regulamentação global-local e as políticas públicas internacionais-nacionais para promover IA que estimulem a justiça social e os nossos problemas sociais mais urgentes, nomeadamente, mas não limitados, à crise climática, à saúde global, às desigualdades de gênero, à integridade da informação e ao trabalho digno.

Estes valores e pilares normativos são essenciais para estabelecer uma governança da IA que não seja opressiva, mas sim emancipatória, promovendo laços de solidariedade e garantindo o desenvolvimento seguro e fiável e a implantação destas tecnologias numa perspectiva ecológica. Acima de tudo, visam prevenir a exacerbação das desigualdades e a potencial emergência do apartheid epistemológico entre o Norte e o Sul globais.