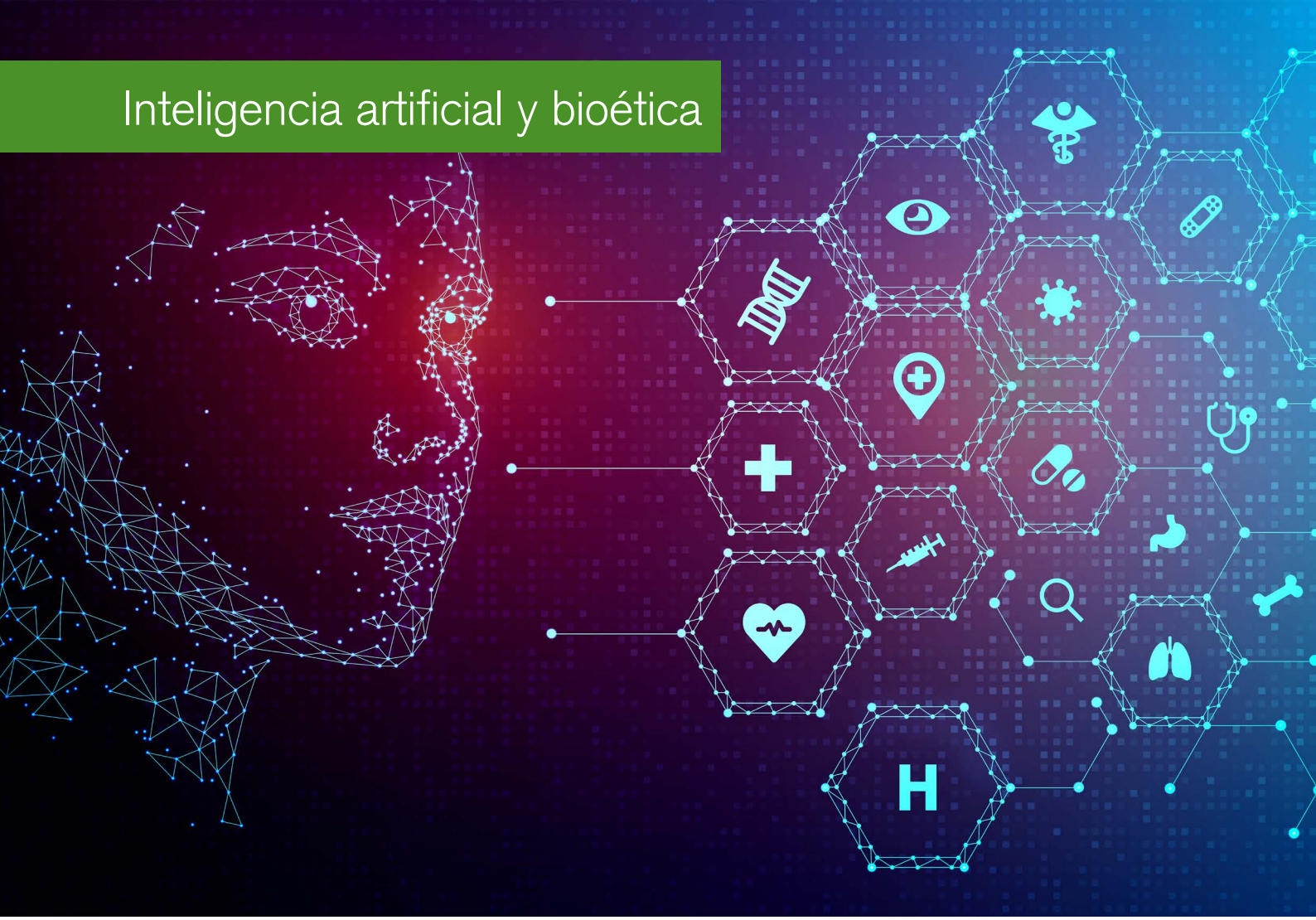

Inteligência artificial e injustiça: para além de uma análise ética encantada dos algoritmos

Artificial Intelligence and Injustice: Beyond an Enchanted Ethical Analysis of Algorithms

Autores

Leonardo Cambraia
Universidade de Brasília
Email: leoccambraia@gmail.com
https://orcid.org/0000-0002-4049-3220

Monique Pyrrho
Universidade de Brasília
Email: pyrrho.monique@gmail.com
https://orcid.org/0000-0003-1000-6361

Resumo

Não há limites para as expectativas em relação às Inteligências Artificiais (IAs). O que não é consenso é se suas promissoras aplicações, caso se realizem, contemplarão a todos. O objetivo deste trabalho é refletir sobre os desafios éticos na utilização de IA, especialmente aqueles relacionados à justiça. Para isso, após apresentar os aspectos mais comumente levantados na discussão ética em relação a IA, será discutida a maneira diversa com que tais tecnologias podem interagir com contextos e corpos periféricos, representando risco de incremento de desigualdades. A partir dessa constatação, pretende-se não apenas descrever tais riscos, mas também vislumbrar possíveis recursos para enfrentar efeitos negativos das IAs em contextos historicamente vulnerados.

Abstract

There are no limits to expectations regarding Artificial Intelligences (AIs). What there is no consensus on is whether their promising applications, if realized, will benefit everyone. The goal of this paper is to reflect on the ethical challenges of using AI, especially those related to justice. To this end, after a brief presentation of the aspects most commonly raised in ethical discussions in relation to AI, the work will discuss the diverse ways in which these technologies can interact with peripheral contexts and bodies, posing a risk of increasing inequalities. Based on this observation, the aim is not only to describe these risks, but also to look at possible resources for dealing with the negative effects of AIs in historically vulnerable contexts.

Key words

Bioética; inteligência artificial; injustiça algorítmica.

Bioethics; artificial intelligence; algorithmic injustice.

Fechas

Recibido: 07/02/2024. Aceptado: 19/05/2024

1. Introdução

As Inteligências Artificiais (IAs) atravessam o campo da saúde, desde a pesquisa até a prática clínica. Os benefícios esperados incluem: a automação e a simulação de experimentos, descoberta de novos medicamentos, avanços diagnósticos, interpretação de genomas, seleção de tratamento, monitoramento de pacientes e etc. O argumento é que, em comparação com análises humanas, aquelas realizadas por IA apresentam diferenciais em fatores como desempenho, custos e disponibilidade, o que poderia contribuir para o aumento do acesso a cuidados em saúde (Secinaro et al., 2021).

A Inteligência Artificial se origina como uma disciplina, ainda em 1956, motivada pela expectativa de que as pesquisas no campo da cognição humana tornariam todos os aspectos relativos à inteligência e aos processos de aprendizagem tão bem conhecidos, e descritos com tamanha precisão, que uma máquina seria capaz de replicá-los

Sem dúvida, o alcance dos usos dessas tecnologias já justificaria a necessidade de análises éticas. Munn (2023), contudo, argumenta sobre a inutilidade da ética em IA. A discussão, baseada na proposição de princípios, mais do que inútil, denuncia ele, é um verniz usado pelas empresas para driblar a necessidade de regulação.

Mesmo que se desenvolva com o objetivo de promover o bem-estar humano, a atividade tecnocientífica comporta riscos e com as IAs isso não é diferente. A utilização de IA é um fenômeno tecnocientífico e econômico de grande magnitude, mas também social, de dimensões culturais, epistêmicas e éticas que receberam, proporcionalmente, ainda pouca reflexão acadêmica (Cotton et al., 2023; Jobin et al., 2019).

Muito da abstrata discussão em torno das IAs tem como função operar como uma cortina de fumaça para deslocar o foco da necessidade premente de regulamentação das *Big Techs* (Munn, 2023) ou da concretude com que corpos periféricos já sofrem consequências negativas das IAs (Mohamed et al., 2020). Propomos aqui uma abordagem ética centrada na preocupação com os efeitos das IAs nas configurações de poder e injustiça. Para tanto, argumenta-se pela necessidade de uma análise crítica de muitos dos pressupostos que sustentam a compreensão tanto do que são as IAs, quanto do que é uma discussão ética a respeito delas.

1.1. De que IA estamos falando?

A Inteligência Artificial se origina como uma disciplina, ainda em 1956, motivada pela expectativa de que as pesquisas no campo da cognição humana tornariam todos os aspectos relativos à inteligência e aos processos de aprendizagem tão bem conhecidos, e descritos com tamanha precisão, que uma máquina seria capaz de replicá-los (Dick, 2019). No entanto, as IAs atuais não funcionam bem assim. Se no começo os esforços da disciplina estavam voltados para fazer uma máquina que pensasse como um humano, hoje as abordagens pretendem ultrapassar as limitações humanas quanto

à capacidade de processamento de informação. Trata-se, assim, de fazer o que é humanamente impossível (Dick, 2019).

O genérico nome "Inteligência Artificial" na verdade é um guarda-chuva que abriga algoritmos codificados em linguagens de programação diversas, com complexidades e finalidades muito variadas. O imbróglio é tão grande que o livro "Inteligência Artificial", escrito por Russell e Norvig (2013), considerado uma referência introdutória para o campo, não consegue oferecer uma definição suficientemente clara e unívoca de uma tecnologia que promete transformar nossas vidas nos próximos anos. Ainda que o principal interesse no uso de algoritmos seja aquele de processar uma quantidade humanamente impeditiva de dados, a execução de um tipo de atividade similar à cognição humana ainda é identificada como atributo central nas explicações sobre o campo.

Os recentes desenvolvimentos das inteligências do tipo LLM (*Large Language Models*, ou grandes modelos de linguagem, em português) levou à expectativa, inclusive por parte de alguns desenvolvedores, de que as IAs têm ou estão perto de adquirir consciência

Tecnicamente, IAs são algoritmos, ou seja, sequências de instruções. A particularidade daquilo que chamamos de IA se dá porque estas, ao serem alimentadas por um conjunto de dados, possuem a capacidade de analisar e identificar relações significativas entre dados, extraindo automaticamente informações sobre associações e tendências desconhecidas, para então, a partir dos padrões apreendidos, performar uma função a qual foram destinadas desde seu desenvolvimento. Em outras palavras, em vez de desempenhar uma tarefa com um código previamente fornecido, como faria uma calculadora simples, por exemplo, uma IA realiza a tarefa para a qual foi desenvolvida através de uma solução que ela mesma desenvolveu a partir do que aprendeu com os dados que lhe foram fornecidos. Ou seja, se uma IA é feita para gerar textos, ela, analisando textos, aprenderá a gerar textos; se ela é construída para gerar imagens, ela, a partir da análise de um banco de imagens, irá gerar imagens; e, assim por diante.

Os recentes desenvolvimentos das inteligências do tipo LLM (*Large Language Models*, ou grandes modelos de linguagem, em português) levou à expectativa, inclusive por parte de alguns desenvolvedores, de que as IAs têm ou estão perto de adquirir consciência. Embora o diálogo humano simulado seja um feito impressionante e crescentemente verossímil quando comparado à linguagem humana, os geradores de textos apenas fazem a tarefa para a qual são concebidas, reproduzindo padrões encontrados nos bancos de dados.

Iniciar a discussão acerca dos riscos das IAs pressupõe como primeira distinção importante a diferença entre o que se tem chamado de IAs gerais e estreitas. As IAs estreitas, também chamadas de IAs fracas, utilizam algoritmos de aprendizado profundo para analisar grandes volumes de dados e fazer predições de comportamento em tarefas específicas. A consequência é que essas IAs não desempenham em outros contextos a não ser aqueles para os quais tenham sido treinadas. Já as IAs gerais, também chamadas de IAs fortes, teriam outro nível de inteligência que, por sua vez, permitiria atingir objetivos em diferentes ambientes e contextos. Com isso, uma IA geral não seria voltada para tarefas específicas, seria autônoma e poderia aprender de

forma não supervisionada nos mais diversos âmbitos, se aproximando à consciência humana (McLean et al., 2023).

De alguma forma, a plurivalência e a complexidade de associações e habilidades são características constituintes da inteligência humana. Para adquirir consciência, no entanto, as IAs precisariam bem mais do que se tornar multitarefas. A consciência e agência, tais quais adquirimos como espécie, no entanto, dependem de muitos outros fatores. Elas foram construídas ao longo de longo processo evolutivo, em toda sua aleatoriedade de desafios; dependem de nosso estar, contatar e compreender sensorialmente o mundo; do multifatorial processo de construção de memórias individuais, que não envolve apenas a cognição; de estruturas neurais com trajetórias recíprocas de informação; de nossa complexa organização pluricelular; e de nossa existência coletiva e interação com outros agentes conscientes (Aru et al., 2023).

De alguma forma, a plurivalência e a complexidade de associações e habilidades são características constituintes da inteligência humana. Para adquirir consciência, no entanto, as IAs precisariam bem mais do que se tornar multitarefas

Sem dúvida, a presença ou ausência de atributos como a consciência, é moralmente muito significativa. Atualmente, IAs não possuem sequer a habilidade de desenvolver múltiplas tarefas. Seu desenvolvimento não conta com quaisquer desses atributos e trajetórias que conformaram a consciência humana (Aru et al., 2023). Os exercícios especulativos gozam de prestígio nas análises de novas tecnologias, mas interessa mais pensar nos efeitos do real alcance e capacidade das IAs.

O segundo aspecto significativo é referente aos pressupostos epistêmicos, nem sempre verdadeiros ou ao menos quase nunca verificáveis, que guiam as avaliações éticas quanto às IAs. A esperança é que o conhecimento sobre todas as áreas estaria de alguma forma oculto na avalanche de dados que, graças às máquinas, fomos capazes de coletar. A expectativa é que sejam também as máquinas aquelas capazes de identificar, naquilo que nos parece um caótico, incompreensível e gigantesco mar de dados, as informações relevantes. Serão as máquinas a traduzir esses dados em informações e as informações, em conhecimento, nisso estão todas as fichas do ramo de IAs, e as promessas aqui são infinitas.

Se quanto mais dados, melhor, qualquer impedimento em relação a uma sempre mais pervasiva coleta de dados é considerado um empecilho para o progresso. O problema, no entanto, é que a limitação do conhecimento humano não se extingue com a abundância de dados. Ao contrário, a capacidade de reconhecer quais relações e informações não são pertinentes e precisam ser desconsideradas pode ser até mais importante para o conhecimento do que a habilidade de acumulá-las. É preciso se questionar se são sempre mais dados aquilo de que precisamos. É necessário indagar se durante a atenção a um infarto do miocárdio é assim tão relevante o número de passos que o paciente deu naquele dia, uma quantidade enorme de informações, ou se aquilo que é preciso material e equipamentos para cuidados críticos em saúde (Neff, 2013).

Cientistas não leem todos os artigos científicos, não somente porque não podem, mas simplesmente porque nem tudo que é escrito e publicado é igualmente válido e merece atenção. Saber quais deles ler é chave importante para produzir ciência de qualidade.

Da mesma forma, a opção por reforçar o que pensa a maioria também pode não ser a resposta. Somente identificar e repetir o que é dito pela maioria dos artigos e textos científicos minaria a capacidade de inovar, e no passado teria impedido as grandes revoluções científicas.

O acúmulo do conhecimento nos possibilita elaborar uma hipótese ou tese a ser testada, quer empiricamente, quer no embate com pares. O conhecimento depende da interação entre observador e observado (quer sujeito, quer objeto), depende da cultura em seu tempo e suas relações de poder. Um conjunto maior de dados e uma capacidade maior de analisá-los não substituem a capacidade de criar hipóteses, testá-las, e julgar quais variáveis são relevantes. O que é mais crucial, porém, e segue inquestionado, é que a expectativa de lidar com uma máquina preditiva ou prescritiva que recebe um *input* e gera um *output* sem que seus processos possam ser averiguados é muito problemático. Ainda que abandonemos às máquinas a tarefa de conhecer, não conhecemos como elas produzem esse conhecimento, não somos capazes de identificar erros em seus funcionamentos, vieses em suas decisões. Não somos capazes de prever seus comportamentos. Não somente porque muitos desses algoritmos são propriedade intelectual protegida, mas porque, frequentemente, nem os desenvolvedores têm controle sobre os processos de aprendizado de máquina e de decisão. Em um mundo em que IAs que interagem e alteram comportamentos humanos passam a interagir com outras IAs que também fazem isso, como ocorre no mercado financeiro, por exemplo, não saber como as máquinas operam enquanto guiam nossos passos e esperanças por um mundo melhor e de mais conhecimento não é só temerário (Rahwan et al., 2019), assemelha-se perigosamente a um pensamento mágico.

A esperança/crença/ilusão é a de que estamos diante de uma inteligência superior, como a de oráculos ou alguma divindade, de onde emana a verdade, um mistério absoluto e inquestionável

A esperança/crença/ilusão é a de que estamos diante de uma inteligência superior, como a de oráculos ou alguma divindade, de onde emana a verdade, um mistério absoluto e inquestionável (Floridi, 2015; McArthur, 2023a; McArthur, 2023b).

Para performar uma análise ética dessas novas tecnologias, portanto, é preciso desafiar esses pressupostos.

1.2. De que ética estamos falando?

As análises éticas têm como objetivo identificar os posicionamentos e argumentos mobilizados em torno de objetos de interesse — frequentemente novas tecnologias com aplicação no campo da saúde, quando falamos de bioética — de modo a compreendê-los em sua complexidade, auxiliando e informando assim o processo de tomada de decisões (Hottois, 2020).

Sobretudo, a bioética surge como campo num momento em que o desenvolvimento da ciência pareceu vertiginoso e ameaçador. Na segunda metade do Século XX, já havia ficado claro que a busca por conhecimento não podia ser feita em detrimento dos indivíduos. Que o progresso da ciência e todo o benefício esperado para a humanidade

encontrava seu limite no consentimento, na autonomia e no bem-estar dos sujeitos. De forma mais grave, os sujeitos mais vulneráveis eram aqueles mais vitimizados por abusos nas práticas científicas e, por isso, era o respeito a eles e elas as medidas para a efetividade das proteções e garantias. Por esse motivo, as análises bioéticas no Brasil, e na América Latina, em geral, assumem uma preocupação especial com as iniquidades relativas ao acesso à saúde e com a distribuição justa dos benefícios proporcionados pelo desenvolvimento tecnocientífico. Mais do que isso, o escopo alargado da disciplina comporta trocas entre ciências da saúde e da vida, ciências humanas e sociais e pauta novas tecnologias, mas relacionando-as às dimensões sociais, ambientais presentes e futuras (Garrafa, 2022).

A dimensão dos riscos e o descompasso entre a velocidade do desenvolvimento tecnológico e das pesquisas acerca das dimensões éticas das IAs evidenciam a importância de maior investimento acadêmico para a compreensão dessas tecnologias e seus impactos

Para elaborarmos princípios, normas e mecanismos de fiscalização necessários para guiar eticamente as IAs, é antes de tudo necessário compreender o que são, quais são suas características, operações e os conteúdos, suas finalidades pretendidas e a fonte de informações usadas para seus treinamentos (Siau e Wang, 2020). Diante das tendências crescentes de um campo da saúde cada vez mais baseado em dados e algoritmos treinados com aprendizado de máquina, a discussão bioética torna-se indispensável. A manipulação de informações e do comportamento de populações apresentam uma diversidade de riscos, apresentando-se, inclusive, como novos fatores de adoecimento em uma sociedade digital (Pyrrho et al., 2022a).

No geral, a promessa é que as IAs reduzam o trabalho humano. Entretanto, é preciso se perguntar se essa redução, promessa que acompanhou o surgimento de grande parte das novas tecnologias, é verdade para todos os tipos de trabalho. Mais do que não atuar em todos os campos das atividades humanas — o que nos convida a pensar sobre os campos nos quais se quer desempregar humanos — é preciso imaginar campos nos quais o trabalho, mesmo árduo, continua sendo manual e insalubre.

A dimensão dos riscos e o descompasso entre a velocidade do desenvolvimento tecnológico e das pesquisas acerca das dimensões éticas das IAs evidenciam a importância de maior investimento acadêmico para a compreensão dessas tecnologias e seus impactos.

De tal forma, o objetivo deste trabalho é desenvolver uma reflexão a respeito dos desafios éticos na utilização de IAs, especialmente aqueles relacionados à justiça. Após breve síntese dos aspectos éticos mais comumente associados à coleta e uso de dados, o texto aponta os riscos de que a utilização de IA, conforme é concebida, perpetue e intensifique desigualdades já existentes. Reflete-se então acerca dos resultantes dessas novas tecnologias em contextos e corpos periféricos. Por fim, são apresentados exemplos de estratégias já utilizadas para mitigar alguns desses desafios.

2. Dimensões éticas de uma tecnologia baseada em dados

Longe de ser uma revisão exaustiva, até porque as tecnologias continuam impondo novos desafios, o texto aponta em uma divisão esquemática os dois principais focos a serem abordados em uma reflexão ética sobre o uso de dados por IAs. O primeiro deles concerne a origem de informações fornecidas aos algoritmos.

A vigilância e a coleta massiva de dados para alimentação de IAs podem ser apontados como os mais elementares riscos associados ao desenvolvimento desse tipo de tecnologia. Isso se dá, porque tudo, incluindo pesquisas científicas, dados recolhidos por órgãos públicos, posts públicos, mensagens, e-mails, fotos capturadas e armazenadas por celulares, repositórios de documentos na nuvem, atividades bancárias, registros médicos, registros fotográficos por câmeras de segurança em cada esquina, e até mesmo rastros digitais como os deixados ao nos comunicarmos com uma assistente digital pessoal, ou seja, qualquer registro digital de existência é fonte para a agregação de dados.

Os rastros digitais resultantes das mais diversas atividades humanas, inclusive aqueles referentes à privacidade dos indivíduos, quando expropriados e convertidos em agregados massivos de dados, têm se tornado uma das *commodities* mais amplamente negociadas no mundo.

Aqui se encontra uma produção de valor não consentida e não remunerada. Os rastros digitais resultantes das mais diversas atividades humanas, inclusive aqueles referentes à privacidade dos indivíduos, quando expropriados e convertidos em agregados massivos de dados, têm se tornado uma das *commodities* mais amplamente negociadas no mundo.

Enquanto somos levados a crer que a privacidade não tem mais valor e que o melhor a fazer é compartilhar informações íntimas, médicas até, para o bem comum, sendo convencidos a tratá-las como *bens comuns*, a realidade é que elas têm grande valor, como o próprio termo "sociedade da informação" deveria nos fazer desconfiar. Este valor é expropriado e comercializado, se tornando a origem de todas as grandes fortunas geradas neste milênio (Zuboff, 2019). Contrastando com discursos que são comunitários somente em aparência, as finalidades e formas de obtenção são mantidas como segredos comerciais e as intimidades de indivíduos vendidas a granel. A concentração de poder político e econômico aqui não é em nada negligenciável. Não é à toa que tudo isso ocorre diante de Estados nacionais impotentes na tarefa de controlar e regulamentar a atividade das *Big Techs*, mesmo quando elas interferem em processos democráticos em diversos graus e em todo o globo (Zuboff, 2019; Pyrrho et al., 2022a; Pyrrho et al., 2022b).

A segunda dimensão se refere à aplicação desses dados. Aqui, as preocupações e desafios são relacionados com a opacidade e falta de auditabilidade das operações algorítmicas, impactos no mundo do trabalho, desinformação e manipulação de comportamentos (Cabitza et al., 2017; Vayena et al., 2018; Jobin et al., 2019). Isso sem mencionar a dimensão mesma do misto de encanto e espanto diante das novas tecnologias generativas, das quais as produtoras de texto talvez sejam as mais emblemáticas. Um exemplo disso foi que no começo, e ao longo do último ano, os

noticiários, por meses a fio, só falavam do ChatGPT. Escrever poemas e textos científicos, mas também pintar e desenhar, mesmo atuar, compor músicas ou peças para teatro, tudo isso agora é coisa de máquina. Em tempos que existem imagens realistas de pessoas que nunca existiram; nascer, viver e produzir enquanto humano parece estar relegado à obsolescência. Essa, porém, é a discussão mais abstrata. Concretamente, grupos historicamente vulnerabilizados, como as mulheres, passam a ter seus corpos objetificados e sexualizados em montagens realistas. Novos meios, velhas injustiças.

Concretamente, grupos historicamente vulnerabilizados, como as mulheres, passam a ter seus corpos objetificados e sexualizados em montagens realistas. Novos meios, velhas injustiças

A discussão ética é desenvolvida tendo como premissa majoritariamente indiscutida a maneira como essas tecnologias impactarão um ser humano universal. Tudo se passa como se o fenômeno ameaçasse indistintamente, um horizonte distópico que nos espera a todos igualmente na próxima esquina. Não é bem assim.

3. Desigualdade, Injustiça e IA

Além dos mais frequentemente discutidos problemas referentes aos dados, novos desafios surgem como a injustiça algorítmica, termo que descreve a compreensão de que, ao reproduzir preconceitos e vieses, algoritmos não somente atuam de forma inadequada ou insuficiente, mas são agentes perpetuadores de injustiça, fazendo com que aqueles mais vulneráveis sejam afetados de forma desproporcional por seu uso crescente (Birhane, 2021).

Para abordar desigualdade e IA é necessário observar ao menos três características dessa tecnologia, por vezes sobrepostas, que contribuem com a perpetuação ou incremento de danos a grupos historicamente vulnerabilizados: representatividade dos dados, viés algorítmico e opacidade tecnológica.

No que diz respeito à representatividade dos dados, o treinamento dos algoritmos de IA podem refletir preconceitos e vieses derivados das formas de construir os bancos de dados (Mittelstadt e Floridi, 2016). O problema não é apenas a existência de vieses embutidos, mas que isso acontece enquanto as IAs se apresentam como um remédio numérico à limitação da cognição e subjetividade humanas. A neutralidade e a objetividade atribuídas aos algoritmos são alguns dos mais fortes obstáculos para o combate a vieses (Buolamwini e Gebru, 2018). Se os dados utilizados como parâmetros clínicos, por exemplo, sub-representam gênero ou grupos étnicos, as decisões automatizadas em escala tendem a errar mais sobre esses corpos. Já foram relatados, por exemplo, casos em que a automatização empregada ao diagnóstico de hipóxia por oxímetro negligenciou padrões de mensuração diversos, relacionados a absorção de luz em interação com a melanina, e subestimou o grau de hipóxia dos pacientes negros, elevando os riscos das sequelas por falta de oxigenação (Sjoding et al., 2020). Outro estudo mostrou que o uso de reconhecimento facial por veículos

automotivos autônomos tende a falhar proporcionalmente mais na identificação de pedestres negros e de tons mais escuros de pele, o que ameaça incrementar ainda mais a já maior incidência de agravos por atropelamento nesse grupo da população (Wilson et al., 2019). Ainda, algoritmos utilizados para análise de currículos com fins de seleção de profissionais preterem mulheres, mesmo quando são programados para não considerar a variável gênero (Dastin, 2018). A realidade a ser descrita digitalmente é plena de vieses e dita qual informação tem valor e como coletá-la. Os dados irrevogavelmente refletirão esse quadro. Isso é tão significativo que em casos como o anterior, como o da seleção de currículos, os algoritmos resistem às tentativas de eliminar vieses. Se as prescrições e predições algorítmicas se baseiam em parâmetros usados até hoje, instruções simples para evitar critérios preconceituosos serão dribladas e as posições de poder e privilégio serão mantidas por outros cálculos. A ingenuidade é acreditar que *prompts* simples poderiam resolver o sexismo.

Os efeitos desses processos de concentração de poder, da operação de algoritmos com vieses embutidos e a modulação de comportamentos em contextos periféricos ainda não foram suficientemente investigados

Esses preconceitos e injustiças são exacerbados pela opacidade digital, ou seja, pela ausência de explicabilidade e rastreabilidade dos processos com os quais as máquinas geram os *outputs*. Não é dado a saber que processos e operações levaram à prescrição ou decisão oferecida pela máquina, muitas vezes se desconhece inclusive com base em quais dados (Dalton-Brown, 2020). Do ponto de vista ético, quem seria responsável pelas decisões dessas máquinas e por suas consequências?

Os efeitos desses processos de concentração de poder, da operação de algoritmos com vieses embutidos e a modulação de comportamentos em contextos periféricos ainda não foram suficientemente investigados.

Países e corpos periféricos são afetados de forma ainda mais intensa por essas tecnologias. Além de efeitos universalmente distribuídos, o reforçamento das injustiças diárias, a concentração de poder e aspectos específicos relacionados a situações de vulnerabilidade podem agir como obstáculo à mera compreensão dessas novas tecnologias e seus riscos, o que significa que estes corpos são menos aptos a se protegerem (Cambraia et al., 2023).

O que dizer então de sua capacidade de produzir tecnologia, fazer dela algo útil e realmente representativo? A sujeitos periféricos só é consentido desejar consumi-las, almejando esse status supostamente universal de civilizados digitais. As tecnologias são desenvolvidas de forma a mascarar esse imperialismo moral. Tudo se opera de maneira similar à descrita por Quijano (2000), em que aos corpos periféricos só é oferecido maneiras inadequadas de se enxergar, trata-se de um espelho que só oferece um reflexo distorcido, feio e inadequado de nós. Aqui a tecnologia não faz mais do que produzir um desejo de não coincidir com a própria pele, nos fazer querer ser outra coisa, encaixarmo-nos em outros padrões. Novos meios, mesmas injustiças.

Enquanto isso, os minerais necessários para a produção dos dispositivos são explorados em países periféricos. Tirar essa matéria prima da terra, cansa e adoece os

corpos de quem vive ali, corrói as condições ambientais, mas também investe corpos de toda a sociedade periférica nas tramas políticas e guerras civis desenvolvidas ao redor desse interesse econômico (Faustino e Lippold, 2022).

A maior parte das análises éticas se desenvolve como se todos fossem igualmente explorados e expostos ao risco, quando, na verdade, as diferenças são marcantes. Os algoritmos exponenciam vieses humanos, incrementam a injusta divisão internacional do trabalho e a distribuição do impacto ambiental do uso de recursos naturais, ao mesmo tempo, que por serem baseados em números, numa fetichização incrementada da tecnologia, guardam um verniz brilhante de neutralidade.

Quando o assunto é o uso de algoritmos em saúde coletiva, há ainda mais desafios. Adotados largamente em planos de saúde nos EUA, algoritmos usam parâmetros como o presente uso de recursos terapêuticos de cada paciente para determinar o direcionamento futuro de atenção à saúde

Talvez não se trate de decidir se é mais grave o fim da privacidade ou a obsolescência humana diante de máquinas mais inteligentes. É preciso atentar para como a utilização de IA incrementa e reinveste as desigualdades por meio da concentração de poder, das alterações do trabalho humano, e do acobertamento e neutralização tecno-numérica de persistentes vieses e preconceitos. É nesse momento que a pausa é necessária para darmos conta do perigo do uso de uma automação que faria crescer exponencialmente os impactos do uso de pressupostos falsos e racistas em saúde por IAs, por exemplo.

Pressupostos como o humano universal da indústria farmacêutica, que na realidade é um homem branco de 70 quilos, ou aquele de que as pessoas negras têm maior massa muscular, assunção falsa que pode decorrer em danos renais a depender do tratamento (Cerdeña et al., 2020; Khazanchi et al., 2023) informam intervenções clínicas há muito tempo. A automatização em grande escala desses padrões racistas em saúde é assustadora (Birhane, 2021).

Quando o assunto é o uso de algoritmos em saúde coletiva, há ainda mais desafios. Adotados largamente em planos de saúde nos EUA, algoritmos usam parâmetros como o presente uso de recursos terapêuticos de cada paciente para determinar o direcionamento futuro de atenção à saúde. Tal mecanismo tem como objetivo a prevenção e a redução de custos, e parte de uma pressuposta relação direta entre a gravidade da doença e seu custo. Maior o uso de recurso, maior a gravidade, maior necessidade de prevenção, menores os custos futuros: essa é a lógica. O que se esquece de dizer é que, com a mesma gravidade clínica, um paciente latino ou negro nos EUA recebe menor investimento clínico, quer porque tem pouco acesso à saúde, ou porque, mesmo quando tem acesso, não conta com a mesma empatia e atenção dos profissionais de saúde. A prática considerada objetiva, neutra e eficiente do ponto de vista econômico, porque foi prescrita por algoritmos, nada mais é do que o velho "dar mais a quem tem mais" (Obermeyer et al., 2019).

Já existem soluções voltadas para práticas mais justas nas aplicações de IAs. Gebru e colaboradores (2021), reconhecendo que é improvável que uma IA seja adequada do ponto de vista ético sem que os dados que a alimentam o sejam, propõem que os agregados de dados sejam acompanhados por uma ficha técnica (*datasheet)*.

Essa conteria informações quanto à motivação, composição e representatividade dos dados, processo de coleta, processamento, limpeza, classificação, usos pretendidos e consentidos, distribuição e manutenção. Para construir essa ficha, os autores propuseram um questionário com perguntas para que, durante a criação dos mecanismos de coleta de dados, os desenvolvedores reflitam sobre esses requisitos, contemplando-os em seus bancos. A ficha pode ser considerada como um atestado de qualidade daqueles dados, uma espécie de selo para garantir que qualquer IA que vier a utilizá-los apresente certo grau de transparência, rastreabilidade e representatividade.

Silva (2020) sintetizou em categorias as diversas práticas digitais de "microagressões" raciais em mídias digitais. Ofensas verbais, comportamentais e ambientais comuns, intencionais ou não, que comunicam desrespeito e racismo são categorizadas. Nomear e classificar as práticas têm uma importante dimensão de denúncia, conscientização e educação. Categorias como "suposição de criminalidade", "suposição de inferioridade intelectual", "exotização", dentre outros expõem os problemas em sua concretude e permitem sua mais fácil identificação (Silva, 2020), inclusive pelos desenvolvedores que criam a tecnologia e não pensam cuidadosamente sobre o racismo sistêmico, perpetuando essa e outras formas de injustiça algorítmica (Benjamin, 2019).

4. Conclusão

É indispensável revisar expectativas e pressupostos epistêmicos que baseiam nossa avaliação ética das IA. Isso é necessário para ponderar a relação entre os riscos à privacidade e os benefícios esperados. Porém, isso pode não ser suficiente.

É premente pensar como essas tecnologias interferem nas dinâmicas de poder e justiça. Hoje, a principal frente para mitigar injustiças tem sido o esforço de inclusão e incremento de representatividade de grupos demográficos marginalizados nas bases de dados, o que é muito importante, mas ainda insuficiente. Apenas passar a coletar dados de vulneráveis, sem imaginar maneiras de distribuir benefícios e diminuir desigualdades, seria só mais uma prática exploratória.

A expectativa de que as IAs trarão automaticamente a verdade e a justiça em saúde, ou em qualquer outra área, não somente é irreal, mas contribui para a perpetuação da desigualdade

Compreender o funcionamento das tecnologias, mas também as formas com que elas podem potencializar a injustiça, é um passo para construção de mecanismos efetivos destinados a expor e combater vieses e preconceitos. Formas mais conscientes e rastreáveis de construção de bancos de dados garantem não apenas o consentimento e o uso adequado dos dados, mas também a representatividade da diversidade humana. Categorizar preconceitos, auxiliando a conscientização, educação e defesa de grupos marginalizados também é um passo inicial para caminharmos na direção de tecnologias digitais menos danosas.

A expectativa de que as IAs trarão automaticamente a verdade e a justiça em saúde, ou em qualquer outra área, não somente é irreal, mas contribui para a perpetuação da desigualdade.

Referências

Aru, J., Larkum, M. E., & Shine, J. M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. *Trends in Neurosciences, 46*(12), 1008-1017. https://doi.org/10.1016/j.tins.2023.09.009

Benjamin, R. (2019). Assessing risk, automating racism. *Science, 366*(6464), 421-422. https://doi.org/10.1126/science.aaz3873

Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns, 2*(2), 1-9. https://doi.org/10.1016/j.patter.2021.100205

Buolamwini, J., & Gebru, T. (2018). Gender shades: intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. *Proceedings of Machine Learning Research, 81*, 77-91. https://proceedings.mlr.press/v81/buolamwini18a.html

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *The Journal of the American Medical Association, 318*(6), 517-518. https://doi.org/10.1001/jama.2017.7797

Cambraia, L., Pyrrho, M., & Manchola-Castillo, C. (2023). *Big Data e saúde: uma análise bioética*. Teseo Press. https://doi.org/10.55778/ts911693147

Cerdeña, J. P., Plaisime, M. V., & Tsai, J. (2020). From race-based to race-conscious medicine: how anti-racist uprisings call us to act. *The Lancet, 396*(10257), 1125-1128. https://doi.org/10.1016/S0140-6736(20)32076-6

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Dalton-Brown, S. (2020). The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics, 29*(1), 115-121. https://doi.org/10.1017/S0963180119000847

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/idUSKCN1MK0AG/

Dick, S. (2019). Artificial Intelligence. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.92fe150c

Faustino, D., & Lippold, W. (2023). *Colonialismo digital: por uma crítica hacker-fanoniana*. Boitempo Editorial.

Floridi, L. (2015). Singularitarians, aitheists, and why the problem with artificial intelligence is HAL (humanity at large), not HAL. *Philosophy and computers, 14*(2), 8-11.

Garrafa, V. (2022). Bioética y transdisciplinariedad como puentes de diálogo entre las ciencias de la salud, las ciencias sociales y/o humanas en el contexto de la evaluación ética de investigaciones. *Salud colectiva, 18*, e4177. https://doi.org/10.18294/sc.2022.4177

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86-92. https://doi.org/10.1145/3458723

Hottois, G. (2020). *¿Qué es la bioética?* Universidad del Bosque.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence, 1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2

Khazanchi, R., Soled, D. R., & Yearby, R. (2023). Racism-conscious praxis: a framework to materialize anti-oppression in medicine, public health, and health policy. *The American Journal of Bioethics, 23*(4), 31-34. https://doi.org/10.1080/15265161.2023.2186521

McArthur, N. (2023a). *AI worship as a new form of religion*. PhilPapers. https://philarchive.org/rec/MCAAWA

McArthur, N. (2023b, March 15). Gods in the machine? The rise of Artificial Intelligence may result in new religions. *The Conversation.* https://theconversation.com/gods-in-the-machine-the-rise-of-artificial-intelligence-may-result-in-new-religions-201068

McLean, S., Read, G. J., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with Artificial General Intelligence: a systematic review. *Journal of Experimental & Theoretical Artificial Intelligence, 35*(5), 649-663. https://doi.org/10.1080/0952813X.2021.1964003

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of Big Data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics, 22*(2), 303-341.

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*, 659-684. https://doi.org/10.1007/s13347-020-00405-8

Munn, L. (2023). The uselessness of AI ethics. *AI and Ethics, 3*, 869-877. https://doi.org/10.1007/s43681-022-00209-w

Neff, G. (2013). Why Big Data won't cure us. *Big Data, 1*(3), 117-123. https://doi.org/10.1089/big.2013.0029

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Pyrrho, M., Cambraia, L., & de Vasconcelos, V. F. (2022a). Privacy and health practices in the digital age. *The American Journal of Bioethics, 22*(7), 50-59. https://doi.org/10.1080/15265161.2022.2040648

Pyrrho, M., Cambraia, L., & de Vasconcelos, V. F. (2022b). Response to open peer commentaries on "Privacy and health practices in the digital age". *The American Journal of Bioethics, 22*(12), W5-W8. https://doi.org/10.1080/15265161.2022.2127972

Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. Em E. Lander (ed.), *La colonialidad del saber: eurocentrismo y ciencias sociales. Perspectivas Latinoamericanas* (pp. 118). CLACSO.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... & Wellman, M. (2019). Machine behaviour. *Nature, 568*(7753), 477-486. https://doi.org/10.1038/s41586-019-1138-y

Russell, S., & Norvig, P. (2013). *Inteligência Artificial.* Elsevier.

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of Artificial Intelligence in healthcare: a structured literature review. *BMC medical informatics and decision making, 21*, 125. https://doi.org/10.1186/s12911-021-01488-9

Siau, K., & Wang, W. (2020). Artificial Intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74-87. https://doi.org/10.4018/JDM.2020040105

Silva, T. (2020). Racismo algorítmico em plataformas digitais: microagressões e discriminação em código. Em T. Silva (ed.), *Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos* (pp. 121-135). LiteraRUA.

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine, 383*(25), 2477-2478. https://doi.org/10.1056/NEJMc2029240

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. *PLoS medicine, 15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive inequity in object detection. arXiv preprint arXiv:1902.11097. https://doi.org/10.48550/arXiv.1902.11097

Zuboff, S. (2019). *A era do capitalismo de vigilância*. Intrínseca.