

A INTELIGÊNCIA ARTIFICIAL SOB A ÉGIDE DA ÉTICA

ARTIFICIAL INTELLIGENCE UNDER THE AEGIS OF ETHICS

Lucia Santaella¹

<https://orcid.org/0000-0002-0681-6073>

RESUMO

A questão da ética tornou-se inseparável da inteligência artificial. Entretanto, a inteligência artificial não pode ser tratada como um monolito, mas como um campo heteróclito, especialmente depois do advento da inteligência artificial Generativa, que tem provocado um grande impacto nas sociedades em suas mais variadas esferas. Este artigo objetiva chamar a atenção para as diferenças entre a inteligência artificial preditiva e a inteligência artificial generativa, tendo em vista evidenciar que essas diferenças afetam o modo como a ética deve ser acionada em cada um desses sistemas.

Palavras-chave: inteligência artificial preditiva; inteligência artificial generativa; riscos; ética.

ABSTRACT

The issue of ethics has become inseparable from artificial intelligence. However, Artificial Intelligence cannot be treated as a monolith, but as a heterogeneous field, especially after the advent of Generative Artificial Intelligence, which has had a great impact on societies in their most varied spheres. This article aims to draw attention to the differences between predictive Artificial Intelligence and generative Artificial Intelligence, with a view to highlighting that these differences affect the way in which ethics should be activated in each of these systems.

Keywords: Predictive Artificial Intelligence; Generative Artificial Intelligence; risks; ethics.

¹ Pesquisadora 1A do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e livre-docente em Ciências da Comunicação pela USP. E-mail: lbraga@pucsp.br

RESUMEN

La cuestión de la ética se ha vuelto inseparable de la inteligência artificial. Sin embargo, la inteligência artificial no puede tratarse como un monolito, sino como un campo heterogéneo, especialmente tras la llegada de la inteligência artificial Generativa, que ha tenido un gran impacto en las sociedades en sus más variados ámbitos. Este artículo tiene como objetivo llamar la atención sobre las diferencias entre la inteligência artificial predictiva y la inteligência artificial generativa, con el fin de resaltar que estas diferencias afectan la forma en que se debe activar la ética en cada uno de estos sistemas.

Palabras clave: inteligência artificial Predictiva; inteligência artificial Generativa; riesgos; ética.

INTRODUÇÃO

Todas as tecnologias, melhor dizendo, próteses, que estendem capacidades ou habilidades humanas, são ambivalentes. Trazem benefícios, mas também riscos em um jogo de ganhos e perdas. Que nos sirvam de lição os diálogos platônicos com seus lamentos acerca da perda da memória provocada pela implantação do alfabeto escrito. De fato, as memórias individuais fragilizam-se quando podem ser transferidas para um dispositivo externo. Entretanto, se há perdas em nível individual, há ganhos no plano coletivo. Veio disso, no caso dos gregos, a chance de podermos ler esses filósofos até hoje.

A lição é importante como transferência para nos darmos conta das ambiguidades, ambivalências e contradições tecnológicas no contexto em que vivemos hoje. Conforme discutido anteriormente (Santaella, 2022, p. 288-290), em uma de suas avaliações preditivas, Flusser (1999, p. 204) afirmava que as memórias eletrônicas são simulações de algumas das funções do nosso cérebro. “Uma simulação é uma imitação em que alguns aspectos são exagerados, outros ignorados. [...] As memórias eletrônicas exageram e projetam algumas das funções cerebrais para fora do cérebro em objetos inanimados de modo a serem observados e manipulados”. Assim, essas invenções aperfeiçoam grandemente as funções da memória humana. Mas há uma consequência não intencional: permitir atingirmos uma distância crítica dessas funções. “Não resta dúvida de que o aperfeiçoamento das funções da memória irá transformar a cultura no futuro. Contudo, serão as consequências não intencionais da simulação da memória que irão provocar as maiores mudanças” (Flusser, 1999, p. 204).

Está prognosticada aí justamente a experiência que nos força a constatar que as memórias eletrônicas nos colocam face a face com a aquisição de informação, armazenamento, processamento e transmissão como um processo que depende de objetos, memórias externas como

hardwares computacionais ou corpos humanos. Não faz sentido reificar esse processo, querendo colocá-lo em um *medium*. “É por isso que todas as noções reificadas (tais como ‘alma’, ‘espírito’, ‘identidade’, ‘ego’ ou ‘self’ devem ser banidas). Uma nova antropologia deve ser desenvolvida: a habilidade peculiar de nossa espécie para armazenar, processar e transmitir informação adquirida tem que ser compreendida.” (Flusser, 1999, p. 205).

Em concordância com Flusser, Escobar (2000, p. 56) afirmava que “qualquer tecnologia representa uma invenção cultural, no sentido de que tecnologias trazem consigo um mundo; elas emergem a partir de condições culturais particulares e, por sua vez, ajudam a criar novas situações sociais e culturais”. Por isso, segundo o autor, os antropólogos deveriam estar particularmente preparados para compreender esses processos, precisando para isso estarem “receptivos à ideia de que a ciência e a tecnologia se constituem em uma arena crucial para a criação da cultura no mundo de hoje”.

Vinte e tantos anos se passaram e o Google fala por si, como representante do ponto a que chegou a memória cognitiva externalizada do *Sapiens*. Segundo Oliveira (2009, p. 332), a proliferação de ciberdispositivos, de que os fluxos de dados são gerados, operados e transmitidos, conduziu à conexão não apenas entre humanos, mas também de humanos com máquinas, máquinas com máquinas, objetos com máquinas e humanos com objetos, em ambientes sensíveis e responsivos que configuram um sistema cognitivo exteriorizado, distribuído e cooperativo em escala suficiente para se tornar invisível.

O que está acima brevemente delineado é o contexto tecnológico em que a inteligência artificial (IA) se destaca como uma das principais protagonistas, um contexto cuja ambivalência magna revela-se na inserção da IA em um planeta hiper conectado, com seus tentáculos voltados para o espaço sideral, enquanto populações inteiras desse mesmo planeta ainda vivem em condições sub-humanas e desumanas, questão exemplar da mais profunda contradição.

Não obstante o necessário desmascaramento dos processos de poder promovidos pela inteligência algorítmica, para dar conta da lógica ambivalente e paradoxal que hoje domina, é preciso reconhecer que estamos vivendo uma nova ordem cognitiva em expansão, uma escala informacional que a mente humana em si não mais alcança. Do poder de computação à alimentação de dados, os sistemas de IA cresceram tremendamente nos últimos anos. A extensa escala de implantação e responsabilidades atribuídas à IA agora envolve variados aspectos da tecnologia no cenário. Seja aprendizado profundo ou dimensionamento de qualquer outra tecnologia avançada que envolva o uso de IA, a situação escapou das capacidades de compreensão até mesmo dos profissionais mais proficientes.

Embora os benefícios da IA não sejam tão frequentemente lembrados quanto suas externalidades negativas, costumo dizer que os benefícios falam por si e obviamente não precisam de defensores, enquanto os aspectos negativos da IA precisam ser combatidos. Dos rios de discursos que vêm sendo produzidos acerca da IA, é possível extrair alguns temas-chave. Entre aqueles que comparecem com mais frequência estão os vieses que ferem os direitos fundamentais, os riscos, inclusive existenciais e, sobretudo, a necessidade de regulamentação tendo como horizonte a ética.

De fato, as questões relativas à regulamentação têm mobilizado vários países do mundo. A União Europeia, depois de pelo menos três anos de debates, chegou ao *AI Act*, uma regulamentação

abrangente que estabelece padrões rigorosos de proteção de dados e categorizações de risco. Os Estados Unidos implantaram os princípios gerais do *AI Blueprint*. A China também se mobiliza nessa direção, assim como outros países que buscam não perder de vista a inovação harmonizada com regulações setoriais. No Brasil, o debate é levado a cabo por uma comissão temporária. Portanto, também estamos caminhando nessa direção.

OBJETIVOS DESTA ARTIGO

Embora defenda com veemência a necessidade de regulamentação da IA e seus sucedâneos e tenha curiosidade em me informar sobre isso, não sou uma especialista no tema. No dia 15 de outubro de 2022, Antonio Lavareda (2022) publicou um artigo na *Folha de S. Paulo* cuja introdução coincide com uma questão que tenho defendido à sociedade, a saber, a leviandade que deve ser evitada de nos pronunciarmos sobre temas dos quais não somos especialistas. Em primeiro lugar, porque temos de prezar a ética da vida intelectual. Em segundo lugar porque, nesta era da desinformação, supostos e pretensos especialistas, mesmo que de modo não deliberado, também ajudam a alimentar os imbróglis da desinformação. Diante disso, decidi enfrentar para este artigo um tema ao qual tenho me dedicado e que dá alguma legitimidade para meus argumentos. Esse é o tema da ética, em seus modos de interferência na IA.

Antes de tudo, é relevante o sinal de alerta enunciado por Katherine Hayles (2017, p. 4):

[...] enquanto as investigações em ética tradicional colocavam seu foco no ser humano considerado como um sujeito possuidor de vontade livre, essas perspectivas me parecem inadequadas para lidar com os dispositivos técnicos que operam autonomamente, assim como em assemblagens humano-técnicas complexas nas quais os poderes de cognição e de tomadas de decisões distribuem-se através de todo o sistema.

Já existe um grande número de títulos sobre IA e ética. Em 2011, quando a IA apenas começava a despontar na onipresença que viria ocupar, Bostrom e Yudkowsky (2014) já antecipavam a importância que o tema estaria fadado a desempenhar, ao discutir tanto os riscos que a IA apresenta de ferir os direitos fundamentais quanto a controversa questão do estatuto moral das máquinas nelas mesmas. Daí para a frente, a ética passou a ocupar a preocupação e a agenda de todos aqueles envolvidos tanto na prática quanto na avaliação crítica da IA. Duas obras fundamentais foram lançadas pela Oxford Press em 2020: um manual editado por Duber, Pasquale e Das (2020), e o livro editado por Liau (2020), que recebeu ampla circulação. Segundo Kaufman (2021, p. 158),

Composta de dezessete ensaios inéditos, produzidos por cientistas e filósofos, agrupados em quatro seções, a coletânea aborda dilemas-chave para uma IA ética, dentre outros, os impactos da automação inteligente no mercado de trabalho; o viés contido nos dados que perpetuam os preconceitos da sociedade; a ética envolvida em aplicações como carros autônomos, sistemas de vigilância, armas autônomas; robôs sexuais; direitos e consciência da IA; e status moral. Numa perspectiva futura, os ensaios da terceira seção refletem sobre os riscos da “superinteligência”.

A Unesco (2022) pronunciou-se sobre o tema em 2022, ao propor suas *Recomendações sobre a ética da inteligência artificial*. Começando com a discussão dos valores e dos princípios éticos, desenvolveu, em um panorama amplo, os diferentes aspectos da ação política voltados para: 1) Impacto ético, 2) Governança e gestão, 3) Política dos dados, 4) Desenvolvimento e cooperação internacional, 5) Meio ambiente e ecossistemas, 6) Gênero, 7) Cultura, 8) Educação e pesquisa, 9) Comunicação e informação, 10) Economia e trabalho, 11) Saúde e bem-estar social.

Dada a sua relevância, a ética da e na IA tem ocupado o cenário de modo cada vez mais proeminente. Dois filósofos de prestígio no campo da IA lançaram suas obras para discutir a questão: Coeckelberg (2020) e Floridi (2023). Para o primeiro, o *hype* que circunda a IA fez emergir todos os tipos de especulações não só a respeito do futuro da própria IA, mas, sobretudo, sobre o futuro do que significa ser humano. Para isso, o autor dedica todo um capítulo sob o título de “Tudo sobre o humano”, em que são discutidas as diferenças entre humanos e máquinas. Diante da onipresença da IA na vida humana, o autor coloca em relevo a importância da ética para enfrentar as vulnerabilidades provocadas pela IA. Floridi, por seu lado, na primeira parte do livro explora o passado, o presente e o futuro da IA. Na segunda parte desenvolve longas discussões sobre a distinção, para mim equivocada (Santaella, 2023, p. 100-105), entre agenciamento e inteligência.

De modo sintético, Pujara (2022, s.p) afirma que quando se trata de “Ética na IA”, o termo significa investigar e questionar constantemente as tecnologias que podem dificultar a vida humana. Seja substituindo humanos por máquinas inteligentes ou preocupações relacionadas ao compartilhamento de informações pessoais em sistemas alimentados por IA. Pode-se dizer que a era do aprendizado de máquina e do aprendizado profundo se viu acompanhada pela preocupação com a ética. A partir disso se pode concluir que, não sem razão, IA e ética são inseparáveis.

Este artigo não tem a intenção de discutir a literatura sobre IA e ética, mas sim, pinçar, dentro desse tema, uma questão que me parece primordial. Para quaisquer áreas de aplicação da IA, seja na medicina, no direito e especialmente na educação, tenho defendido a necessidade de se diferenciar a IA que se desenvolve sob o nome de IA preditiva, daquela mais recente que vem sendo conhecida como IA generativa. Essa distinção é de extrema relevância porque, embora apresentem semelhanças gerais nos seus aspectos positivos e negativos, quando se leva a questão para a ética é preciso estabelecer algumas diferenças sem as quais a questão cai em generalidades que não ajudam na detecção dos problemas éticos que rondam cada um desses tipos de IA.

Tendo isso em vista, no que se segue este artigo está dividido em três partes. A primeira voltada para evidenciar as diferenças entre IA preditiva e IA generativa. Logo depois, procede-se a refletir sobre a ética da IA, à luz da ampla literatura sobre o tema, visando perquirir, inclusive, a possibilidade de transferência do saber acumulado sobre a preditiva para a generativa. Enfim, pretende-se trazer à tona os modos diferenciados de funcionamento das mesmas para que se possa enxergar o grau de interferência, das constatadas diferenças, no tratamento da ética em uma e na outra.

A IA PREDITIVA E A IA GENERATIVA

Compreender os efeitos produzidos pela IA implica compreender, mesmo que de modo leigo, a maneira como ela funciona. Segundo Medha (2024), a IA preditiva ingere grandes volumes de dados históricos de diferentes fontes, relevantes para o problema que lhe é colocado. Então os algoritmos de aprendizado de máquina analisam esses dados buscando tendências, padrões e relacionamentos entre variáveis.

Para isso, ela depende da modelagem estatística, ou seja, várias técnicas estatísticas e de aprendizado de máquina são usadas para, a partir dos dados, treinar modelos que sejam preditivos, modelos que sejam treinados com o propósito de alcançar determinado resultado. Os métodos de modelagem mais utilizados são análise de regressão, árvores de decisão, redes neurais, previsão de séries temporais e modelagem de conjuntos. Então, vem a fase da validação do modelo. Para isso, a exatidão e a precisão dos modelos devem ser não apenas rigorosamente testadas, quanto os modelos refinados até que o nível desejado de desempenho preditivo seja alcançado. A seguir, com os modelos razoavelmente precisos, passa-se para a simulação de cenário, quando diferentes cenários são simulados para o ajustamento dos parâmetros de entrada de modo a estimar previsões sob diversas condições. A etapa seguinte é a da implantação do modelo em ambientes de produção, o que não impede que novos dados sejam continuamente inseridos nos modelos para gerar *insights* preditivos atualizados.

Por fim, vem a integração de processos dos *insights* preditivos “aos processos de negócios e fluxos de trabalho por meio de painéis, alertas APIs etc., para permitir a tomada de decisões orientada por dados com base nas previsões do modelo” (Medha, 2024, s.p). Todo esse percurso torna a IA preditiva poderosa e valiosa para as corporações e organizações atuais, com o *surplus* de que os modelos tornam-se mais inteligentes com o tempo, à medida que processam mais informações (Medha, 2024, s.p).

Como se pode ver, são várias e diferenciadas as fases para o desenvolvimento da IA preditiva. Se houver falhas em qualquer uma delas, o resultado sofrerá contaminações, o que significa que princípios éticos serão feridos, conforme será discutido mais à frente.

Sem que se possa considerar como abordagens isoladas pois ambas, IA preditiva e IA generativa, pertencem ao mesmo campo de pesquisa e possíveis aplicações, a IA generativa, por seu lado, diferencia-se da IA preditiva porque segue preceitos relativamente distintos, embora ainda se utilize de aprendizagem de máquina e de redes neurais. Segundo Bharath (*apud* Lawton, 2023), a generativa está voltada para a criação de conteúdo novo e original, como imagens, texto e outras mídias, aprendendo com os padrões de dados existentes. Por isso, pode-se dizer que auxilia em campos criativos e na resolução de novos problemas.

A IA generativa é um subconjunto do aprendizado profundo, mas de um tipo diferente, chamado de Modelo Gerativo que aprende com um conjunto subjacente de dados para gerar novos dados que imitam de perto os dados originais. Por meio do emprego de aprendizagem não supervisionada, esses modelos são usados principalmente para criar novos conteúdos, como imagens, texto ou até mesmo música, semelhantes àquilo que pode ser criado por humanos.

Os modelos geradores de textos estão baseados em grandes modelos de linguagem (*Large Language Models* – LLM) que é um tipo de aprendizagem de máquina treinado em um grande conjunto de dados de texto e que usa arquiteturas de redes neurais avançadas para gerar ou prever textos semelhantes aos humanos. Os LLMs estão ligados ao processamento de linguagem natural (*Natural Language Processing* – NLP), que é um subconjunto da IA focado na interação entre computadores e humanos por meio da linguagem natural. Ele está habilitado a entender a língua humana e a se comunicar conosco na mesma língua (preferencialmente o inglês, vale lembrar), graças a algoritmos que ajudam os computadores a entenderem o contexto e o sentimento por trás das palavras e das sentenças.

Embora breves, as explicações acima são capazes de evidenciar o funcionamento distinto de cada uma das IAs. A diferença maior, entretanto, reside no fato de que, enquanto a IA preditiva existe na relação dos desenvolvedores com as corporações, a IA generativa, por seu lado, caiu diretamente no colo dos usuários em conexões de fácil utilização. Isso traz consequências profundas para a ética da IA generativa que não são exatamente as mesmas que operam na IA preditiva, senão vejamos.

A ÉTICA NA IA PREDITIVA

Na sua base, o tema da ética tem uma natureza filosófica. Há de fato, uma ética estritamente filosófica, que já teve seu nascimento em Aristóteles (ed. port. 2015) e que, séculos depois, Kant (ed. port., 2018) a tornaria paradigmática no Ocidente, desdobrando-se nas obras de muitos outros filósofos. Há também éticas teóricas situadas em campos de conhecimento específicos, como uma ética sociológica e uma ética jurídica. Este artigo não irá tomar esse caminho, pois, de uma forma breve, já o desenvolvi em outro trabalho (Santaella, 2021). Embora os pressupostos filosóficos e teóricos sejam tomados como referência em éticas práticas ou aplicadas, estas levantam questões que podem ser chamadas de questões situadas. Esse é o caso da ética da IA, que apresenta questões e levanta discussões que lhe são específicas.

No campo da IA preditiva, as questões da ética costumam ser tratadas a partir de uma categorização de princípios fundamentais que não devem ser feridos. Na maior parte das publicações, os princípios que comparecem são os seguintes: 1) Equidade (*Fairness*), 2) Confiabilidade e Segurança (*Reliability and Safety*), 3) Impacto Social (*Social Impact*), 4) Responsabilidade (*Accountability*), 5) Privacidade e Segurança (*Privacy and Security*) e, por fim, 6) Transparência (*Transparency*). Não há dúvida de que são princípios fundamentais. No entanto, sem minimizar o funcionamento desses princípios como sinais de alerta intransponíveis, o que se tem aí são grandes e valiosas palavras que, infelizmente, quando são postas sob reflexão, revelam-se recheadas de ambiguidades. São essas ambiguidades que entram em ação quando se trata de colocá-las em prática. A partir disso, tenho tomado como rota de pensamento para a ética da IA preditiva vasculhar essas ambiguidades com mais cuidado. Já o fiz com o tema da equidade em cuja semântica está implícito o sentido de igualdade (Santaella, 2021). Agora farei o mesmo com o tema da privacidade.

Antes que a IA explodisse na ciência e na sociedade, desde 2007, já tinha a questão da privacidade no centro de minhas preocupações (Santaella, 2007). Tanto é que publiquei capítulos sobre as dificuldades que rondam o sentido do conceito (Santaella, 2007;2010; 2013). Em lugar de tomar o significado de privacidade *prima facie*, tratei de explorar as ambivalências, os paradoxos e os desafios da privacidade no mundo contemporâneo capturado na teia das redes. Com a explosão da IA, a questão mudou um pouco de figura, ou melhor, os desafios se intensificaram, conforme será brevemente apresentado no que se segue.

Conforme já discutido em Santaella (2022, p. 284-286), curioso ou não, os textos, que mais têm chamado atenção, relativos à batalha contra as ameaças à privacidade, têm sido escritos por mulheres, a começar por Doss (2020) na sua busca por respostas para três grandes questões: De quem é a propriedade dos nossos dados? Por que devemos nos importar com isso? O que podemos fazer contra isso? Suas explicações não deixam por menos – o rastreamento de dados é um problema real que afeta cada um de nós e cada atividade mediada por meios digitais que realizamos. Pouco importa quem somos nós, qual é a nossa idade, quais são as nossas preocupações, nossos dados estão sendo coletados e nosso perfil desenhado, com o intuito de ser acessado de modo a analisar e prever nosso comportamento para fins que não deveríamos ignorar.

Segundo a autora, de um lado, a tecnologia evolui com rapidez crescente, de outro, as leis e as políticas se transformam com lentidão. Em função disso, é preciso, antes de tudo, empoderar os indivíduos para que lutem por sua autodefesa como um modo de pressionar a legislação regulamentadora. Sem negar os benefícios que o *big data* também pode trazer para a sociedade, Doss (2020) alerta para o fato de que os olhos devem ser mantidos sempre abertos e os riscos administrados.

Na esteira deixada por Zuboff (2019), no seu celebrado livro *Capitalismo de vigilância*, Véliz (2021a, 2021b) vai ainda mais longe ao chamar as *big techs* de abutres, pois quaisquer que sejam os aparelhos, televisão, computador, *smartphone*, eles disfarçadamente funcionam como espões de cada pequeno ato que realizamos. O capital procria por meio do poder que assumiu sobre nossas vidas, uma condição que foi estudada por Eugenio Bucci (2021) como uma superindústria do imaginário. Com isso, Eugenio Bucci apresenta o capital em uma perspectiva de captura do desejo pelo olhar, bem mais engenhosa do que o discurso repetitivo de muitas autoras e autores, que mudam apenas a modulação das palavras. De qualquer maneira, acredita-se que denunciar esses subterfúgios insidiosos e proteger nossos dados e nossa privacidade é um modo de lutar pela democracia. Trata-se, portanto, de uma política da privacidade que não se restringe à noção individualista e mesmo egoísta da privacidade, mas se alarga até a defesa de democracias saudáveis.

Sob esses discursos, existe, portanto, o defrontamento com a necessidade de redefinição das tradicionais concepções de privacidade. Essa necessidade, imposta pela era da dataficação, é atendida na obra de Richards (2022) dividida em duas partes. A primeira parte está voltada para a explanação sobre o que a privacidade é e não é. São desmistificados os argumentos que visam enfraquecer ou negar a existência da privacidade em tempos atuais, como se ela não tivesse mais importância entre os jovens ou se limitasse ao controle e à propriedade individual dos dados. Na segunda parte, são discutidos quatro valores: identidade, liberdade, proteção e confiança. Embora o autor seja cuidadoso, não se pode deixar de sublinhar o quanto esses conceitos são

escorregadios, podendo levar à defesa ingênua de noções iluministas que foram suficientemente criticadas pela filosofia contemporânea e desgastadas pela fragilidade social de que eram portadoras. O que fica de relevante, entretanto, é assumir a releitura da privacidade como uma forma de contrapoder em prol da construção de sociedades dataficadas mais justas.

Obra feminina que chamou atenção internacional pela denúncia ética que ela encerra em relação aos efeitos culturais da IA, *Discriminating data*, foi publicada por Chun (2021). Em comentário ao seu livro, Galloway (2022) retomou os argumentos da autora para desconstruir a crença epistemológica na correlação, evidenciando que, na verdade, o poder dos dados no hiper capitalismo hodierno opera sobre os alicerces da correlação, da discriminação, da autenticação e do reconhecimento. A discriminação, por exemplo, é uma patologia social, mas também é parte integrante da racionalidade discreta. Em suma, as *big techs* são o mecanismo pelo qual o poder opera, “e o poder hoje é racista, misógeno, repressivo e excludente. O poder não incita o desejo tanto quanto o sufoca e o disciplina” (Galloway, 2022, s.p).

São bem-vindas as discussões em profundidade das grandes categorias éticas que têm norteado a ética da IA preditiva, bem como têm sido tomadas como fundamentos para os modelos de regulamentação da IA em vários países. Embora essas mesmas categorias estejam também nas bases da ética na IA generativa, há, nesta, problemas adicionais que precisam ser considerados.

A ÉTICA NA IA GENERATIVA

Os problemas éticos da IA generativa operam em nível mais micro do que na IA preditiva. A IA generativa é de uso pessoal. Qualquer pessoa minimamente equipada com uma prótese computadorizada, ao acionar o ChatGPT ou congêneres, tem à sua disposição, inclusive gratuita, um multitarefairo solícito, disponível a responder com palavras, imagens, vídeos e sons a quaisquer tipos de comandos do usuário. Isso, em princípio, é irresistível.

Vários sites (*Ethics in...*, por exemplo) têm alinhado os problemas éticos mais evidentes que essa nova interface humano-máquina é capaz de trazer. Falar como o humano, desenhar como o humano, gerar música como o humano podem aumentar a produtividade dos negócios. Entretanto, do outro lado da moeda, também aumentam o potencial de geração de conteúdo prejudicial, ofensivo e falso. Isso se tornou flagrante com a disseminação das *deep fakes*, antes de produção mais precária e agora habilitadas para gerar imagens, sons e verbos convincentes, mas paradoxalmente falsos.

Do mesmo modo que outros modelos de IA, a generativa é alimentada e treinada com um grande volume de dados. Uma vez que seu *output* é similar a textos e imagens produzidos por humanos, a IA corre o risco de infringir os direitos autorais e direitos de propriedade intelectual de empresas ou de pessoas. Além disso, os dados de treinamento subjacentes podem conter informações confidenciais, incluindo informações de identificação pessoal. Isso implica o controle diligente dos dados do usuário, devido aos riscos decorrentes da divulgação inadvertida de informações confidenciais.

Já amplamente divulgado no campo da ética da IA preditiva, também na IA generativa, se os dados de treinamento carregarem preconceitos predominantes na sociedade, o modelo assim se comportará. Preocupação bastante lembrada diz respeito ao aumento e à aceleração da produtividade prometidos pela IA que agem como uma faca de dois gumes, pois igualmente prometem a perda de empregos, uma vez que são necessários menos seres humanos para realizar a mesma quantidade de trabalho.

Em suma: o poder que essa nova tecnologia, que simula ser humana, carrega é tão significativo quanto os riscos éticos que aciona, incluindo a distribuição de conteúdos nocivos, violações de direitos de autor, violações da privacidade de dados e a amplificação dos preconceitos existentes.

Como se tudo isso não bastasse, o problema mais tipicamente generativo diz respeito à dificuldade ainda não ultrapassada de se reconhecer se um texto foi escrito por IA ou por um ser humano. Várias pesquisas foram realizadas evidenciando essa dificuldade, o que sinaliza consequências éticas ainda imprevistas para os mais variados campos profissionais e níveis educacionais cujo material de trabalho é a linguagem.

Em meio a incertezas, já se pode comprovar que se tornou raro encontrar alguém confesso ou inconfesso nesses ambientes que não esteja fazendo experiências com o GPT, especialmente o GPT 4, que chega a assustar com seu desempenho à imagem e semelhança do humano. Sua solicitude para realizar tarefas é tão atraente que se torna difícil resistir. Até que ponto essa IA se apresenta como auxiliar e parceira, e até que ponto seu uso é feito para enganar o outro e a si mesmo, é uma questão crucial que se coloca. O que temos aí é uma encruzilhada que não nos dá outro caminho a não ser aquele de pregar a necessidade mandatória da ética, a saber: não apenas a urgência de uma educação para a ética, mas também de introjeção da ética no foro íntimo de cada um de nós.

REFERÊNCIAS

- ARISTÓTELES. *Ética a Nicômaco*. Luciano Ferreira de Souza (trad.). São Paulo: Martin Claret, 2015.
- BOSTROM, Nick; YUDKOWSKY, Eliezer. The Ethics of Artificial Intelligence. In: FRANKISH, Keith; RAMSEY, William (eds.). *Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, New York, 2014. p. 316-334.
- BUCCI, Eugenio. *A superindústria do imaginário*. Belo Horizonte: Autêntica, 2021.
- CHUN, Wendy H. K. *Discriminating Data: Correlation, neighborhoods, and the new politics of recognition*. Cambridge, Mass.: MIT Press, 2021.
- COECKELBERG, Mark. *AI ethics*. Cambridge, Mass.: MIT Press, 2020.
- DOSS, April Falcon. *Cyber Privacy: Who has your data and why you should care*. Dallas, Texas: BenBella, 2020.
- DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (eds.). *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 2020.
- ESCOBAR, Arturo. Welcome to cyberia: Notes on the anthropology of cyberspace. In: BELL, David; KENNEDY, Barbara M. (eds.). *The cybercultures reader*. London e New York: Routledge, 2000. p. 56-76.
- ETHICS in generative AI. Disponível em: <https://www.datacamp.com/tutorial/ethics-in-generative-ai>, s.d. Acesso em: 20 jun. 2024.
- FLORIDI, Luciano. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press, 2023.
- FLUSSER, Vilém. Memories. In: DRUCKREY, Timothy (ed.). *Ars Electronica: Facing the future. A survey of two decades*. Cambridge, Mass.: MIT Press, 1999. p. 202-206.
- GALLOWAY, Alexander. Big Bro: Review of Wendy Hui Kyun Chun, *Discriminating Data Correlation, Neighborhoods, and the New Politics of Recognition*. B20. 10 jan. 2022. Disponível em: <https://www.boundary2.org/2022/01/alexander-r-galloway-big-bro-review-of-wendy-hui-kyun-chun-discriminating-data-correlation-neighborhoods-and-the-new-politics-of-recognition/>. Acesso em:

HAYLES, Katherine. *Unthought: The power of the cognitive nonconscious*. Chicago and London: University of Chicago Press, 2017.

KANT, Immanuel. *Lições de ética*. Bruno Leonardo Cunha e Charles Feldhaus (trads.). São Paulo: Unesp, 2018.

KAUFMAN, Dora. Resenha do livro *Ethics of artificial intelligence*, de S. Matthew Liau. *Revista Digital de Tecnologias Cognitivas*, n. 23, jan./jun. 2021, p. 157-163.

LAVAREDA, Antonio. Lavareda responde a ataque a pesquisas e explica discrepâncias. *Folha de São Paulo*, 15 de outubro de 2022.

LIAU, S. Matthew (ed.). *Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2020.

LAWTON, George. Generative AI vs. predictive AI: Understanding the differences. 2023. Disponível em: <https://www.techtarget.com/searchenterpriseai/tip/Generative-AI-vs-predictive-AI-Understanding-the-differences>. Acesso em: 02 abril, 2024.

MEDHA. *Demystifying predictive AI: Definition and use*. 2024. Disponível em: <https://fireflies.ai/blog/predictive-ai>. Acesso em: 15 abr. 2024.

OLIVEIRA, Luiz Alberto. Homo Civilis (ou Homo Sapiens 2.0). In: NOVAES, Adauto (org.). *A condição humana: As aventuras do homem em tempo de mutação*. São Paulo: Sesc, 2009, p. 317-339.

PUJARA, Hima. *Artificial Intelligence: Underlining The 7 Most Common Ethical Issues*

2022. Disponível em: <https://medium.datadriveninvestor.com/artificial-intelligence-underlining-the-7-most-common-ethical-issues-cb257d5d905>. Acesso em: 10 nov. 2023.

RICHARDS, Neil. *Why Privacy Matters*. Oxford: Oxford University Press, 2022.

SANTAELLA, Lucia. *Linguagens líquidas na era da mobilidade*. São Paulo: Paulus, 2007.

SANTAELLA, Lucia. *A ecologia pluralista da comunicação. Conectividade, mobilidade, ubiquidade*. São Paulo: Paulus, 2010.

SANTAELLA, Lucia. *Comunicação ubíqua. Consequências na cultura e na educação*. São Paulo: Paulus, 2013.

SANTAELLA, Lucia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis Santiago; SANTAELLA, Lucia; KAUFMAN, Dora; CANTARINI, Paola. *Direito e inteligência artificial: Fundamentos*, vol 1 – inteligência artificial, Ética e Direito. Rio de Janeiro: Lumens Juris, 2021. p. 109-136.

SANTAELLA, Lucia. *Neo-Humano*. A sétima revolução cognitiva do Sapiens. São Paulo: Paulus, 2022.

SANTAELLA, Lucia. *Há como deter a invasão do ChatGPT?* São Paulo: Estação das Letras e Cores Ed., 2023.

UNESCO. *Recomendações sobre a ética da Inteligência Artificial*. 2022. Disponível em: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>. Acesso em: 10 out, 2023.

VÉLIZ, Carissa. Moral zombies: why algorithms are not moral agents. *AI & Society*, n. 36, p. 487-497, 2021a.

VÉLIZ, Carissa. *Privacy is power: Why and how you should take back control of your data*. London: Bantam Press, 2021b.

ZUBOFF, Shoshana. *The age of surveillance capitalism*. The fight for a human future at the new frontier of power. London: Profile Books, 2019.