

INTELIGÊNCIA ARTIFICIAL, ÉTICA E TECNOLOGIA

a contribuição da filosofia da tecnologia
para a discussão da ética em IA

ARTIFICIAL INTELLIGENCE, ETHICS AND TECHNOLOGY
a contribution of philosophy of technology to the discussion
of ethics in artificial intelligence

Martha Vieira Ribeiro¹
Universidade de Brasília

¹ Graduanda em Filosofia pela Universidade de Brasília (UnB).
E-mail: marthavi@gmail.com.

RESUMO: Os avanços da inteligência artificial e suas consequências trazem possibilidades de realizações positivas para a humanidade, mas também trazem ameaças. Tal potencial ambíguo exige um exame dessa tecnologia, a quem ela serve e que poder ela representa, abrangendo da ética à política, já que as mudanças trazidas implicam a necessidade de se investigar qual tipo de desenvolvimento da IA pode proporcionar um progresso sustentável e positivo para o indivíduo e a sociedade. Nesta relação do ser humano com máquinas que apresentam vasto conhecimento, capacidade de aprendizado e raciocínio, surge uma pressão sobre o ser humano para que encontre uma forma bem-sucedida de compartilhar o mundo com a inteligência artificial. Esse trabalho traz reflexões sobre ética e filosofia da tecnologia, na visão dos filósofos Mark Coeckelbergh e Andrew Feenberg, na busca de fundamentar uma resposta a essa questão.

Palavras-chave: Inteligência artificial. Ética. Filosofia da tecnologia.

ABSTRACT: The advances in artificial intelligence and their consequences bring possibilities for positive achievements to humanity but also pose threats. Such an ambiguous potential requires an examination of this technology, whom it serves, and what power it represents, ranging from ethics to politics, as the changes it brings imply the need to investigate what kind of development of artificial intelligence can provide sustainable and positive progress for individuals and society. In the relationship between humans and machines that possess vast knowledge, learning and reasoning abilities, there arises a pressure on humans to find a successful way to share the world with artificial intelligence. This work presents reflections on ethics and philosophy of technology, from the perspective of philosophers Mark Coeckelbergh and Andrew Feenberg, as an effort to establish a foundation for addressing this issue.

Keywords: Artificial intelligence. Ethics. Philosophy of technology.

INTRODUÇÃO

As novas tecnologias são tanto consequência como causa de mudanças marcantes nas sociedades humanas. A revolução agrícola neolítica e a primeira revolução industrial, nos séculos XVIII e XIX, são exemplos de avanços tecnológicos que causaram profundas mudanças econômicas, sociais, políticas e ambientais em suas épocas. Na atualidade, o advento da inteligência artificial, que assume dimensão global, pode ser considerado revolucionário, gerando iniciativas de acadêmicos, pesquisadores, corporações e governos para responder aos desafios impostos pelo seu crescente desenvolvimento. Em 1950, o matemático inglês Alan Turing já tratava do assunto, como mostra o artigo “Computing Machinery and Intelligence” (Turing, 1950), mas oficialmente as pesquisas sobre inteligência artificial foram inauguradas em 1956, quando estudiosos se reuniram no *Dartmouth College* para explorar esta área (Kaufman, 2019, p. 21). Com o progresso das ciências computacionais ao longo das últimas oito décadas, a inteligência artificial passou por períodos de estagnação e avanço, mas chegou à terceira década do século XXI sendo usada diariamente em todo o mundo por pessoas que muitas vezes não estão cientes de que a utilizam.

Um dos motivos do uso da inteligência artificial ter passado despercebido por algum tempo decorreu das diferentes características que assume, de acordo com sua aplicação. Nas últimas três décadas, diferentes modalidades de inteligência artificial já estavam sendo utilizadas para diversas atividades, como cirurgias robóticas, tradução e revisão de textos, direção de carros autônomos, simulação de processos bioquímicos e celulares, vendas, filmes, fotografias, georreferenciamento, entre outras, mas com a chegada dos robôs de conversação do tipo *chatbots de linguagem generativa pré-treinada (GPT)*, que constituem o objeto principal do presente estudo, o uso da inteligência artificial ganhou maior visibilidade. Os *chatbots* vinham sendo utilizados há algum tempo, embora com um repertório limitado, respondendo dúvidas ou direcionando clientes em menus de atendimentos pré-programados. Na última década, as grandes empresas de tecnologia digital, ou *Big Techs*, lançaram assistentes virtuais com funções mais complexas, que apoiavam o usuário em buscas na internet, no agendamento de compromissos, lembretes, acionamentos remotos de aparelhos domésticos e respostas a comandos e perguntas pontuais, tais como os assistentes *Alexa (Amazon)*, *Siri (Apple)* e *Google Assistant*. Com o lançamento do *Chat GPT 3.5 (Generative Pre-trained Transformer)* pela empresa *Open AI* em setembro de 2022, um *chatbot* que, por meio de conversação em linguagem natural é capaz de prover uma interação entre o ser humano e máquina muito semelhante à conversa que ocorre entre duas pessoas, houve difusão e impulso significativos à inteligência artificial generativa. O termo *generativa* denota a capacidade de geração de conteúdo autonomamente a partir de

treinamento prévio, o qual também possibilita o aprendizado autônomo de padrões, informações e estruturas provenientes de um grande conjunto de dados textuais, associado à capacidade de prever qual a próxima palavra de uma sentença. Isso resulta em um texto com continuidade lógica, além de permitir que o modelo melhore suas respostas à medida que é utilizado. Este tipo de inteligência artificial usa a arquitetura de modelo de *transformers* para o processamento de linguagem natural, possibilitado por um mecanismo de atenção que permite focar nas diversas partes que compõem um texto, e assim compreender qual o contexto e a relevância de cada palavra (Vaswani, 2017). Tais avanços foram viabilizados pelo surgimento dos *Large Language Model* (LLM), que são grandes modelos de linguagem baseados em redes neurais com múltiplas camadas profundas, o que permite que uma quantidade imensa de dados seja usada no seu treinamento, resultando em capacidade aumentada de processamento de linguagem natural. Ao conversar com o ser humano, a máquina, além de simular um diálogo, responde questões que lhe são propostas com uma velocidade muito superior à humana. O treinamento com grande quantidade de dados faz com que os LLM possam ser usados em diferentes tipos de tarefas, com necessidade mínima de ajuste, sendo por isso chamados modelos de base (IBM, 2024). Todas essas características fazem com que o uso dos *chatbots* GPT ocupe cada vez mais o cotidiano das pessoas. Com potencial para trazer tanto consequências positivas como ameaças à humanidade, é necessário que se examine a inteligência artificial do ponto de vista ético e político, uma vez que as mudanças produzidas podem afetar profundamente o indivíduo e a sociedade. Interessa investigar qual tipo de desenvolvimento dessa tecnologia pode proporcionar um progresso sustentável e positivo, que permita ao ser humano encontrar uma forma bem-sucedida de compartilhar o mundo com a inteligência artificial, considerando que é provável que ocorra uma pressão crescente para a humanidade se adaptar ao convívio cotidiano com máquinas que apresentam vasto conhecimento e capacidade de aprendizado e raciocínio.

Na busca de fundamentar uma resposta a essa questão, esse trabalho explora o pensamento dos filósofos Mark Coeckelbergh e Andrew Feenberg, estando dividido em três partes e conclusão. A primeira parte explora os potenciais riscos e benefícios do uso da inteligência artificial generativa, a segunda traz a visão de Andrew Feenberg sobre filosofia da tecnologia e como seus conceitos podem ser aplicados à inteligência artificial. A terceira parte incorpora a questão da ética na tecnologia, com foco no desenvolvimento tecnológico que propiciou o surgimento e o avanço da inteligência artificial, e é seguida pela conclusão, que mostra os possíveis caminhos para a aplicação do pensamento desses filósofos a essa nova tecnologia.

1 IA: RISCOS E BENEFÍCIOS

Os usos potenciais da inteligência artificial reanimam questões filosóficas antigas e ensinam outras, que passam pelo conceito de inteligência, pela relação do ser humano com máquinas inteligentes, pela possibilidade de máquinas se programarem autonomamente, ou mesmo a hipótese de uma máquina, no futuro, atingir o status de autoconsciência. São temas instigantes, mas ainda distantes da realidade atual da inteligência artificial (Coeckelbergh, 2023, p. 63). Há questões mais urgentes nos âmbitos individual, social e no tocante à tecnologia em si. Há vieses presentes no treinamento dos modelos de linguagem generativa, a produção de conteúdos falsos de textos, sons e imagens, que se refinam a cada dia, borrando progressivamente os limites entre o mundo da matéria, ou físico, composto por átomos, e o mundo virtual, composto por *bíts*. Por outro lado, existem limitações inerentes aos modelos de LLC atuais, como a elaboração de respostas sem sentido, a invenção de conteúdos, comumente chamada de “alucinação”, o uso excessivo de certos termos, e a possibilidade de colaborar com instruções de cunho nocivo, a despeito de treinamentos que enfatizem a moderação de tais conteúdos (Open AI, 2024).

Em termos individuais e sociais, há a questão da privacidade, uma vez que a exemplo do que já acontece nas redes sociais ou na própria navegação pela internet, a interação com os *chatbots* pode fornecer dados sobre o comportamento, preferências e escolhas dos usuários, de uma maneira mais rápida e mais completa. Esses dados, uma vez tratados, produzem informações amplamente comercializadas pelas empresas de tecnologia como sendo mercadoria valiosa, chamada por Shoshana Zuboff (2021, p. 22) de superávit comportamental. O crescimento das *Big Techs* foi impulsionado pela venda de dados fornecidos pelo usuário sem o seu conhecimento, e a esta nova forma de reunir capital foi dado o nome de capitalismo de vigilância: “O capitalismo de vigilância (CV) reivindica de maneira unilateral a experiência humana como matéria prima gratuita para a tradução em dados comportamentais.” (Zuboff, 2019, p. 22). A autora argumenta que o capitalismo de vigilância não se baseia simplesmente na venda desses dados, mas sim na extração de comportamentos preditivos para modificar e influenciar ações futuras.

O mais preocupante, portanto, é a ação em grande escala, moldando o comportamento da sociedade, suas opiniões, escolhas, interferindo na política e na economia de forma massiva: “O capitalismo de vigilância gera uma nova espécie de poder que chamo de instrumentalismo. O poder instrumentário conhece e molda o comportamento humano em prol das finalidades de terceiros.” (Zuboff, 2019, p. 23). Esse mercado extremamente lucrativo promove a concentração de dinheiro e poder na mão de um número cada vez menor de pessoas, o que acentua as iniquidades sociais. Juntam-

se a isso as modificações no mercado de trabalho trazidas por uma automação movida pela inteligência artificial, uma vez que muitas atividades hoje executadas por seres humanos poderão em breve ser substituídas por máquinas, com o consequente desaparecimento de vários postos de emprego. A avaliação sobre quais ocupações têm maior risco de desaparecer devido à inteligência artificial envolve muitas variáveis, mas há uma tendência para a automação substituir tarefas rotineiras e repetitivas, o que afeta precisamente as ocupações que exigem menor treinamento ou escolaridade para sua execução (Reino Unido, 2019), e que habitualmente são desempenhadas por pessoas em situação econômica menos favorecida na sociedade.

Desta forma, a despeito dos benefícios que a inteligência artificial possa trazer à humanidade, surgem problemas éticos que precisam ser abordados. A intenção de quem desenvolve a tecnologia pode ser boa, mas nada impede que ela seja usada para fins nocivos. É interessante analisar criticamente para quem a inteligência artificial traz benefícios, se para o cidadão ou para o governo, para o criminoso ou para a polícia, se para a empresa ou para o usuário. Além de ética, é uma questão de cunho político, dado o impacto que a nova tecnologia tem nas vidas individuais e na sociedade (Coeckelbergh, 2023, p.17). Essa questão pode se beneficiar do olhar da filosofia da tecnologia.

1 A FILOSOFIA DA TECNOLOGIA E A INTELIGÊNCIA ARTIFICIAL

O filósofo canadense Andrew Feenberg (2013, p. 51) explica que, nas sociedades tradicionais, são os mitos e costumes que moldam a forma das pessoas pensarem. Já na sociedade moderna, a ênfase passou a recair sobre a utilidade dos costumes e das instituições, que aliada à crença na ciência e na tecnologia, formou a base do modelo de pensamento que propiciou o surgimento da sociedade industrial. Quase três séculos depois, o mundo se vê às voltas com os resultados do modelo de desenvolvimento tecnológico que adotou, ao lidar com questionamentos que ultrapassam a esfera do utilitarismo. A filosofia da tecnologia surge a partir da necessidade de obter respostas para os problemas que despontam em uma sociedade com base tecnológica, tais como o tipo de mundo e de vida que ela oferece, cujos padrões de conforto, segurança e satisfação de desejos podem se mostrar predatórios, além de se distribuírem de forma notavelmente desigual entre seus membros, criando um paradoxo que desafia a viabilidade do próprio mundo que criou². A filosofia da tecnologia busca integrar a

² O paradoxo advém de que o avanço tecnológico, guiado por uma visão por demais utilitarista, pode inviabilizar a civilização que propiciou este avanço. Um exemplo é o câmbio climático ligado ao consumo de combustíveis fósseis e à destruição de florestas e outros ecossistemas, que em última análise pode comprometer a sobrevivência de inúmeras formas de vida no planeta, inclusive a dos seres humanos e sua civilização.

autoconsciência dessa sociedade, ao promover a reflexão sobre pontos normalmente considerados fora de questão, abrindo espaço para o pensamento crítico sobre a tecnologia e oferecendo opções para um desenvolvimento sustentável e socialmente mais justo (Feenberg, 2013, p. 52).

Para compreender o caminho da transformação social ocorrida a partir da idade moderna, o autor resgata o conceito clássico grego de *technè*, o qual se refere ao conhecimento que permite a criação ou produção humana (*poiesis*) de um artefato que tem propósito e significado. Diferentemente da natureza (*physis*), que cria a si mesma e traz à existência simultaneamente um ser e sua essência, na *technè* há uma essência pré-existente, que é então capturada pelo ser humano encarregado de concretizá-la no mundo material. Esta essência independe tanto do produto como do produtor, e sua criação obedece a um desígnio oculto na natureza, que guarda sentido e significado. A modernidade modificou tal visão, ao voltar sua racionalidade para um desenvolvimento tecnológico que, ao cumprir prioritariamente os desígnios da vontade humana, passou a subjugar e se apoderar da natureza de forma excessivamente utilitária.

A relação com uma essência previamente existente se esvaziou progressivamente, não importando mais o que um artefato é, e sim como ele funciona, ou seja, como pode ser útil, sem considerar o objeto em si um resultado do trabalho humano, o qual se impregna no artefato junto com a essência capturada, o que de certa forma o torna único. Para o autor (Feenberg, 2013, p. 55), quando o ser humano via sentido e significado tanto nas coisas produzidas por ele como nas produzidas pela natureza, isso o conduzia a uma percepção análoga de si como agente realizador de um mundo que possuía significado próprio. O ser humano partilhava dessa realidade maior e se envolvia com a essência à qual ele devia dar forma. A independência entre os fins e os meios, trazida pela modernidade, fez da natureza unicamente matéria-prima destinada a ser explorada para a satisfação das necessidades humanas, que se revelam ilimitadas. Embora o avanço tecnológico possibilite que o ser humano supra suas necessidades e viva materialmente bem, seguro e confortável, não perceber a essência daquilo que ele produz gera perda de sentido: sabe-se como chegar a um destino, mas não o porquê, ou para quê.

Para criticar o modelo da contemporaneidade tecnológica, Feenberg (2013, p. 57) apresenta esquematicamente as diferentes correntes da filosofia da tecnologia na forma de quadrantes, resultantes do cruzamento de dois eixos ortogonais. No eixo vertical, discute-se a questão da tecnologia ser neutra ou carregada de valores que influenciam a vida dos seres humanos, em nível individual e coletivo. O eixo horizontal questiona se a tecnologia é autônoma, ou seja, se o caminho para seu desenvolvimento é regido por leis intrínsecas à própria tecnologia, vedando a possibilidade de alternativas mais econômicas ou menos poluentes, ou se o ser humano pode assumir o controle dessa tecnologia por meio de suas escolhas:

| Quatro perspectivas da tecnologia, segundo Feenberg: | | |
|--|----------------|------------------------|
| Neutra | Autônoma | Humanamente controlada |
| | Determinismo | Instrumentalismo |
| Carregada de Valores | Substantivismo | Teoria Crítica |

Fonte: Própria autoria (2024).

A primeira teoria, chamada *determinista*, parte da visão de que a tecnologia é neutra e autônoma, e, portanto, controla o avanço da humanidade, que deve se adaptar às exigências tecnológicas em troca de maior poder e eficiência, sendo comumente adotada pelas ciências sociais. A segunda corrente postula que a tecnologia é neutra em termos de valores, mas é humanamente controlada, sendo conhecida como teoria *instrumentalista*. É a concepção do senso comum na sociedade atual, para o qual a tecnologia é apenas uma ferramenta para ser usada pelos seres humanos na satisfação de suas necessidades, sem que percebam criticamente se ela carrega valores que determinam seu modo de vida ou suas escolhas. Quando se admite que a tecnologia é carregada de valores, mas autônoma, surge a corrente *substantivista*, para a qual a tecnologia se assenta em atributos como eficiência e poder, em detrimento de outros valores importantes para o bem-estar humano. Como é autônoma, ela impõe estilos de vida opressivos, que levam a um estado de submissão da humanidade, com o apagamento de elementos vitais como a liberdade, a possibilidade de escolher, criticar e criar. O substantivismo tem a mesma visão da inexorabilidade tecnológica apresentada pela teoria determinista, mas com um viés mais pessimista e distópico. A quarta corrente, defendida por Feenberg, é chamada *teoria crítica da tecnologia*³. Embora concorde com o substantivismo quanto às consequências nefastas da tecnologia para a vida do indivíduo e da sociedade, essa teoria afirma que o controle humano sobre a tecnologia é possível, o que pode propiciar uma vida melhor e mais livre à humanidade, embora observe que não é isso o que ocorre no atual modelo tecnológico. O controle proposto pela teoria crítica da tecnologia deve advir de um processo democrático, tanto no desenho

³ O uso do termo teoria crítica por Feenberg revela a influência de filósofos da Escola de Frankfurt e sua crítica à sociedade capitalista, tais como Adorno e Horkheimer, especialmente em relação ao tema tecnologia. Estes autores, assim como Martin Heidegger e Jaques Ellul, expressaram grande pessimismo em relação ao papel da tecnologia, que chega a oprimir a essência humana no indivíduo e na sociedade, o que caracteriza a visão substantivista. Posteriormente, Herbert Marcuse criticou a visão extremamente negativa encabeçada por esses filósofos, mas não chegou a formular uma teoria alternativa. Coube a Feenberg, que foi seu aluno, elaborar essa crítica, com uma filosofia que implica em ação social, política, político-cognitiva e cultural (Neder, 2013, pp. 13-16)

como no desenvolvimento das tecnologias, afirmando que a tecnologia pode incorporar valores que vão além dos que estão contidos em conceitos como eficiência e domínio:

[...]. A tecnologia não molda apenas um, mas muitos possíveis modos de vida, cada um dos quais reflete escolhas distintas de objetivos e extensões diferentes da mediação tecnológica. A palavra *moldar* é usada propositadamente. Todos os quadros em um museu têm molduras, mas não é por essa razão que ali estão. As molduras são limites e contêm o que está dentro delas. De modo semelhante, a eficiência *molda* todas as possibilidades da tecnologia (Feenberg, 2013, p. 62).

A tecnologia não é vista como instrumento pela teoria crítica, mas como elemento estruturante de maneiras de viver. A escolha de como se dá o desenvolvimento tecnológico baseia-se no reconhecimento da conexão existente entre os meios e os fins, e deve ser submetida a um controle democrático que ateste sua adoção pela sociedade. Esta proposta remete à conexão existente entre o produtor e sua produção, a uma relação que dá sentido e significado à vida de quem produz artefatos, que existia na Grécia antiga e que de certa forma vigorou até a revolução industrial. Deste modo, a teoria crítica não vê a tecnologia separada nem da política nem da sociedade que a utiliza, nem do fazer tecnológico ou da matéria prima necessária, e reconhece que há diferentes opções disponíveis para se desenvolver uma determinada tecnologia, que são escolhidas de acordo com os objetivos e valores que saírem vitoriosos na competição pelo seu desenvolvimento.

A tecnologia da inteligência artificial generativa vem avançando de forma acelerada na última década. A vastidão de informações contidas nos LLC, associada à capacidade de composição de conteúdo de texto, som e imagem, conversação em voz natural e codificação de programas, tem alavancado o interesse pelos *chatbots* GPT. A discussão sobre os riscos inerentes a esses modelos é abafada em meio ao seu uso cada vez mais popularizado, prevalecendo na sociedade uma atitude passiva, que aceita o determinismo do tipo de tecnologia oferecida pelas empresas. O avanço da inteligência artificial ditado por elas é encarado como inevitável, obedecendo a leis intrínsecas que não podem ser diferentes do que são, e resta à sociedade se adaptar (Feenberg, 2002, p. 138). Essa concepção não ajuda na busca de soluções para os riscos que o desenvolvimento da inteligência artificial possa trazer à humanidade e inclusive às demais formas de vida. É preciso romper com essa visão estreita da tecnologia e admitir que ela pode ser desenvolvida de diversas maneiras, o que somado ao seu caráter relacional, deve incorporar aspectos éticos, sociais e políticos em sua formulação. Este é um caminho possível para o resgate do sentido e significado do ser humano em relação à sua produção tecnológica.

2 INTELIGÊNCIA ARTIFICIAL: DIÁLOGO ENTRE ÉTICA E TECNOLOGIA

Segundo Feenberg (2013, p. 69), a tecnologia gera poder nas sociedades pós-industriais, a ponto de as lideranças dos sistemas técnicos superarem em poder as instituições de governo, quando determinam como e quais são as tecnologias usadas cotidianamente em nossa sociedade. Como exemplo, os meios de transporte, os tipos de materiais de construção, o urbanismo, as ações em saúde, em educação e na segurança pública, todos foram influenciados pela incorporação das tecnologias desenvolvidas ao longo dos últimos séculos e moldaram o funcionamento da sociedade moderna. Contudo, este tipo de poder se apoia na crença de que política, economia e tecnologia podem ser tratadas como coisas independentes, aderindo com naturalidade à convicção de que a economia deve ser regida essencialmente pelo mercado. Marx detectou a conexão entre tecnologia e produção na economia da sociedade moderna ainda no século XIX (Feenberg, 2013, p. 104), postulando a necessidade de participação da população nas decisões de produção como verdadeiro exercício de cidadania. Embora hoje sociedades de base socialista e democrática não rejeitem essa necessidade, a democratização dos processos produtivos não tem ocorrido como teoricamente se esperava. É como se qualquer controle sobre a produção, inclusive a tecnológica, ameaçasse a democracia a partir do enfraquecimento de suas bases econômicas, o que reforça o ponto de vista determinista e substantivista da tecnologia. Não é de se estranhar que donos de *Big Techs* se oponham frontalmente ao controle das atividades de suas empresas por parte do governo ou qualquer outra instância da sociedade, alegando que “...a regulação sempre é uma força negativa que impede a inovação e o progresso, e que a ausência de legislação é o contexto necessário para a ‘inovação tecnológica’.” (Zuboff, 2019, p. 134). Esse apelo encerra a, por parte das *Big Techs*, de que é natural priorizar os objetivos da empresa, que só respeitará os interesses dos usuários se estes não interferirem em seu lucro.

A essa certeza se opõe a constatação de que uma tecnologia eficiente pode ser alcançada por diferentes conformações no uso dos recursos, e que escolhas sociais interferem tanto na delimitação do problema como em sua solução:

A tecnologia é não-determinada pelo critério da eficiência e responde aos vários interesses e ideologias particulares selecionados entre as opções. A tecnologia não é racional, no sentido antigo do termo positivista, mas socialmente relativa; o resultado de escolhas técnicas é um mundo que dê sustentação à maneira de vida de um ou de outro grupo social influente. Assim, as tendências tecnocráticas das sociedades modernas poderiam ser interpretadas como um efeito de limitar os grupos que intervêm nos projetos, desde os peritos técnicos até as elites corporativas e políticas, às quais essas tendências servem. Na formulação desta tese, consideramos que a intervenção dos interesses e das ideologias não reduz, necessariamente, a eficiência,

mas inclina sua realização de acordo com um programa social mais largo (Feenberg, 2013, p. 103).

Há muitas iniciativas acontecendo no âmbito global para o desenvolvimento de uma inteligência artificial confiável e boa para a humanidade e para o planeta, no campo governamental, acadêmico, técnico e mesmo empresarial. É relevante que vários segmentos da sociedade estejam constatando a necessidade de se estabelecer parâmetros regulatórios e éticos para o desenvolvimento, funcionamento e uso da inteligência artificial (Coeckelbergh, 2023, p. 162). É essencial alinhar os objetivos dados à inteligência artificial com aqueles que mais interessam ao bem-estar e florescimento da humanidade. Como determinar este alinhamento é um dos grandes desafios filosóficos e práticos a ser enfrentado no percurso do desenvolvimento da inteligência artificial. Mesmo as *Big Techs*, cuja orientação para o lucro e concentração de poder político, econômico e social as fazem pender para o desenvolvimento da tecnologia em detrimento de questões éticas, buscam o alinhamento, revelando a percepção de que é necessário um limite, além do qual haveria perdas para toda a humanidade. Todavia, a grande discussão quanto ao tema alinhamento concerne quais valores éticos devem integrar o referencial de valores aos quais a inteligência artificial deve se alinhar. A multiplicidade de interesses nos diversos segmentos da sociedade evidencia a dificuldade dessa questão, que exige ser abordada de maneira ampla e transectorial.

Feenberg (2013, p. 64) reconhece que a intervenção democrática na tecnologia pode salvar a humanidade de si mesma, em um momento em que a tecnologia é usada não somente para maior eficiência, mas também para concentrar poder, influência, lucro e riquezas nas mãos de um número cada vez menor de pessoas, com aumento da iniquidade social. O autor argumenta que o caminho é a participação da população em decisões sobre o desenho e o desenvolvimento das tecnologias e não apenas na sua utilização. Mas, como motivar esse tipo de ação sem que a sociedade abra espaço para o envolvimento, o conhecimento e a discussão de novas tecnologias, o que a tornaria capaz de pressionar por modelos tecnológicos que concorressem para o bem da humanidade em geral, e não apenas para o de poucos indivíduos? Esse é um obstáculo para que a intervenção democrática saia do plano teórico e se concretize, o que exige ações concretas para conscientizar a sociedade sobre o seu poder de influenciar o desfecho das tecnologias. No caso da inteligência artificial generativa, a própria tecnologia pode ser usada para a difusão do pensamento crítico, de ideias que promovam o questionamento e a proposta de alternativas à corrente hegemônica que dirige o desenvolvimento tecnológico. Mas é preciso saber utilizar a tecnologia para poder aplicá-la, sendo imperioso que as pessoas aprendam como utilizar a inteligência artificial generativa, para que possam obter os resultados que almejam.

Retornando às correntes da filosofia da tecnologia, vemos que o determinismo se fortalece na ignorância da sociedade sobre as conexões entre seu cotidiano e as tecnologias que o cercam. Ao ser considerado unidirecional e autônomo, como nas correntes substantivista e determinista, o desenvolvimento de uma tecnologia tende a se restringir aos engenheiros, tecnólogos, políticos e grandes empresários da tecnologia. Ignorando que podem interferir nesse processo, entre as pessoas nos demais segmentos sociais prevalece o ceticismo, que as leva à paralisia. Na teoria crítica da tecnologia, existe a perspectiva da ação, a possibilidade de interferência e mudança a partir da compreensão do processo como um todo: os atores envolvidos no desenvolvimento da tecnologia e seus respectivos interesses, a matéria-prima necessária, a mão de obra utilizada, o impacto ambiental, o funcionamento da tecnologia em si e seus possíveis benefícios e riscos. Apesar das dificuldades naturais para sua concretização, o caminho está em adquirir e popularizar o conhecimento para que seja possível reivindicar voz ativa na formulação das tecnologias e estar na linha de frente por uma regulação de seu uso. Para que haja uma mudança do caráter tecnológico é necessária uma mudança concomitante na política e na sociedade, movida pela criação de consciência, sendo importante destacar que a própria tecnologia pode ser usada como instrumento para essa conscientização⁴.

A história mostra que a evolução humana é dinamizada em momentos de crise, os quais pressionam por inovação e mudança diante de dificuldades. O surgimento dos modelos generativos de inteligência artificial, como argumentado acima, parece ser um desses momentos. Considerando-se que sua utilização ainda percorre estágios iniciais, a hora é propícia para que sejam elaboradas políticas de natureza ética e regulatória. Coeckelbergh (2023, p. 157) fala sobre a inovação responsável, que insere a ética no desenvolvimento tecnológico por meio da governança inclusiva, a qual requer debates públicos e intervenções da sociedade em suas etapas iniciais, levando-se em consideração as opiniões e os interesses dos diversos atores envolvidos. Contudo, hoje a discussão sobre ética e regulação em inteligência artificial ocorre primariamente no nível das corporações técnicas, políticas e empresariais, em uma iniciativa “de cima para baixo”, sem a participação efetiva da sociedade:

Com relação à pergunta ‘quem’ no âmbito da ética na IA, precisamos de mais espaço para as abordagens de baixo para cima, ao lado das abordagens de cima para baixo, no sentido de escutar mais os pesquisadores e profissionais que trabalham com IA na prática e, certamente, as pessoas potencialmente prejudicadas por ela. Se defendemos o ideal da democracia e se tal conceito envolve inclusão e participação na tomada de decisão sobre o futuro de nossas sociedades, a escuta das vozes das partes interessadas não é uma opção, mas um requisito ético e político (Coeckelbergh, 2023, p. 158).

⁴ A investigação de casos icônicos do uso pernicioso da tecnologia da informação, como o da *Cambridge Analytica* e o do algoritmo de previsão de reincidência criminal, o *Compas*, se apoiou no uso da própria tecnologia da informação. Para mais detalhes, consultar Sumpter (2019, pp. 38-60).

O aspecto sociopolítico deve, portanto, ser evidenciado em todas as discussões sobre inteligência artificial. Um controle voluntário agora, quando a inteligência artificial está se desenvolvendo e ocupando importância crescente no mundo, aumenta as chances de alinhamento, de torná-la parceira na construção de uma sociedade mais benigna, ao encontrar formas de diminuir a concentração de poder e riqueza, de coibir a formulação e divulgação de conteúdos falsos, de promover os governos democráticos, e não os autoritários. Ao persistir a oposição a qualquer tipo de regulação, sem priorizar o alinhamento dos objetivos da inteligência artificial ao melhor interesse da humanidade, dificilmente o progresso da inteligência artificial poderá atender às necessidades humanas de vínculo, afeto, estabilidade material, criatividade, deixando-a à mercê de estímulos provenientes de uma tecnologia determinada pelo interesse de grupos cada vez mais minoritários, que se tornam oligarquias tecnológicas.

Ao continuar como um ator inconsciente, que serve a comandos ditados por corporações poderosas de inteligência artificial, qual destino que o ser humano dará à sua sociedade? As chances de ações disruptivas aumentam: conflitos civis, religiosos, de costumes, desestruturação das democracias e deflagração de guerras que, em última análise, podem travar o avanço tecnológico e agir como força regressiva, e assim levar a um retrocesso civilizatório e tecnológico, ou seja, a um mundo distópico. As distopias, encenadas em ficções literárias e cinematográficas e que atraem um público considerável, mostram cenários assombrosos com os quais a humanidade tem que lidar, geralmente causados por uma catástrofe tecnológica, bélica ou ambiental, resultantes de ações nefastas de seres humanos, e simbolizam o choque entre as forças criativas e destrutivas existentes nos indivíduos e na sociedade. Uma fantasia distópica frequente é aquela encenada em um mundo em que sobreviventes humanos de um apocalipse lutam contra zumbis. Vervaeke, Mastropietro e Miscevic (2017, p. 3) apontam que “Zumbis nos têm pressionado com os perigos de um momento único no tempo, e se tornaram a metáfora mais duradoura, expressiva e consumada de nossa crise de significado”⁵. Em que pese o caráter ficcional dessas obras, por um lado elas refletem o temor de um futuro distópico, e por outro suscitam interesse, ao oferecerem uma possibilidade criativa de superação da distopia. Feenberg (2013, p. 64) fala sobre as transformações pelas quais as pessoas afetadas por uma tecnologia podem passar, interferindo incisivamente no futuro desta⁶. Porém, essas transformações dependem da tomada

⁵ “Zombies have pressed us with the dangers of a unique moment in time, and they have become the most enduring, expressive and consummate metaphor for our crisis in meaning” (tradução nossa).

⁶ Feenberg aqui ressalta que os idealizadores do *e-mail* o destinaram inicialmente a usuários qualificados, e não pensaram que se tornaria a função mais usada na internet. Sua popularização fez com que se tornasse uma das maiores contribuições do computador para a vida das pessoas. Outro exemplo de como a produção/oferta de tecnologias é influenciada pela sociedade encontra-se no movimento das pessoas com deficiência física nos Estados Unidos, nos anos 1970. Elas denunciavam que a urbanização, os meios de transporte e as máquinas fabricadas as impediam de se deslocar ou trabalhar, excluindo-as da vida pública

de consciência, por essas pessoas, das conexões existentes entre o modelo tecnológico e os interesses em jogo pelos diferentes atores em cena: quem desenvolve a tecnologia, quem a financia, quem a produz, quem a utiliza, assim como os governos e as forças do mercado que controlam o que é ofertado e ambicionam determinar as necessidades dos usuários da tecnologia, além de moldar suas vidas. Essa consciência é a mola que impulsiona as pessoas a expressarem uma qualidade que o autor considera como inerente à sociedade humana, que é uma autorregulação própria, capaz de orientá-la no sentido de uma vida mais plena e satisfatória, pelo bom uso da tecnologia: “Se puderem conceber e perseguir os seus interesses intrínsecos em paz e, além disto, alcançar sua realização por via do processo político, as pessoas assumirão a questão da tecnologia inevitavelmente junto com muitas outras aspirações que hoje se mantêm meras expectativas (Feenberg, 2013, p. 64)”.

Coeckelbergh (2023, p. 165) enfatiza o poder da educação na formação de uma sociedade mais atuante e crítica, e propõe uma abordagem transdisciplinar ao discutir ética em inteligência artificial, após constatar a distância entre a formação de pessoas nas áreas das ciências humanas daquelas formadas nas ciências naturais e tecnologia. Tal distância não favorece que a inteligência artificial avance com ética, pois falta a cada lado conhecer a importância do outro. A formação em humanidades deve evidenciar a necessidade de conhecer melhor os aspectos técnicos do funcionamento da inteligência artificial, a história de seu desenvolvimento e as perspectivas de seu avanço. O estudante de ciências humanas não deve, portanto, se limitar a utilizar a tecnologia da inteligência artificial, mas também conhecê-la para que possa fazer análises e críticas com embasamento. Por sua vez, a formação dos engenheiros computacionais e demais tecnólogos deve incluir disciplinas de humanidades, que promovam o desenvolvimento do pensamento crítico e ensinem a identificar as implicações éticas e sociais de certas tecnologias. Essa troca e incorporação de saberes pode formar pessoas com uma base mais robusta para a concretização de uma inteligência artificial na qual a ética não seja algo marginal, e sim fundamental.

A consciência, por parte do utilizador, da sua capacidade influenciadora sobre determinada tecnologia potencializa sua vontade de agir, de se organizar. Tomando posse da consciência enquanto ator dentro do cenário do desenvolvimento tecnológico, ele pode efetivamente contrabalançar as forças corporativas que almejam impor o domínio da sociedade por uma inteligência artificial com vieses opressivos e iníquos. Ao mesmo tempo preserva-se o impulso do desenvolvimento tecnológico, sem a postura regressiva de frear este avanço. A ideia não é barrar o avanço tecnológico, muito menos revertê-

(Cupani, 2016, p.155). Observou-se, ao longo das décadas seguintes, uma progressiva adaptação dos equipamentos públicos e a criação de novas próteses e outros artefatos tecnológicos para a inclusão das pessoas com deficiência na sociedade, não apenas nos Estados Unidos, como também nos demais países, o que tem modificado substancialmente a vida dessas pessoas.

lo. A ideia é cooptá-lo, por meio de ações coletivas organizadas, que possibilitem o desenvolvimento de uma sociedade tecnológica mais justa e ambientalmente sustentável.

CONCLUSÃO

A humanidade enfrenta, no atual momento civilizatório, o desafio de incorporar a inteligência artificial de forma benéfica para si, para as demais espécies e para a vida no planeta. Segundo a teoria crítica da tecnologia, que considera que a tecnologia é carregada de valores e pode ser controlada pela vontade humana, a decisão sobre o bom ou mau destino do ser humano perante o avanço da inteligência artificial não parte da tecnologia em si, mas dele mesmo, ao escolher como a utilizará. Esta teoria fundamenta sua argumentação na capacidade da sociedade em mudar formas de tecnologia hegemônicas por outras, antes referidas como inviáveis econômica ou tecnologicamente, mas que passam a se desenvolver, dependendo da pressão social, ambiental, política ou econômica, podendo conquistar a posição hegemônica. Este é um percurso conhecido do desenvolvimento tecnológico, mas o que a teoria crítica da tecnologia faz é conectar os avanços das tecnologias às consequências que trazem aos diferentes setores da sociedade, os quais habitualmente não são considerados, consultados, ou informados sobre elas. Adicionalmente, esses mesmos segmentos, que compõem majoritariamente a sociedade, não são ouvidos quando do desenvolvimento de uma nova tecnologia, e recebem um novo artefato tecnológico finalizado, o que exige que cada indivíduo se adapte e adeque ao seu uso, inconsciente da cadeia de produção e eventos que o trouxeram até suas mãos, em um padrão que corresponde à visão instrumentalista na filosofia da tecnologia.

No entanto, a efetivação da proposta da filosofia da tecnologia depende da vontade da sociedade em se engajar, aprender, debater, negociar, protestar e influenciar o uso e o desenvolvimento da inteligência artificial. Tal movimento ultrapassa o escopo da filosofia da tecnologia e exige o apoio de outras áreas da filosofia, especialmente a ética, além das ciências sociais e econômicas, da ciência política, do direito, entre outras, em um esforço de caráter transdisciplinar, que possa iluminar as conexões existentes entre a produção tecnológica, o consumo dos itens produzidos, as condições de sua produção e os interesses que a dominam. Em relação à inteligência artificial, isso inclui a busca de patamares éticos favoráveis ao alinhamento da inteligência artificial com o ser humano. Mas essas escolhas éticas não devem ser impostas de cima para baixo à sociedade, e sim incluir as escolhas vindas de forma ascendente, das bases da sociedade para os níveis decisórios das empresas de tecnologia, o que pode ser facilitado e incentivado por meio de informação, discussão e de propostas vindas de grupos representativos de segmentos sociais organizados.

Este tipo de intervenção requer um cenário político que permita a discussão e a participação popular nas leis, ações e políticas governamentais, o que só é possível em regimes democráticos. A inteligência artificial pode ser um instrumento de fragilização da democracia, como parecem atestar as tendências políticas autoritárias que vêm despontando no mundo, mesmo em países europeus, tradicionalmente democráticos. A força que a inteligência artificial exerce sobre sistemas eleitorais tem potencial para subvertê-los, ao distorcer fatos, impulsionar notícias tendenciosas ou mesmo falsas, e desmontar o processo democrático de forma velada, mas efetiva. Esse aspecto destrutivo da organização política de países democráticos, assim como o perigo de manipulação do poder político, com sua concentração cada vez maior nas mãos de cada vez menos atores, é muito preocupante. Isso conduz o olhar para o outro lado, para como a inteligência artificial pode ser usada no fortalecimento da democracia em outras bases, diferentes de sua atual aliança com o liberalismo econômico, que favorece o surgimento e empoderamento da tecnocracia das *Big Techs*.

O filósofo Mark Coeckelbergh vem se dedicando a estudar e propor caminhos para a concretização desse potencial da inteligência artificial na proteção, fortalecimento e promoção da democracia. Como visto acima, conhecer e se apoderar dos recursos oferecidos pela inteligência artificial é um caminho para esse fortalecimento. Seu poder de penetrar em todos os estratos sociais e difundir todo o tipo de informação abre espaço para a divulgação de ideias que promovam a conscientização dos seres humanos, levando-os a perceber o jogo de forças em meio ao qual estão vivendo, e o que cada força significa. A inteligência artificial oferece essa possibilidade, a qual deve ser utilizada para a construção de uma democracia que inclua em sua balança, além do capital e do lucro, valores humanos e sociais que conduzam a uma vida mais gratificante, à proteção do meio ambiente, a tecnologias sustentáveis. As ideias defendidas por Feenberg e Coeckelbergh podem tornar viável uma inteligência artificial justa e benéfica, tanto social como ambientalmente, o que pode permitir ao ser humano resgatar seu sentido e significado existenciais e evitar as distopias antevistas pelo substantivismo. Uma consideração mais detalhada e aprofundada das propostas defendidas por esses dois autores é recomendável, dada a atualidade e premência de se vincular o avanço tecnológico com o avanço ético no uso da inteligência artificial.

Ensaio recebido em: 20/01/2025

Ensaio aceito em: 28/03/2025

Ensaio publicado em: 31/03/2025

REFERÊNCIAS

- COECKELBERGH, Mark. *Ética na Inteligência Artificial*. São Paulo/Rio de Janeiro: Ubu Editora/Editora PUC-Rio. 2023.
- CUPANI, Alberto. *Filosofia da tecnologia: um convite*. 3. ed. Florianópolis: Editora da UFSC, 2026. Versão digital.
- FEENBERG, Andrew. *Transforming technology - a critical theory revisited*. New York: Oxford University Press. Edição digital. 2002.
- FEENBERG, Andrew. O que é filosofia da tecnologia. In: NEDER, R.T. *A Teoria Crítica de Andrew Feenberg: racionalização democrática, poder e tecnologia*. Brasília: Observatório do Movimento pela Tecnologia Social na América Latina/CDS/UnB/Capes, 2013 (2ª edição), pp. 49-66.
- IBM. *Como funciona um modelo LLM?* Disponível em: <https://www.ibm.com/br-pt/topics/large-language-models>. Acesso em: 14 jun. 2024.
- KAUFMAN, Dora. *A inteligência artificial irá suplantará a inteligência humana?* Barueri, SP: Estação das Letras e Cores, 2019.
- NEDER, Ricardo T. Apresentação: o que (nos) quer dizer a teoria crítica da tecnologia?. In: NEDER, R.T. *A Teoria Crítica de Andrew Feenberg: racionalização democrática, poder e tecnologia*. Brasília: Observatório do Movimento pela Tecnologia Social na América Latina/CDS/UnB/Capes, 2013 (2ª edição), pp. 07-24.
- OPEN AI. *Introducing Chat GPT 4: limitations*. Disponível em: <https://www.ibm.com/br-pt/topics/large-language-models>, acesso em 17 jun. 2024.
- REINO UNIDO. Office for National Statistics. *Which occupations are at highest risk of being automated?* Publicado em 25 mar. 2019. Disponível em: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/whichoccupationsareathighestriskofbeingautomated/2019-03-25>. Acesso em: 23 set. 2024.
- SUMPTER, David. *Dominado pelos números: do Facebook e Google às fake news, os algoritmos que controlam nossa vida*. Rio de Janeiro: Bertrand Brasil, 2019. Versão digital.
- TURING, Alan M. Computing machinery and intelligence. *Mind*, v. 59, n. 236, p. 433-460, out. 1950. Disponível em: <https://doi.org/10.1093/mind/LIX.236.433>.
- VASWANI, Ashish, et al. Attention is all you need. *31st Conference on Neural Information Processing Systems* (NeurIPS 2017), Long Beach, CA. USA. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 05 jun. 2024.
- VERVAEKE, John; MASTROPIETRO, Christopher; MISCEVIC, Filip. *Zombies in Western Culture: a twenty-first century crisis*. Cambridge, UK: Open Book Publishers, 2017. Versão digital.
- ZUBOFF, Shoshana. *A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Editora Intrínseca. Edição digital. 2021.

