Síntese: Revista de Filosofia adere a uma Licença
Não comercial 4.0 Internacional - Creative Commons

APRESENTAÇÃO

Desde seus primórdios, a inteligência artificial (IA) se estruturou como uma empreitada dupla. A simulação computacional de capacidades cognitivas atribuídas à inteligência humana permitiria, ao mesmo tempo, explicar essas capacidades e atribuí-las a outros substratos, tais como computadores de silício. Dessa forma, surgiu no horizonte uma empreitada científica voltada a explicar o que é a mente por meio de modelos computacionais — constituindo os primórdios do que hoje denominamos ciências cognitivas ou, mais recentemente, neurociências cognitivas — e uma empreitada de engenharia dedicada a criar mentes artificiais.

Contudo, no âmbito científico, a mente humana se mostrou resistente à análise computacional então empregada e, embora as ciências cognitivas ainda façam amplo uso de modelos computacionais, o papel que eles ocupam na pesquisa contemporânea é significativamente distinto daquele de algumas décadas atrás – a ponto de alguns pesquisadores considerarem-nos dispensáveis. Dificuldades semelhantes se manifestaram no âmbito da engenharia, ampliando a percepção de que as promessas grandiosas envolvendo computadores inteligentes não seriam cumpridas. Aos poucos, ficou claro que a enormidade do desafio havia sido subestimada e a empreitada adentrou o chamado "inverno da IA" – um período de descrédito, desinteresse e parco financiamento.

Apesar de breves períodos de otimismo ao longo das últimas décadas, foi apenas muito recentemente, nos anos 2010, que a IA voltou a ocupar um lugar central na pesquisa e na indústria. Diversos fatores contribuíram para essa nova fase, entre eles a disponibilidade de um poder computacional antes impensável e de um volume de dados inimaginável antes da Internet. Ambos viabilizaram técnicas antes impraticáveis.

Com efeito, vivemos hoje um renovado interesse pela IA enquanto empreitada de engenharia, mas com diferenças importantes. A preocupação central não é mais a de criar mentes artificiais (ao menos não em um sentido literal), tampouco a de criar sistemas computacionais que sirvam como modelos explicativos das capacidades humanas. Tal como no caso de máquinas de lavar louças, os esforços se concentram em obter resultados semelhantes aos dos seres humanos em determinadas atividades, mas sem se preocupar com a simulação dos meios pelos quais os seres humanos os alcançam.

Essa é uma das lições deixadas pela derrota de Kasparov para o *Deep Blue* no xadrez, em 1997: se, por um lado, é evidente que a forma como aquela IA "raciocinava" sobre o jogo não serve de modelo para a maneira como seres humanos pensam e jogam xadrez, por outro, parece igualmente evidente que podem existir formas distintas – inumanas – de jogar bom xadrez. A IA contemporânea generaliza essa lição para outros domínios, como o reconhecimento de padrões em imagens, vídeos e textos. Além disso, aprendemos que as técnicas usadas para identificar padrões podem ser aplicadas para produzi-los, originando assim os sistemas de maior impacto até aqui: os modelos generativos capazes de gerar textos, imagens, áudios e vídeos.

Renovou-se assim a promessa de desenvolver sistemas capazes de substituir seres humanos em vários domínios de atividade. Além disso, o cenário contemporâneo viabilizou a oferta de ferramentas e aplicações ao público, consolidando a IA como uma tecnologia de propósito geral, cujas potencialidades e riscos estamos apenas começando a conhecer. Não obstante, tais sistemas já assumem papéis decisivos: mobilizam nossa atenção, selecionam as notícias que leremos, deliberam sobre nossos perfis comportamentais e tomam decisões que afetam nossa saúde, nossa segurança e o funcionamento de nossas instituições. Ao adentrarem dessa forma múltiplos domínios sociais, econômicos e políticos, desafiam conceitos tradicionais de moralidade, agência e justiça.

Este é o cenário em que o presente número temático se insere. Seu objetivo é iluminar as interações, por vezes permeadas de tensão, entre tecnologia e atividades humanas, convidando os leitores a considerar não apenas o que as IAs podem fazer, mas sobretudo o que devem fazer – bem como o que nós devemos fazer com elas. Essas são as questões que estruturam a *Ética da IA*, um campo recente que busca discutir e formular quadros normativos para a produção, a aplicação e o uso de sistemas de IA nas diversas esferas da atividade humana.

Trata-se de um debate necessariamente transdisciplinar, no qual os esforços de filósofos precisam articular-se aos de cientistas da computação, juristas, profissionais da saúde e especialistas de todas as áreas afetadas pela IA. Embora cada uma dessas disciplinas concentre sua reflexão em aspectos específicos, se deixadas atuar de forma isolada, correm o risco de distorcer o próprio fenômeno que buscam compreender e sobre o qual pretendem intervir – risco que, em alguns casos, já se concretizou.

Além disso, é importante que esse debate seja situado e atento às assimetrias históricas, políticas e econômicas que marcam, por exemplo, o Norte e o Sul globais. As condições de desenvolvimento tecnológico, os regimes regulatórios, as formas de exploração do trabalho e as vulnerabilidades específicas de cada contexto - inclusive do Brasil - influenciam tanto os riscos quanto as oportunidades associados à IA.

Ao reunir perspectivas variadas sobre diferentes dimensões desses desafios, esta edição temática busca incentivar pesquisadores a participar ativamente do debate, além de contribuir para o seu aprofundamento, promovendo uma compreensão mais apurada dos dilemas éticos atuais. Com isso em mente, passemos agora à apresentação das contribuições que compõem este volume.

No artigo "Agora os objetos me percebem — automação, percepção e controle", as autoras **Luciana Nacif** e **Anna Luiza Coli** argumentam que as tecnologias contemporâneas distorcem de modo profundo nossa apreensão da realidade. Isso é feito pela articulação de três quadros críticos: a usurpação do esquematismo em Adorno e Horkheimer, a máquina de visão de Virilio e o presente operacional de Hansen. A partir dessa análise, as autoras caracterizam uma crescente expansão do controle exercido pelo capitalismo tardio, que ultrapassa a dominação ideológica e se mostra capaz de reconfigurar os fundamentos da experiência humana. Contemporaneamente, sistemas computacionais em geral, e a IA em particular, atuam como lentes algorítmicas que filtram o real antes que ele se registre como experiência consciente. Esse cenário levanta questões importantes sobre consentimento e manipulação do comportamento por parte das instituições e empresas que implementam tais sistemas, uma vez que eles refletem e promovem seus interesses financeiros e políticos, bem como seus valores e sua visão de mundo.

Outra preocupação com os impactos da tecnologia aparece no artigo "Virtudes e vícios epistêmicos e o ChatGPT", de **Verônica de Souza Campos**. Sob a lente da epistemologia da virtude, a autora concentra sua análise nos efeitos dos grandes modelos linguísticos — em especial o ChatGPT — sobre nossas virtudes epistêmicas. Embora reconheça benefícios práticos, como rapidez, acessibilidade e apoio pedagógico, argumenta que, na maioria das circunstâncias, a ferramenta tende a fomentar vícios epistêmicos, gerando déficits cognitivos, enfraquecendo a memória e o pensamento crítico, além de perpetuar vieses sociais. Em vez de ampliar a capacidade cognitiva humana, tais sistemas favorecem a passividade, a desonestidade intelectual e a homogeneização do discurso. Desse modo, para que modelos linguísticos contribuam positivamente para o nosso florescimento epistêmico, é imprescindível utilizá-los de forma regulada e reflexiva, preservando os processos cognitivos que cultivam virtudes e mitigando os vícios que a delegação acrítica tende a instaurar.

A epistemologia da virtude também oferece o arcabouço teórico para a análise de **Tiegue Vieira Rodrigues** no artigo "An epistemic virtue-theoretic solution to the problem of AI authorship". Nele, o autor investiga como a prática emergente da escrita distante — a produção textual mediada por uma IA generativa — desafia concepções tradicionais de autoria e de responsabilidade epistêmica. Contrariando a ideia de que o uso de IAs generativas anularia o papel do autor, Rodrigues sustenta que, embora a

geração de texto seja delegada ao modelo, a competência epistêmica humana permanece decisiva em ações de segunda ordem, como a definição de metas, a elaboração de *prompts* e a curadoria do conteúdo produzido. Dessa forma, torna-se possível ao escritor assumir plena responsabilidade e reivindicar autoria, mesmo quando o texto final é produto de uma IA.

Em sua contribuição "Inteligência artificial emocional: uma análise ética", os autores **Marco Aurélio Sousa Alves** e **Ariel Ribeiro Tolentino** examinam a emergência de *chatbots* utilizados como companheiros artificiais íntimos. O texto se dedica ao caso dos *chatbots* relacionais, que conjugam as capacidades de grandes modelos linguísticos com elementos de computação afetiva. Tais sistemas buscam acomodar tanto o reconhecimento de emoções humanas expressas textual ou oralmente quanto a produção de respostas que o usuário reconheça como afetivamente adequadas. Nessa análise, os autores argumentam que há um grande hiato entre as alegadas capacidades empáticas desses sistemas e as capacidades atuais da computação afetiva. Existe, portanto, um profundo desalinhamento entre o modo antropomorfizado como esses sistemas são compreendidos por seus usuários e suas reais capacidades de operar de maneira afetivamente responsável, o que gera diversos riscos — entre eles, o de que a IA venha a incentivar comportamentos antiéticos e/ou prejudiciais ao próprio usuário. Para os autores, mitigar esses riscos exige uma combinação de medidas, tais como a reavaliação do escopo de aplicação desses sistemas e a implementação de uma regulamentação robusta.

No artigo "What is right for 'this' patient? Autonomy, trust and transparency in the era of medical AI", **Anna Smajdor**, **Arnoldo Frigessi** e **Manuela Zucknick** examinam e comparam recomendações terapêuticas geradas por algoritmos de IA e por médicos humanos. Os autores argumentam que, em virtude da opacidade das bases decisórias e da dependência de metas rigidamente definidas, os sistemas atuais comprometem o respeito à autonomia individual. Ou seja, embora os algoritmos possam oferecer maior consistência e explorar grandes volumes de dados, eles falham tanto em incorporar metas de tratamento personalizadas quanto em comunicar, de maneira compreensível, seus objetivos e os níveis de incerteza envolvidos. Conclui-se, portanto, que confiar exclusivamente em uma IA para decisões clínicas viola princípios de autonomia e transparência. Profissionais de saúde devem mediar a interação entre IAs e pacientes, de modo a garantir que o resultado final reflita, de fato, os valores e as preferências individuais de cada pessoa.

A tradicional conexão entre responsabilidade moral e autonomia é problematizada e reavaliada por **Delmo Mattos da Silva** em seu artigo "Autonomia e responsabilidade moral no contexto dos sistemas autônomos". O autor sustenta que autonomia e responsabilidade são conceitos sem conexão intrínseca: é possível ser autônomo sem ser moralmente responsável e, inversamente, ser responsável mesmo sob forte limitação da autonomia.

A partir desta análise, Silva busca mostrar que, ao invés de atribuir intenções morais a máquinas ou de culpar injustamente programadores e usuários sem controle efetivo, nossas práticas de responsabilização devem ser repensadas. Em vez de identificar indivíduos a quem se deve imputar responsabilidade moral, é preciso organizar institucionalmente a gestão de riscos e danos. Sua posição converge, assim, para um modelo de governança que busca redistribuir os custos associados aos erros de agentes artificiais autônomos por meio de mecanismos como corresponsabilidade institucional, seguros obrigatórios, regulação orientada ao risco e mecanismos de prevenção e compensação.

Por sua vez, **Daniel de Luca-Noronha** e **Bruno Dornelas** revisitam um tema fundacional: o que significa afirmar que um sistema computacional é inteligente? O texto "A IA vai a julgamento: as IAs podem ser realmente inteligentes?" mobiliza a metodologia de Garth Hallett — em particular o "Princípio de Semelhança Relativa" (PSR) e a "Norma de Correspondência Linguística" (NCL) — para enfrentar a polarização entre duas posições: de um lado, a rigidez conceitual de Peter Hacker, que rejeita qualquer aplicação não metafórica de predicados psicológicos a máquinas; de outro, a permissividade pragmática de Daniel Dennett, que aceita tal linguagem mediante a adoção da postura intencional. Os autores sustentam que sua abordagem possibilita uma classificação mais nuançada, capaz de evitar os problemas associados a essas posições extremas ao reconhecer, por exemplo, diferentes sentidos de "inteligência" e mesmo a existência de casos indecidíveis. Além disso, concluem que a maior clareza conceitual, resultante dessa análise, constitui uma ferramenta relevante para as discussões contemporâneas no contexto da IA, tais como as que envolvem autoria e responsabilidade.

Dando sequência às discussões sobre os impactos da IA, **Daniel Piñero Rodriguez** e **Jacopo Morelli** argumentam no artigo "From compliance to justification: addressing digital marginalization in constitutional democracies" que muitos dos riscos éticos associados à IA não decorrem de falhas técnicas, mas da consolidação de uma ordem normativa injusta no espaço digital. Polarização, discriminação algorítmica, racismo estatístico e vigilância aparecem como formas de dominação encobertas pela retórica da neutralidade e da eficiência. Para mitigar esses problemas, os autores defendem a existência de um direito moral e político fundamental à justificação: é preciso que as decisões mediadas por IA sejam justificáveis em termos que as pessoas afetadas possam compreender e avaliar criticamente. Sob essa perspectiva, a mera transparência quanto às técnicas utilizadas na construção e implementação desses sistemas é insuficiente para proteger democracias constitucionais, o que desafia uma posição comum entre pesquisadores de IA. Com efeito, é necessário que empresas, desenvolvedores e o Estado respondam publicamente pelas razões que orientam o uso desses sistemas, e que estas respostas sejam capazes de resistir a formas de escrutínio razoável.

Os efeitos da IA sobre o mundo do trabalho são o tema do artigo de **Anastasia Siapka**, intitulado "Reframing AI governance: capabilitarian insights for the future of work". A autora discute o futuro do trabalho a partir da Abordagem das Capacidades, de Amartya Sen e Martha Nussbaum, compreendendo trabalho e lazer como dimensões centrais do florescimento humano. A tese principal de sua proposta é que a automação do trabalho é desejável apenas na medida em que as mudanças ampliam as liberdades substantivas: a capacidade de trabalhar "tal qual um ser humano", de dispor de tempo e meios para o lazer e de participar da vida política. Com isso, Siapka defende a existência de fortes deveres por parte do Estado, das empresas e de órgãos reguladores, no sentido de distribuir os ganhos de produtividade e desenhar leis trabalhistas — bem como instituições — orientadas à expansão de capacidades, e não apenas ao crescimento econômico.

Por fim, o artigo "A primazia da *phronesis*: um modelo formativo para a governança ética", de **Everton de Jesus Silva** e **Fernanda Monzato Machado de Jesus**, se localiza no centro do debate contemporâneo sobre modelos de governança para a IA. No texto, os autores colocam em confronto três tradições éticas clássicas — deontologia, utilitarismo e ética das virtudes aristotélica — e argumentam que apenas esta última oferece um fundamento normativo adequado. A IA não pode ser governada eticamente apenas por regras deontológicas ou por cálculos utilitaristas, pois ambos os modelos, tomados isoladamente, fracassam diante da complexidade e da incerteza inerentes às situações concretas de deliberação moral. Os autores não negam a importância de regras e métricas, mas defendem que elas devem ser subordinadas à *phronesis*, isto é, ao juízo de sujeitos dotados de sabedoria prática, capazes de ponderar conflitos de valores em diferentes contextos. No cenário resultante, a governança da IA deve ser compreendida não apenas como um arranjo regulatório, mas como um empreendimento formativo dos atores envolvidos.

Carlos Henrique Barth
(Pesquisador de pós-doutorado na Faculdade Jesuíta de Filosofia e Teologia – FAJE)

Anna Luiza Coli
(Pesquisadora de pós-doutorado da Cátedra IEAT Darcy Ribeiro, fundada no Instituto de Estudos Avançados Transdisciplinares IEAT/UFMG)

Luciana Nacif
(Pesquisadora de pós-doutorado no Programa de pós-graduação em Filosofia da Universidade Federal de Minas Gerais – UFMG)

Maria Carolina Mendonça de Resende
(Doutora em Filosofia pela UFMG)