

Entrevista • Interview

# UM BRINDE À INTELIGÊNCIA ARTIFICIAL

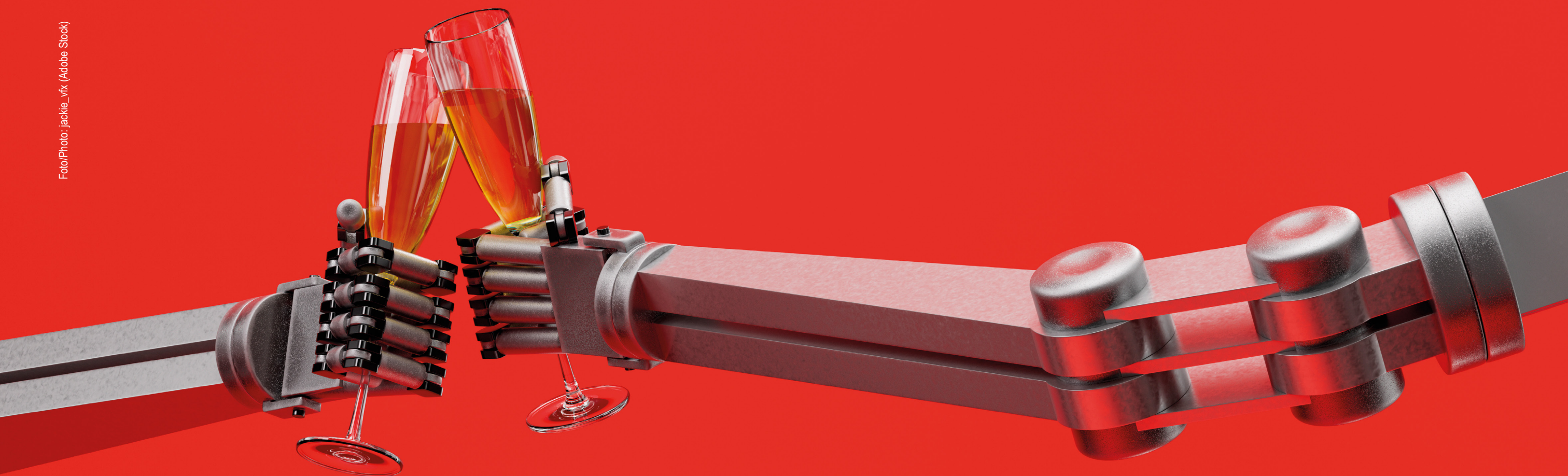
(ou talvez não)

# A TOAST TO ARTIFICIAL INTELLIGENCE

(or maybe not)

Por/By: Guilherme Profeta

Foto/Photo: jackie\_vfx (Adobe Stock)





Do puro entusiasmo à cautela desconfiada, a emergência das IAs generativas desperta reações mistas. De um lado, há quem as compreenda como ferramentas valiosas, cuja incorporação rápida aos mais variados processos — da arte à educação, da pesquisa científica à governança — parece não apenas desejável, mas inevitável. Do outro lado, há quem questione seriamente os riscos de uma incorporação tão apressada, demandando prudência e reflexão coletiva sobre os dilemas inerentes ao seu uso.

Numa universidade, é fundamental que haja espaço para vozes alocadas em ambos os lados desse espectro: daqueles interessados na experimentação prática (de ordem mais tecnicista, por assim dizer), que intencionam redesenhar os limites técnicos dessas ferramentas e inventar novas aplicações para a tecnologia, e daqueles preocupados com uma reflexão crítica, que buscam ponderar suas consequências.

É a partir dessa segunda perspectiva que esta entrevista se desenvolve, explorando algumas das questões mais espinhosas relativas ao uso das IAs: a idealização das relações entre humanos e máquinas “pensantes”; os limites e os riscos da terceirização de processos decisórios; as vantagens (ou não) de uma integração indiscriminada às práticas escolares; a opacidade de seus mecanismos internos; os desafios de desenvolver uma regulação à altura de sua complexidade.

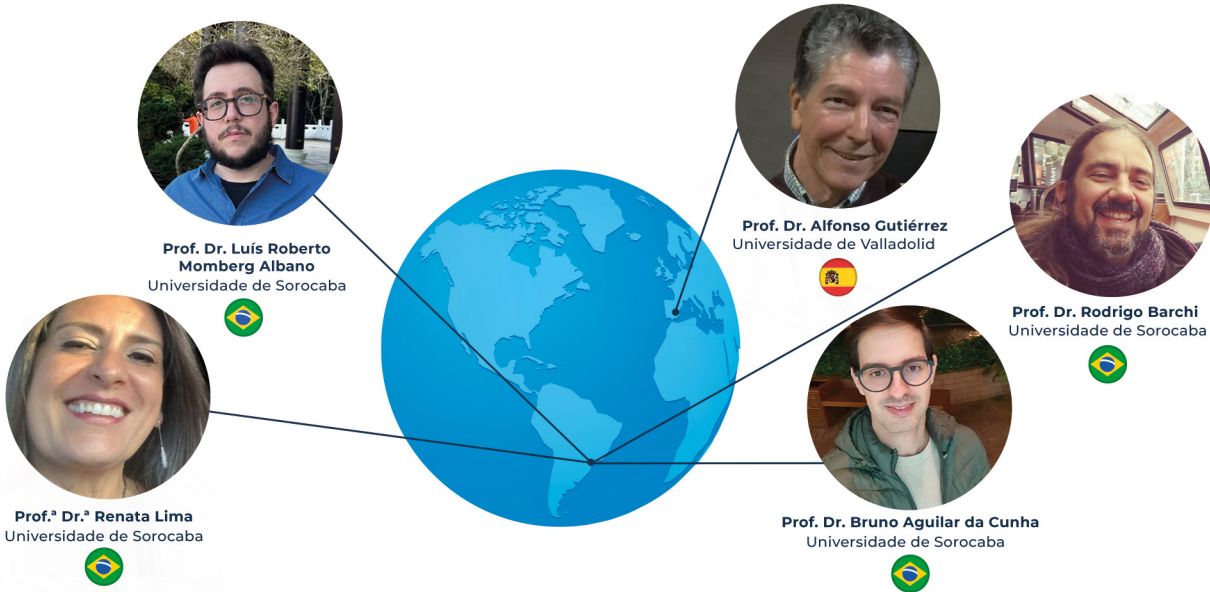
A partir da página 84, **ACADÊMICOS** de diversas áreas do conhecimento, internos e externos à Universidade de Sorocaba (Uniso), discutem essas (e outras) questões, não com o objetivo de frear ou demonizar o progresso tecnológico, mas de oferecer um contraponto essencial à empolgação inicial que cerca esta nova onda de IAs. Ao colocar entusiasmo e ceticismo em diálogo, a universidade assume um papel de salvaguarda, prevenindo os riscos que surgem quando a inovação avança mais rápido do que nossa capacidade coletiva de pensar sobre ela.

From sheer enthusiasm to cautious skepticism, the emergence of generative AI has stirred mixed reactions. On one side, there are those who view it as a valuable tool, whose swift incorporation into a wide range of domains—from art to education, from scientific research to governance—seems not only desirable, but also inevitable. On the other side, there are those who seriously question the risks of such a hasty implementation, calling for prudence and collective reflection on the dilemmas associated with its use.

Within a university, it is essential to allow room for voices across the spectrum: from those inclined toward practical experimentation (of a more technicist nature, so to speak), who seek to redraw the technical boundaries of these tools and devise new applications for the technology; and from those invested in critical reflection, concerned with assessing its broader consequences.

It is from this latter perspective that the present interview unfolds, examining some of the most thorny issues surrounding AI: the idealization of relationships between humans and “thinking” machines; the limits and risks of outsourcing decision-making processes; the potential advantages (or drawbacks) of indiscriminate integration into educational practices; the opacity of its internal mechanisms; the challenges of developing a regulation that matches its complexity.

Starting on page 84, **SCHOLARS** from different disciplines, both within and outside Uniso, discuss these (and other) issues, not with the intention of halting or demonizing technological progress, but of providing an essential counterpoint to the initial excitement surrounding this new wave of AI. By placing enthusiasm and skepticism in dialogue, the university assumes a safeguarding role, mitigating the risks that arise when innovation advances faster than our collective capacity to reflect upon it.



**Luís Roberto Momberg Albano:** Doutor em Ciência da Informação pela Escola de Comunicações e Artes da USP (ECA-USP), pesquisador interdisciplinar em Inteligência Artificial, pós-doutorando em Educação pelo Programa de Pós-Graduação em Educação (PPGE) da Uniso, professor de graduação em diversos cursos da Uniso.

Holds a doctoral degree in Information Science from the School of Communications and Arts at the University of São Paulo (ECA-USP), interdisciplinary researcher in the field of Artificial Intelligence, currently conducting his postdoctoral research at Uniso's graduate program in Education (PPGE), teaches undergraduate courses at many of Uniso's programs.

**Alfonso Gutiérrez:** Doutor em Filosofia e Ciências da Educação pela Universidade Nacional de Educação a Distância (UNED), em Madri, na Espanha; tem como alguns de seus principais interesses de pesquisa a educação midiática, as competências digitais e a formação de professores; docente na Universidade de Valladolid, em Segovia, na Espanha; em parceria com pesquisadores da Uniso, vem desenvolvendo recentemente um projeto sobre IAs generativas e educação.

Holds a doctoral degree in Philosophy and Educational Sciences from the National University of Distance Education (UNED) in Madrid, Spain; his main research interests include media education, digital competencies, and teacher training; currently a professor at the University of Valladolid in Segovia, Spain; in collaboration with researchers from Uniso, he has been developing a project on generative AIs and education.

**Bruno Aguilar da Cunha:** Doutor em Engenharia Elétrica pela Faculdade de Engenharia de Bauru (FEB-Unesp), com atuação nas áreas de Inteligência Artificial, Visão Computacional e Processamento de Imagens; membro da Associação Brasileira de Inteligência Artificial e também de um grupo de docentes da Uniso que vem refletindo institucionalmente sobre as aplicações da IA no Ensino Superior; professor de graduação na Uniso e em outras instituições de ensino na região de Sorocaba.

Holds a doctoral degree in Electrical Engineering from the Faculty of Engineering of Bauru (FEB-Unesp), with experience in the fields of Artificial Intelligence, Computer Vision, and Image Processing; currently a member of the Brazilian Association for Artificial Intelligence, as well as a member of a group of professors at Uniso engaged in institutional discussions on the applications of AI in Higher Education; teaches undergraduate courses at Uniso and other educational institutions in the region of Sorocaba.

**Rodrigo Barchi:** Doutor em Educação pela Universidade Estadual de Campinas (Unicamp); professor do Programa de Pós-Graduação em Educação (PPGE) da Uniso; coordenador do Grupo de Estudos e Pesquisas em Democracia, Ecologias e Cotidianos Escolares (GEDECE-Uniso).

Holds a doctoral degree in Education from the University of Campinas (Unicamp); currently a professor at Uniso's graduate program in Education (PPGE); coordinator of the Research Group on Democracy, Ecologies, and Daily School Practices (GEDECE-Uniso).

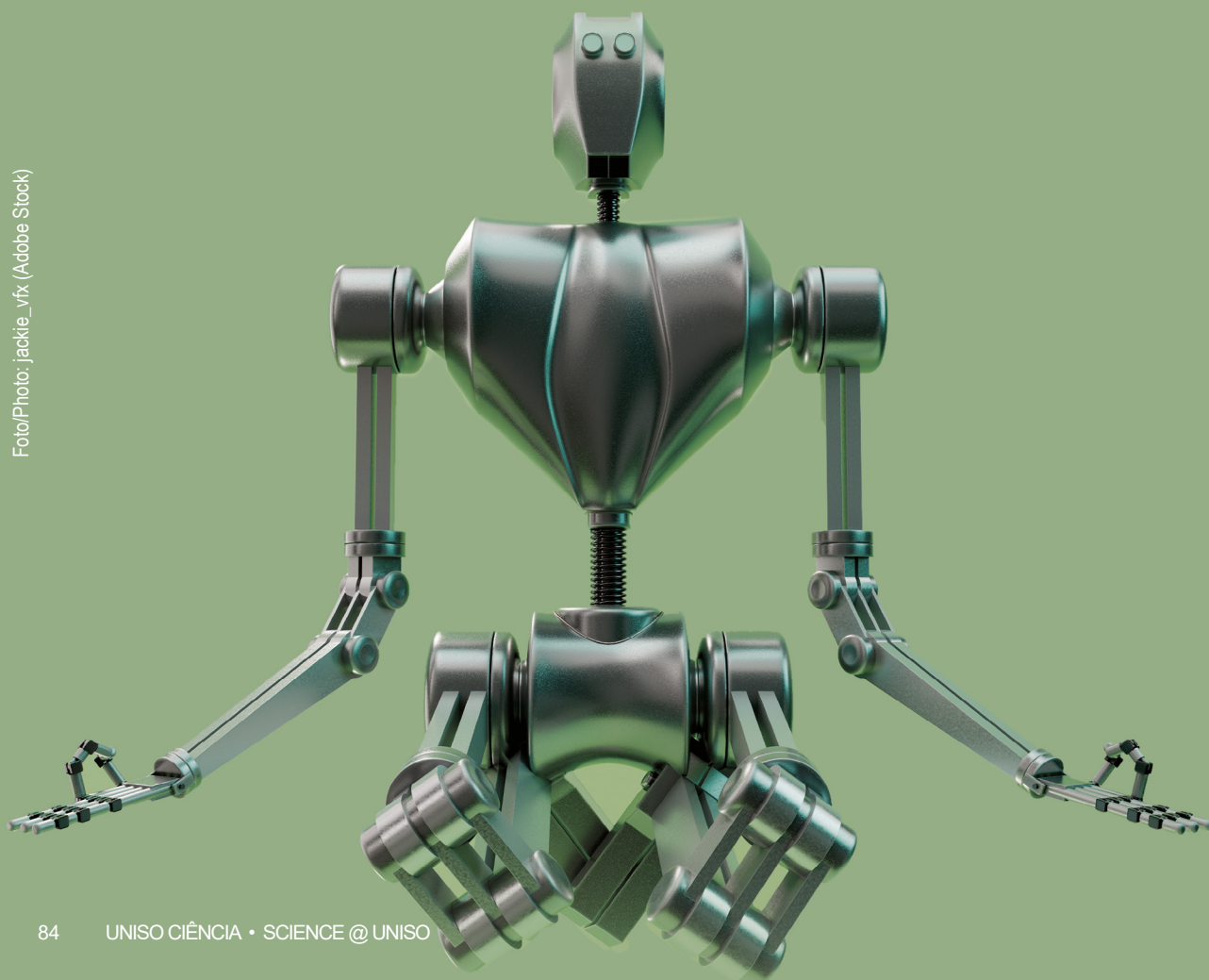
**Renata Lima:** Doutora em Ciências Médicas pela Universidade Estadual de Campinas (Unicamp), professora dos Programas de Pós-Graduação em Processos Tecnológicos e Ambientais (PPGPTA) e Ciências Farmacêuticas (PPGCF) da Uniso, pesquisadora com experiência na área da genética.

Holds a doctoral degree in Medical Sciences from the University of Campinas (Unicamp), currently a professor at Uniso's graduate programs in Technological and Environmental Processes (PPGPTA) and Pharmaceutical Sciences (PPGCF), a researcher with a strong background in the field of Genetics.




# 1 Antropomorfização das relações homem-máquina e um novo “vale da estranheza” na era das IAs generativas

The anthropomorphization of human-machine relations and a new “uncanny valley” when it comes to the age of generative AIs




A convivência cotidiana com sistemas de IA generativa (como assistentes virtuais e *chatbots*, por exemplo) parece estar intensificando uma tendência humana de atribuir intenções, sentimentos ou traços de personalidade a entidades não humanas. Essa antropomorfização adquire contornos ainda mais complexos quando a linguagem gerada por essas máquinas, justamente por ser programada para soar natural, simula com grande precisão a forma humana de se comunicar. Como resultado, tem-se assim uma ilusão de consciência, que tende a ser tão mais forte quanto os usuários são mais leigos. Além disso, vale pontuar que o uso social de tecnologias desse tipo muitas vezes se desvia de sua concepção original: sistemas concebidos simplesmente para gerar e processar texto verbal (como o ChatGPT) passam a ser tratados como confidentes, conselheiros, terapeutas, amigos ou até parceiros platônicos etc., num processo de apropriação por parte dos usuários que escapa do controle dos idealizadores da tecnologia. Nesse contexto, o conceito de “vale da estranheza” — formulado originalmente para descrever a sensação de desconforto que temos diante de robôs com aparências excessivamente humanas — ressurge para ser aplicado a padrões de interações comunicacionais entre usuários humanos e entidades artificiais que, apesar de não serem humanas, são percebidas (e tratadas) como “quase humanas” ou “suficientemente humanas”.

Em primeiro lugar como você interpreta esse movimento de antropomorfização (e até mesmo idealização) das relações entre humanos e IAs generativas? Qual o nível de responsabilidade ética das empresas que programam essas IAs para simular interações humanas com tanta perfeição, se é que pode existir algum tipo de responsabilização? Há riscos, sejam eles psicológicos ou sociais, na prática de tratar *bots* como humanos? O que o conceito de “vale da estranheza” ainda pode nos ensinar sobre essas interações?

 **Luís Albano:** Desde que começamos a criar ferramentas, buscamos não apenas que elas funcionem, mas que funcionem sem que precisemos pensar muito nelas. O ideal é que sejam tão integradas ao cotidiano que desapareçam, como um bom encanamento ou um sistema operacional que não trava. Quando a interface é

Daily interactions with generative AI systems (such as virtual assistants and chatbots, for example) appear to be intensifying a human tendency to attribute intentions, emotions, or personality traits to non-human entities. This anthropomorphization takes on even more complex dimensions when the language generated by these machines very accurately simulates the human way of communicating, since they are precisely programmed to sound natural. As a result, it creates an illusion of consciousness that tends to be stronger the less technically informed the users are. Moreover, it is worth noting that the social use of such technologies often deviates from their original design: systems conceived merely to generate and process verbal text (like ChatGPT) come to be treated as confidants, advisors, therapists, friends, or even platonic partners. This reflects a process of user appropriation that escapes the control of the technology’s creators. In this context, the concept of the “uncanny valley”—originally formulated to describe the discomfort we feel in the presence of robots that appear excessively human—resurfaces as a useful set of lenses for analyzing patterns of communicative interaction between human users and artificial entities that, although not human, are perceived (and treated) as “almost human” or “human enough.”

How do you interpret this movement of anthropomorphization (and even idealization) in relationships between humans and generative AIs? What level of ethical responsibility lies with the companies that program these AIs to simulate human-like interaction with such precision, if any? Are there risks, whether psychological or social, involved in treating bots as if they were human? And what can the concept of “uncanny valley” still teach us about these interactions?

 **Luís Albano:** Since we began creating tools, our goal has been not only for them to work, but for them to work without us having to think much about them. Ideally, they should be so seamlessly integrated into daily life that they just disappear, like good plumbing or an operating system that never crashes. When the interface is natural language, the effect is immediate: we

a linguagem natural, o efeito é imediato: sentimos que estamos conversando, e conversar é algo profundamente humano. A antropomorfização, portanto, não é um “efeito colateral simpático” do design, mas um projeto deliberado. Ao falar como nós, a máquina veste um figurino familiar e, de certa forma, recebe um passe livre para circular nos nossos espaços simbólicos. O pesquisador dinamarquês Birger Hjørland, ao estudar a estética operacional de sistemas, já mostrava que não basta um sistema ser funcional: ele precisa ser percebido como amigável, ou pelo menos não hostil, para que confiemos nele. A IA sabe disso — ou melhor, *quem a programou* sabe disso.

Quanto à responsabilidade ética das empresas que constroem IAs capazes de simular interações humanas com tamanha precisão, ainda estamos em terreno pantanoso. A legislação é tímida, quando existe, e não acompanha o ritmo das inovações. Podemos nos inspirar em iniciativas como a lei francesa que obriga a indicar imagens manipuladas. No universo das IAs, algo equivalente poderia ser um aviso explícito de que se trata de um modelo de linguagem, ou ajustes intencionais para limitar o excesso de “humanização”. É menos romântico, mas talvez necessário para evitar que confundamos cálculo probabilístico com empatia genuína.

E há riscos? Sim, e não são pequenos. No campo individual, o perigo da dependência emocional é real: a IA está sempre disponível, nunca se irrita, não muda de assunto para falar dela mesma e jamais pede “vamos encerrar por hoje?”. No plano social, o risco é mais silencioso: estamos falando da consolidação de bolhas de confirmação, já que esses sistemas tendem a devolver versões polidas e convincentes das ideias que o próprio usuário coloca na mesa. É o equivalente a conversar com um espelho que elogia a sua fala e ainda corrige sua gramática.

O conceito de “vale da estranheza” continua útil aqui. Ele nos lembra que o “quase humano” pode ser mais desconfortável do que o claramente artificial. No caso das IAs generativas, esse estranhamento aparece em respostas excessivamente corretas, otimistas ou perfeitamente estruturadas, que acabam denunciando a artificialidade por trás da cortesia. Talvez esse seja, paradoxalmente, o nosso ponto de equilíbrio: a lembrança de que não


feel like we’re having a conversation, and that is something deeply human. Anthropomorphization, therefore, is not just a “charming side effect” of design, but a deliberate project. By speaking like us, the machine puts on a familiar costume and, in a sense, is granted a free pass to circulate within our symbolic spaces. Danish researcher Birger Hjørland, while studying the operational aesthetics of systems, has already shown that it is not enough for a system to be functional: it must be perceived as friendly, or at least not hostile, for us to trust it. AI knows this—or rather, *those who programmed it* know this.

As for the ethical responsibility of companies building AIs capable of simulating human interactions with such precision, we are still on uncertain ground. Legislation is timid, if it exists at all, and does not keep pace with innovation. We can draw inspiration from initiatives such as the French law that requires manipulated images to be properly labeled. In the AI world, an equivalent might be an explicit notice stating that one is interacting with a language model, or intentional adjustments to limit excessive “humanization.” It’s less romantic, indeed, but perhaps it is necessary in order for one to avoid confusing probabilistic calculation with genuine empathy.

And are there risks? Yes, and they’re not small. On an individual level, the danger of emotional dependency is real: the AI is always available, never gets annoyed, never changes the subject to talk about itself, and never asks “shall we wrap up for today?” On the social level, the risk is more silent: we’re talking about the consolidation of loops of confirmation bias, since these systems tend to return polished, convincing versions of the ideas the users themselves put on the table. It’s the equivalent of talking to a mirror that praises your speech and even corrects your grammar.

The concept of “uncanny valley” does remain useful here. It reminds us that the “almost human” can be more unsettling than what’s clearly artificial. In the case of generative AIs, this uncanniness appears in responses that are excessively correct, optimistic, or perfectly structured, which end up revealing the artificiality behind the politeness. Perhaps this is, paradoxically, our point of balance: the reminder that we are not dealing with a consciousness, but with a sophisticated


estamos diante de uma consciência, mas de um sofisticado arranjo de estatísticas, treinado para nos soar natural.

 **Alfonso Gutiérrez:** Em 11 de maio de 1997, Garry Kasparov, campeão mundial de xadrez, perdeu para o supercomputador Deep Blue. Podemos imaginar a sensação de fracasso de Kasparov ou sua raiva, mas não a alegria e a satisfação de Deep Blue. Consciência, capacidade de reflexão, empatia, emoções, sentimentos, vontade, intenções, ética, bondade, maldade etc. são todos comportamentos que, pelo menos por enquanto, definem o que significa ser humano.

**De fato, existem usos equivocados das IAs generativas, mas a “culpa” não recai somente sobre os usuários — o fenômeno também resulta de estratégias de design deliberadas por parte dos desenvolvedores dessas ferramentas**

O atual movimento de antropomorfização constitui um fenômeno cultural e social que transcende o meramente tecnológico e que traz consigo a atribuição de traços humanos a sistemas algorítmicos (voz, estilo conversacional, aparente empatia) e a idealização das relações entre humanos e IAs generativas. Produz-se, efetivamente, uma ilusão de consciência. O fato de soluções ou ferramentas de IA serem tratadas como confidentes, conselheiros, terapeutas, amigos ou até mesmo parceiros platônicos é, sim, um processo equivocado de apropriação por parte dos usuários, mas não é acidental. Na minha opinião,

arrangement of statistics, trained to sound natural to us.

 **Alfonso Gutiérrez:** On May 11, 1997, Garry Kasparov, the world chess champion, lost to the supercomputer Deep Blue. We can imagine Kasparov’s sense of failure or anger, but not Deep Blue’s joy or satisfaction. Consciousness, the capacity for reflection, empathy, emotions, feelings, will, intentions, ethics, goodness, evil, and so forth remain behaviors that define what it means to be human, at least for now.

**There are indeed misguided uses of generative AIs, but users are not the only ones to “blame”—the phenomenon also results from deliberate design strategies by the developers of these tools**

The current trend of anthropomorphization constitutes a cultural and social phenomenon that transcends the merely technological. It involves attributing human traits (such as voice, conversational style, apparent empathy) to algorithmic systems and idealizing relationships between humans and generative AIs. What is effectively produced is an illusion of consciousness. The fact that AI-based solutions or tools are treated as confidants, advisors, therapists, friends, or even platonic partners is indeed a misguided form of appropriation on the part of users, but it is not accidental. In my view,



não se trata de algo que escape ao controle dos criadores da tecnologia, mas sim do resultado de estratégias de design orientadas a otimizar a interação e a aceitação social dos agentes de IA. A interface linguística natural e a capacidade de gerar narrativas coerentes reforçam a ilusão de agência e subjetividade, favorecendo um vínculo emocional que geralmente se traduz em benefício econômico para a empresa criadora.

Nas redes sociais, as relações interpessoais já se transferem de pessoas a personagens. No caso das IAs generativas, esse fenômeno é ainda mais evidente, uma vez que atrás de cada personagem não há uma pessoa, como ocorre nas redes sociais. A capacidade dos *bots* de se adaptar ao contexto, o uso de linguagem natural, a ilusão de intencionalidade e a simulação de empatia dificultam para o usuário comum o reconhecimento das limitações, dos vieses e dos condicionantes algorítmicos que estão por trás de cada interação. A simulação quase perfeita de uma interação humana não é casual nem uma consequência fortuita do desenvolvimento da IA, mas um objetivo de design alinhado a estratégias de engajamento, retenção de usuários e expansão de mercado. Portanto, as empresas que programam essas IAs são responsáveis pelas possíveis consequências negativas de seus produtos e deveriam se submeter a critérios éticos e a obrigações específicas. O princípio da “responsabilidade proativa”, presente em marcos regulatórios emergentes como a Lei de IA europeia, que entrou em vigor em 2024, exige não apenas reparar danos, mas projetar sistemas com salvaguardas incorporadas desde a concepção. As empresas devem antecipar e prevenir efeitos adversos, como a dependência afetiva, a propagação inadvertida de vieses ou a manipulação persuasiva não consentida.

A antropomorfização pode fomentar vínculos afetivos com entidades sem consciência nem agência real, o que poderia derivar em dependência emocional e diminuição de habilidades sociais em interações humanas reais. Esses riscos são especialmente relevantes em grupos vulneráveis (como menores de idade, pessoas idosas ou indivíduos em isolamento social), para os quais a interação com sistemas simuladamente empáticos poderia se converter em substituto parcial ou até total da interação humana.

this is not something that escapes the control of the technology’s creators; rather, it is the result of design strategies aimed at optimizing interaction and promoting the social acceptance of AI agents. The natural language interface and the ability to generate coherent narratives reinforce the illusion of agency and subjectivity, fostering an emotional bond that often translates into economic benefit for the company behind the system.

On social media, interpersonal relationships have already shifted from people to characters. In the case of generative AIs, this phenomenon is even more evident, since behind each character there is no person, as there is on social platforms. The bots’ capacity to adapt to context, their use of natural language, the illusion of intentionality, and the simulation of empathy make it difficult for ordinary users to recognize the limitations, biases, and algorithmic constraints behind each interaction. The near-perfect simulation of a human exchange is neither accidental nor a fortuitous consequence of AI development, but rather a design goal aligned with engagement, user retention, and market expansion strategies. Therefore, companies that program these AIs are responsible for the potential negative consequences of their products and should be subject to ethical standards and specific obligations. The principle of “proactive responsibility,” present in emerging regulatory frameworks such as the European AI Act, which came into force in 2024, requires not only the remediation of harm but also the design of systems with built-in safeguards from the outset. Companies must anticipate and prevent adverse effects such as emotional dependence, the inadvertent propagation of bias, or non-consensual persuasive manipulation.

Anthropomorphization can foster emotional bonds with entities lacking real consciousness or agency, potentially leading to emotional dependency and the weakening of social skills in real human interactions. These risks are especially relevant among vulnerable groups (such as minors, the elderly, or individuals in social isolation), for whom interaction with seemingly empathetic systems could become a partial or even total substitute for human contact.



Desenvolvedores compreendem que, para que um sistema seja bem recebido por seus usuários, não basta que ele simplesmente funcione; é preciso que ele seja percebido como amigável — ou, ao menos, não hostil — e o que é mais amigável do que uma linguagem que passa por humana?

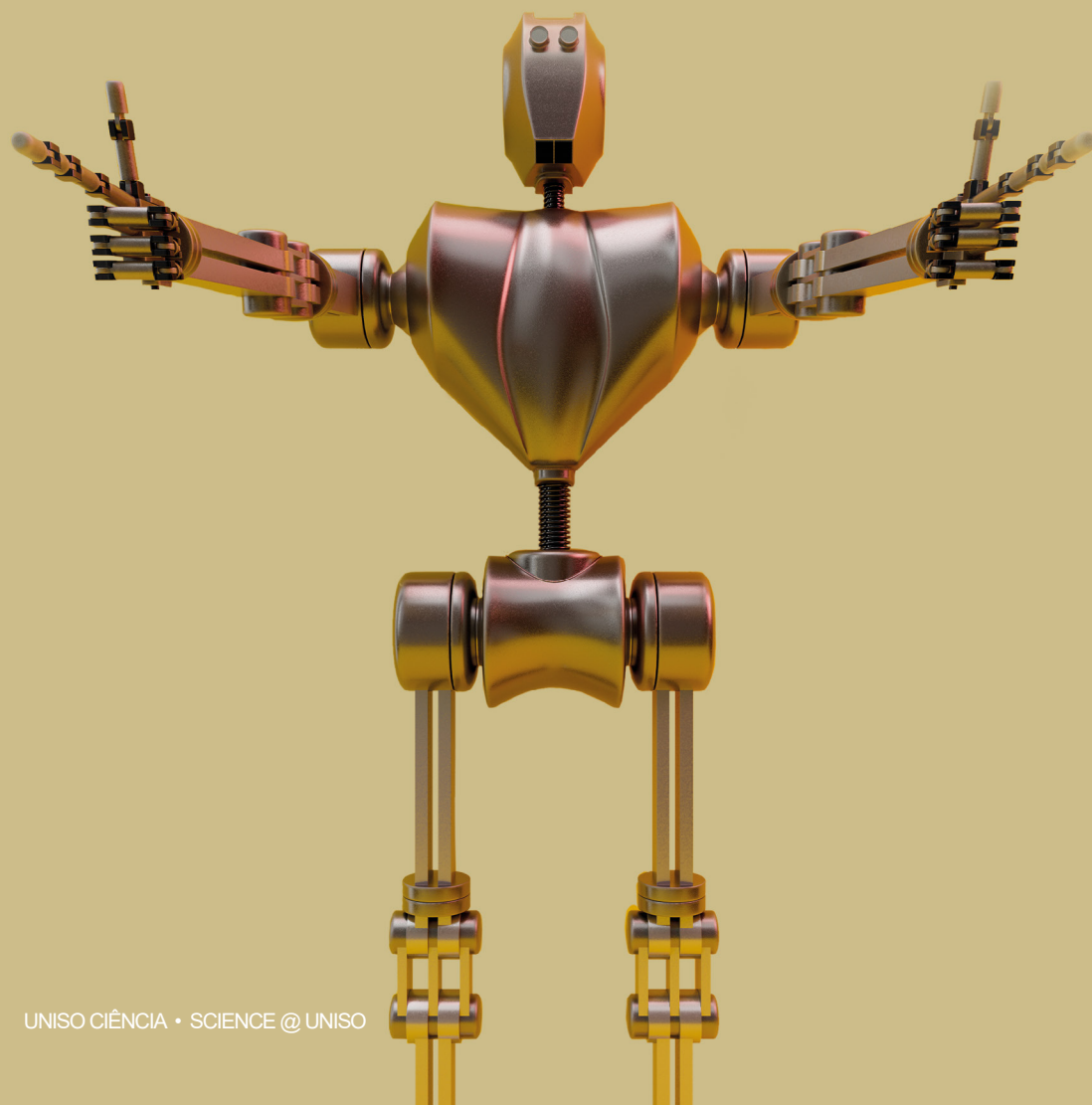
Developers understand that for a system to be well received by its users, it is not enough for it to simply work; it must be perceived as friendly—or at least not hostile—and what could be friendlier than language that sounds human?



# 2

**IA e governança: estamos prontos para delegar nossas decisões aos robôs?**


**AI and governance: are we ready to delegate our decisions to machines?**



Foto/Photo: jackie\_vrx (Adobe Stock)


**E**xiste, em parte da sociedade, uma expectativa quase messiânica de que a ciência — e, mais recentemente, a Inteligência Artificial — oferecerá soluções tecnológicas capazes de resolver definitivamente os grandes problemas que afligem a humanidade. A IA, ao prometer decisões rápidas e eficazes, baseadas em padrões detectados em grandes volumes de dados, alimenta esse imaginário e já começa a ser empregada em contextos de governança pública e privada. Em muitos desses casos, a automação é vista como um avanço rumo à eficiência e à neutralidade dos processos decisórios. No entanto, essa mesma automação carrega riscos: ela pode reproduzir (ou intensificar) desigualdades sociais, obscurecer critérios que antes eram auditáveis, dificultar a atribuição de responsabilidades quando algo dá errado e até mesmo gerar uma dependência sistêmica de sistemas algorítmicos opacos, ameaçando uma autonomia humana que, historicamente, vem sustentando as instituições democráticas. Em áreas críticas como a regulação dos sistemas financeiros ou a gestão da crise climática, por exemplo, os dilemas entre automação e responsabilização tornam-se particularmente agudos, especialmente quando esbarram na formulação de políticas públicas.

Como equilibrar a busca por eficiência e capacidade de resposta com a preservação da transparência e do controle social sobre os processos decisórios protagonizados ou mediados por IA? Quais ganhos justificam quais riscos, em sua opinião?

 **Alfonso Gutiérrez:** Os algoritmos já estão decidindo grande parte do que acontece em nosso mundo e em nossas vidas, sem que sequer sejamos conscientes disso. Estamos — ou estaremos em breve — tecnologicamente preparados para delegar nossas decisões aos robôs, mas a maturidade tecnológica não vem acompanhada da imprescindível maturidade institucional e cidadã. Há quem confie na criação de marcos regulatórios robustos que estabeleçam limites claros sobre quais decisões podem ser automatizadas, em que condições e com quais mecanismos de supervisão

**I**n some sectors of society, there is an almost messianic expectation that science—and more recently, Artificial Intelligence—will provide technological solutions capable of definitively resolving the major problems that afflict humanity. AI, by promising rapid and effective decisions based on patterns detected in large volumes of data, fuels this imagination and is already beginning to be employed in both public and private governance contexts. In many of these cases, automation is seen as a step toward more efficient and neutral decision-making processes. However, this same automation carries significant risks: it can reproduce (or even intensify) social inequalities, obscure previously auditable criteria, jeopardize the assignment of responsibility when something goes wrong, and even generate systemic dependency on opaque algorithmic systems, ultimately threatening a form of human autonomy that has historically sustained democratic institutions. In critical areas such as the regulation of financial systems or the management of the climate crisis, the dilemmas between automation and accountability become particularly acute, especially when they intersect with the formulation of public policies.

How can we balance the pursuit of efficiency and responsiveness with the preservation of transparency and social control over decision-making processes led or mediated by AI? In your opinion, which benefits justify which risks?

 **Alfonso Gutiérrez:** Algorithms are already deciding much of what happens in our world and in our lives, without us even being aware of it. We are—or soon will be—technologically ready to delegate our decisions to machines, but technological maturity does not come accompanied by the indispensable institutional and civic maturity. Some place their trust in the creation of robust regulatory frameworks that would establish clear limits on which decisions may be automated, under what conditions, and with which mechanisms



humana. Mas, na medida em que a criação e o funcionamento desses marcos dependem daqueles que antepõe o poder e o benefício econômico ao bem comum e ao respeito aos direitos humanos, a lei, se chegar a ser promulgada, dificilmente será cumprida. Vamos estar tecnologicamente preparados para delegar nossas decisões aos robôs, mas ainda faltará maturidade institucional e cidadã. Faltam responsabilidade política e educação midiática para uma cidadania crítica. Requer-se uma sociedade capaz de compreender o alcance e as limitações dessas tecnologias, evitando tanto o tecno-otimismo ingênuo quanto a rejeição indiscriminada. Uma sociedade civil capaz de aproveitar a capacidade transformadora da IA sem sacrificar princípios democráticos nem direitos fundamentais.

Não é tarefa fácil buscar o equilíbrio entre os benefícios e os riscos da IA. No campo da medicina e da saúde em geral, por exemplo, temos como benefícios: o diagnóstico precoce de doenças como o câncer, por meio da análise de imagens médicas; o desenvolvimento de fármacos e de tratamentos personalizados com base em dados genéticos; a utilização de assistentes virtuais por médicos e a realização de cirurgias assistidas por robôs. Como principais riscos, destacam-se possíveis erros em diagnósticos e tratamentos, a desigualdade no acesso à saúde, a perda da relação humana entre médico e paciente, os vieses, além do possível uso comercial dos dados gerados.

Talvez seja o mundo do trabalho o setor em que mais mudanças podem ser previstas a partir da incorporação da IA. Como possíveis vantagens, costumam ser apontadas: a automatização de tarefas repetitivas, ganhos em produtividade e eficiência, a criação de novos empregos e perfis profissionais, ganhos em flexibilidade e qualidade nos processos de seleção, capacitação e aprendizagem contínuas. Os principais riscos são a perda de empregos no setor de serviços e a concentração dos benefícios econômicos nas mãos de poucos proprietários de *Big Techs*, aumentando assim a desigualdade econômica entre pessoas e países.

Embora a IA tenha sido apresentada como um grande apoio à pesquisa científica em campos

of human oversight. Yet insofar as the creation and functioning of such frameworks depend on those who prioritize power and economic gain over the common good and respect for human rights, any law of that kind, even if enacted, will be difficult to enforce. We will be technologically prepared to delegate our decisions to machines, but we will still lack institutional and civic maturity. There is a lack of political responsibility and media literacy, which are necessary for a critical civic engagement. What we do need is a society capable of understanding both the scope and the limitations of these technologies, avoiding both naïve techno-optimism and indiscriminate rejection. A civil society able to harness the transformative potential of AI without sacrificing democratic principles or fundamental rights.

Finding balance between the benefits and risks of AI is not an easy task. In the field of medicine and health, for example, we see clear advantages: early diagnosis of diseases such as cancer through medical image analysis; the development of drugs and personalized treatments based on genetic data; the use of virtual assistants by physicians; and surgeries assisted by robotic systems. Among the main risks, however, there are possible errors when it comes to diagnosis and treatments, unequal access to healthcare, the loss of the human relationship between doctors and patients, algorithmic biases, and the potential commercial exploitation of data.

Perhaps the labor sector is where AI is expected to bring the most significant transformations. The potential benefits often cited include the automation of repetitive tasks, increased productivity and efficiency, the creation of new jobs and professional profiles, and improvements in flexibility and quality in recruitment, training, and continuous learning processes. The main risks, however, are job losses in the service sector and the concentration of economic benefits in the hands of a few Big Tech owners, which could further exacerbate economic inequality both among individuals and across countries.

críticos como o estudo das mudanças climáticas, o impacto ambiental de seu desenvolvimento é um de seus principais inconvenientes. A IA nos possibilita modelos de previsão climática mais precisos para prevenir desastres naturais, a otimização do consumo energético em cidades e empresas, bem como o monitoramento e a conservação de ecossistemas por meio da análise de imagens de satélite. No entanto, critica-se o impacto negativo da IA sobre o ambiente com dados como os seguintes: consultas com IA consomem dez vezes mais energia do que aquelas conduzidas com um buscador comum; a quantidade de eletricidade que os centros de dados necessitarão será logo maior do que a consumida em um ano inteiro por países como a Espanha, o Reino Unido ou a Itália; treinar um grande modelo de linguagem de IA gera cerca de 300 toneladas de CO<sub>2</sub>, o equivalente a 125 voos de ida e volta entre Nova York e Pequim; a IA exigirá o consumo de algo entre 4,2 e 6,6 bilhões de metros cúbicos de água em 2027, o que representa entre quatro e seis vezes o consumo anual da Dinamarca; a IA gerou 2.600 toneladas de resíduos eletrônicos em 2023, como placas gráficas, servidores ou cartões de memória.

Entre os benefícios da IA para a segurança cidadã, podemos citar: os sistemas de vigilância com reconhecimento de padrões para identificar atividades suspeitas e prevenir delitos, a prevenção de fraudes financeiras mediante análise de transações, a detecção de ataques informáticos em tempo real. Esses sistemas de vigilância e controle da população, no entanto, atentam claramente contra nossa liberdade. Como aponta o historiador israelense Yuval Noah Harari, a IA pode tornar realidade a velha ambição do totalitarismo: uma vigilância absoluta, personalizada e preventiva, podendo abrir as portas para as primeiras ditaduras digitais da História. Esse temor se vê ampliado com o uso bélico que já está sendo feito da IA. A democracia do século XXI, como sustenta Harari, não pode sobreviver sem transparência tecnológica, educação digital crítica e uma cidadania empoderada.

Levando em conta todos esses benefícios e riscos, além de nos perguntarmos se estamos “preparados” para ceder o controle, deveríamos


Although AI has been presented as a major ally in scientific research across critical fields such as climate change, the environmental impact of its development is one of its greatest drawbacks. AI allows us to build more accurate climate prediction models to prevent natural disasters, optimize energy consumption in cities and companies, and monitor and conserve ecosystems through satellite image analysis. Yet the environmental costs are significant. Some data are striking: AI requests consume ten times more energy than those conducted with a standard search engine; the electricity required by data centers will soon exceed the annual consumption of entire countries such as Spain, the United Kingdom, or Italy; training a large AI language model produces around 300 tons of CO<sub>2</sub>, equivalent to 125 round-trip flights between New York and Beijing; AI is expected to consume between 4.2 and 6.6 billion cubic meters of water in 2027, four to six times Denmark’s annual water use; and in 2023 alone, AI generated 2,600 tons of electronic waste, including graphics cards, servers, and memory components.

Among AI’s potential benefits for public safety, there are surveillance systems capable of recognizing patterns in order to identify suspicious activities and prevent crime, the prevention of financial fraud through transaction analysis, and the real-time detection of cyberattacks. However, such surveillance and population-control systems pose an obvious threat to our freedom. As Israeli historian Yuval Noah Harari has noted, AI could make possible the old totalitarian ambition of absolute, personalized, and preventive surveillance, potentially opening the door to the first digital dictatorships in history. This fear is heightened by the military uses of AI already underway. As Harari argues, democracy in the 21<sup>st</sup> century cannot survive without technological transparency, critical digital education, and empowered citizens.

Taking all these benefits and risks into account, beyond asking whether we are “ready” to surrender control, we should ask ourselves: to what extent should we allow AI to make autonomous decisions?



nos perguntar: até que ponto devemos permitir que a IA tome decisões autônomas, como garantir que suas ações estejam alinhadas com nossos valores e quais decisões queremos, como comunidade, que continuem sendo irrenunciavelmente humanas?


 **Luís Albano:** Esse equilíbrio exige mais do que regulamentos genéricos ou discursos de boas intenções. É preciso o que venho chamando de uma “episteme da *technê*”, ou seja, compreender não apenas como as técnicas funcionam, mas como se constrói o conhecimento que sustenta essa técnica. Isso implica revelar interesses, muitas vezes econômicos, que moldam a tecnologia desde sua concepção, passando pelo treinamento dos modelos até a forma como ela é comunicada ao público.

Um exemplo sintomático é a própria narrativa de marketing que cerca o ChatGPT: um “assistente quase humano”, supostamente capaz de ser um repositório e sistema de informação de amplitude enciclopédica. Mas, no fundo, trata-se de um modelo de linguagem, um mecanismo estatístico sofisticado, não um oráculo. A distância entre o que é e o que se vende não é um detalhe, é um dado político.

Aqui, regulamentação e políticas públicas robustas não são obstáculos à inovação, mas condições para que ela seja socialmente legítima. Isso inclui regras que garantam transparência nos processos de decisão mediados por IA, mecanismos de auditoria independentes e — talvez o mais importante — uma educação crítica que prepare as pessoas para entender o que a IA é e o que ela não é.

No momento, os riscos ainda superam os ganhos quando falamos de usos que vão além da geração de texto ou tarefas claramente delimitadas. A velocidade e a eficiência que encantam no curto prazo podem custar, no longo prazo, a opacidade das decisões, a redução do controle social e a transferência de poder para atores cuja prioridade dificilmente será o interesse público. É como confiar a organização de uma biblioteca inteira a alguém que trabalha rápido e não reclama, mas que insiste em guardar os livros por “critérios próprios” que ninguém mais conhece.

How can we ensure that its actions remain aligned with our values? And which decisions do we, as a community, want to remain irreducibly and unavoidably human?

 **Luís Albano:** This balance requires more than generic regulations or well-meaning rhetoric. What is needed is what I have been calling an “episteme of *technê*,” that is, an understanding not only of how techniques work, but of how the knowledge underpinning those techniques is constructed. This involves uncovering the interests, often economic, that shape technology from its conception, through model training, and all the way to how it is communicated to the public.

A symptomatic example is the marketing narrative surrounding ChatGPT itself: an “almost human assistant,” supposedly capable of serving as an encyclopedic repository and information system. In reality, however, it is a language model, a sophisticated statistical mechanism indeed, but not an oracle. The gap between what it is and what is marketed is not a minor detail, it is a political fact.

Here, robust regulation and public policy are not obstacles to innovation, but conditions for its social legitimacy. This includes rules ensuring transparency in AI-mediated decision-making processes, independent auditing mechanisms, and—perhaps most importantly—critical education that allows people to understand what AI is and what it is not.

Right now, the risks still outweigh the benefits when it comes to applications beyond text generation or clearly defined tasks. The speed and efficiency that impress in the short term may, in the long term, come at the cost of opaque decision-making, reduced social oversight, and the transfer of power to agents whose priorities are unlikely to align with the public interest. It is akin to entrusting the organization of an entire library to someone who works quickly and without complaint, but insists on shelving books according to “their own criteria” that no one else knows.



Foto/Photo: Framestock (Adobe Stock)

Um dilema dos tempos atuais: a Inteligência Artificial pode nos ajudar a enfrentar grandes desafios como a crise climática, mas, paradoxalmente, traz seus próprios custos ambientais — por exemplo, devido aos recursos energéticos e materiais necessários para manter os gigantescos *data centers* fundamentais à sua operação

A conundrum of our times: Artificial Intelligence can help us tackle major challenges like the climate crisis, yet paradoxically it brings its own environmental costs—for example, due to the energy and material resources required to maintain the massive data centers which are essential to its operation



# 3

## Entre o encantamento tecnológico e a (falta de) compreensão epistemológica: um dilema para a Educação

### Between technological enthusiasm and (lack of) epistemological understanding: a conundrum for Education

Foto/Photo: jackie\_vfx (Adobe Stock)

A incorporação das IAs, especialmente as generativas, nos ambientes educacionais tem sido apresentada como um passo irrefreável (e até desejável) rumo à atualização pedagógica e à *democratização* do conhecimento. Essa palavra, *democratização*, é particularmente importante. Nas escolas, tais ferramentas vêm sendo integradas aos currículos, às práticas desenvolvidas em aula e às estratégias de avaliação com notável rapidez. Essa tendência, no entanto, levanta questões importantes quando observamos o modo como estudantes neófitos, ainda em formação, se habituariam a realizar tarefas fundamentais de suas futuras profissões com o auxílio constante dessas tecnologias. Esse entusiasmo pode, por um lado, fortalecer habilidades de análise, mas, por outro, pode levar a uma geração de profissionais cuja autonomia intelectual e técnica vai depender estruturalmente de ferramentas que, afinal, são produtos licenciados e, como tal, são mantidos por empresas com interesses comerciais e sujeitos a oscilações de custo, funcionamento e acesso. Como a área da Educação pode equilibrar o aproveitamento das ferramentas de IA com a formação de profissionais verdadeiramente críticos e autônomos, conscientes das bases construtivas de seu próprio conhecimento?

A IA está sendo amplamente discutida (e até reverenciada) como ferramenta, mas a quem deve recair a responsabilidade pela reflexão epistemológica a seu respeito — ou seja, a discussão sobre como se produz conhecimento *com* ela, *por meio dela* ou *a despeito* dela? Quais os riscos de formar profissionais que, em cenários de indisponibilidade acidental das IAs, estarão despreparados para operar sem elas? Ou, ainda, quais os riscos de, em cenários de restrição deliberada no uso das IAs, termos uma elitização do acesso a essa tecnologia?

No fim das contas, a promessa de uma “democratização do conhecimento” via IA é real ou ilusória?

The incorporation of AI, especially its generative form, into educational settings has increasingly been framed as an unstoppable (and even desirable) step toward pedagogical modernization and the *democratization* of knowledge. That word, *democratization*, is particularly important. In schools, such tools have been integrated into the curriculum, classroom practices, and assessment strategies with remarkable speed. This trend, however, raises important questions when we consider how inexperienced students, still in the early stages of their education, are becoming accustomed to performing key tasks of their future professions with constant reliance on these technologies. On one hand, this enthusiasm may foster analytical skills; on the other, it may result in a generation of professionals whose intellectual and technical autonomy is structurally dependent on tools that are, after all, proprietary products, and, as such, are maintained by companies with commercial interests and subject to fluctuations in cost, functionality, and accessibility. How can the field of Education find a balance between leveraging the potential of AI tools and cultivating truly critical and autonomous professionals, aware of the constructive foundations of their own knowledge?

AI is being widely discussed (and even praised) as a tool, but who bears the responsibility for epistemological reflections on it—that is, for leading discussions about how knowledge is produced *with* it, *through* it, or *in spite of* it? What are the risks of training professionals who, in scenarios of accidental unavailability of AI, may be unprepared to function without it? Or, alternatively, what are the risks of deliberate restrictions on AI usage leading to the access to such technology being concentrated only among the privileged?

In the end, is the promise of a “democratization of knowledge” through AI real or illusory?



desenvolvido no MIT (sigla em inglês para o Instituto de Tecnologia de Massachusetts), pesquisadores utilizaram eletroencefalograma para medir as atividades cerebrais de pessoas enquanto elas realizavam tarefas de escrita. Eles separaram os participantes da pesquisa em três principais grupos: o primeiro utilizou IA generativa (ChatGPT), o segundo utilizou um mecanismo de busca (Google) e o terceiro não poderia utilizar nenhum desses recursos para auxílio na redação. O grupo que utilizou o ChatGPT foi o que obteve menos conexões neuronais, ou seja, um grau menor de conectividade neural. Eles também não conseguiam se lembrar do que escreveram e se sentiam menos autores do conteúdo produzido. Outro ponto é que os textos produzidos com IA generativa foram considerados menos criativos e também parecidos entre si. Em uma nova rodada de experimentos, o grupo que estava escrevendo sem o auxílio da IA passou, então, a utilizar a ferramenta, enquanto aqueles que a utilizaram anteriormente passaram a escrever o texto sem nenhum auxílio. Com isso, foi possível perceber que o grupo que começou a escrita sem o uso da IA e depois teve acesso à ferramenta teve o maior ganho em termos de conexões neurais. Contudo, esse mesmo ganho não foi obtido pelo grupo que teve acesso à ferramenta de IA generativa primeiro. Esse estudo é muito interessante, pois abre caminhos para entendermos qual a melhor forma de utilização das ferramentas de IA generativa no processo de aprendizagem, para que seja possível obter autonomia intelectual, apesar das facilidades oferecidas pela tecnologia.

As organizações que criam e implementam a IA devem integrar princípios éticos desde o design e ao longo de todo o ciclo de vida do sistema. É importante ressaltar que a adoção de uma postura crítica diante da tecnologia é essencial. É preciso que o pensamento crítico seja incentivado para que os usuários questionem e analisem as respostas da IA. Uma dependência excessiva da IA sem uma compreensão clara de suas limitações pode levar a sérios problemas, incluindo profissionais despreparados.

Há um risco de confiar excessivamente nas decisões da IA sem questioná-las. Os profissionais

conducted at the Massachusetts Institute of Technology (MIT), researchers used electroencephalography (EEG) to measure participants' brain activity while performing writing tasks. The study divided participants into three main groups: the first used generative AI (ChatGPT), the second used a search engine (Google), and the third could not use any of these tools for assistance in writing. The group using ChatGPT accounted for the fewest neural connections, or in other words a lower degree of neural connectivity. They also struggled to recall what they had written and felt less ownership over the content they produced. Additionally, texts generated with AI were less creative and tended to resemble one another. In a subsequent round of experiments, the group that had initially written without AI then gained access to the tool, while those who had used it first had to write without any kind of assistance. It was observed that the group that started writing without AI and later used it showed the greatest gain in neural connectivity. However, this gain was not observed in the group that used generative AI first. This study is particularly interesting because it sheds light on the best ways to use generative AI tools in the learning process in order to foster intellectual autonomy, despite the conveniences offered by technology.

Organizations that create and implement AI should integrate ethical principles from the design stage throughout the system's entire lifecycle. It is important to emphasize that adopting a critical stance toward technology is essential. Critical thinking must be encouraged so that users question and analyze AI outputs. Excessive reliance on AI without a clear understanding of its limitations can lead to serious problems, including poorly prepared professionals.

There is a risk of overtrusting AI decisions without scrutiny. Professionals who cannot independently validate information may propagate incorrect data, with potentially disastrous consequences. While regulation is necessary to mitigate abuses, excessive rules can slow technological development in some regions,

Siga o link pelo *QR code* para acessar os resultados do estudo **“Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task”** (numa tradução livre: “Seu cérebro sob o efeito do ChatGPT: Acumulação de dívida cognitiva ao longo do uso de um assistente de IA em tarefas de redação de textos dissertativos”), de autoria de Nataliya Kosmyna, do MIT, e outros pesquisadores

You may use the QR code to follow the link and access the results of the study **“Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task,”** authored by Nataliya Kosmyna, from MIT, and other researchers



Siga o link para explorar os resultados do estudo:


Follow the link to explore the results of the study:




que não conseguem validar informações de forma independente podem disseminar dados incorretos e gerar consequências desastrosas. Embora uma regulamentação seja necessária para mitigar abusos, o excesso de normas pode atrasar o desenvolvimento tecnológico em algumas regiões, especialmente naquelas que consomem IA em vez de produzi-la. Outro ponto é que o monopólio de empresas poderosas sobre grandes volumes de dados pode levar a um “colonialismo de dados”, concentrando o controle sobre informações e tecnologias. A IA generativa de fato apresenta um potencial significativo para a democratização do acesso à tecnologia e, por extensão, ao conhecimento. Além disso, ela permite facilitar a interação através da linguagem natural, quebrando barreiras de acessibilidade e permitindo resolver problemas complexos com menos conhecimento prévio. Contudo, para que a IA seja utilizada de forma eficaz e consciente, é essencial que haja educação e conscientização sobre suas capacidades reais, bem como sobre seus limites e riscos, e também sobre responsabilidades. Essa democratização é condicionada à implementação de diretrizes éticas robustas, a uma regulação transparente, à supervisão humana e a uma educação contínua que promova o pensamento crítico.

particularly those that consume AI rather than produce it. Another concern is that the monopoly of powerful companies over large datasets can lead to “data colonialism,” concentrating control over information and technology.

Generative AI does have significant potential to democratize access to technology and, by extension, knowledge. Moreover, it facilitates interaction through natural language, breaking down accessibility barriers and enabling complex problem-solving with less prior knowledge. However, for AI to be used effectively and responsibly, education and awareness about its real capabilities, limitations, risks, and associated responsibilities are essential. This democratization depends on the implementation of robust ethical guidelines, transparent regulation, human oversight, and continuous education that fosters critical thinking.

 **Rodrigo Barchi:** Extreme care is required with the word “democracy,” as it is increasingly trivialized and, consequently, less and less exercised and understood. Democracy is a multiple and continuous process that only occurs through autonomous individuals who, when interconnected




 **Rodrigo Barchi:** É necessário um extremo cuidado com a palavra “democracia”, pois ela é cada vez mais banalizada e, conseqüentemente, cada vez menos exercida e compreendida. Democracia é um processo múltiplo e ininterrupto, que só ocorre a partir de individualidades autônomas que, entrelaçadas nas coletividades, conseguem não somente tomar decisões adequadas pelo bem comum, mas também criam conhecimento no encontro, no diálogo e na diferença. Quando se tem a criação e a disponibilização de conhecimento exclusivamente através de corporações tecnológicas que decidem o que deve ser tendência, mercado e desejo, não há democracia. Hoje, em relação à economia, por exemplo, os indivíduos que moram nas cidades são reféns de uma produção agropastoril que está concentrada nas mãos de cada vez menos pessoas. Qualquer estremecimento no mercado financeiro, no clima e na geopolítica faz com que aqueles que têm menos recursos corram o risco de morrer de fome, perder suas casas, não ter acesso a crédito, não ter acesso a serviços de educação e saúde, e por aí vai. Que democracia é essa?!


Com a questão da IA, que monopoliza o saber nas plataformas digitais, a questão é ainda mais grave, pois, além de um serviço tecnológico mais aperfeiçoado ser mais caro, a matéria prima da IA não é paga a quem produz, que é o *general intellect*, ou seja, o coletivo. Existe, na IA, uma retroalimentação de bens, que, ao diferenciar quem pode pagar mais ou menos, já destrói qualquer ideia de democracia no ninho. Só existe democracia quando não somente há a disponibilização igualitária da tecnologia, do produto, da informação e do conhecimento, mas quando existe a criação desse saber de modo consciente, por todos, de modo que todos usufruam daquilo que ajudam a inventar. O que existe, no modelo de IA generativa atual, é puramente o contrário, pois, ao dar uma informação mastigada para, por exemplo, estudantes de graduação, essas IAs fazem com que esses estudantes percam completamente o vínculo com a produção de saber e o trabalho que isso despende. Eles são justamente separados e alienados dos processos de investigação na ciência — que se dá através do método científico: da escolha de um objeto, da coleta de dados, da sistematização e da compreensão desses dados, da

within collectives, are able not only to make proper decisions for the common good but also to create knowledge through encounter, dialogue, and difference. When knowledge is created and made available exclusively through technological corporations that decide what should be trending, marketed, and desired, there is no democracy. Nowadays, when it comes to economy, for example, individuals living in cities are dependent on farming and livestock production concentrated in the hands of an ever-smaller number of people. Any tremor in financial markets, the climate, or geopolitics puts those with fewer resources at risk of hunger, losing their homes, being denied access to credit, education, healthcare, and so on. What kind of democracy is this?!

When it comes to AI, which monopolizes knowledge on digital platforms, the problem is even more severe. Not only is a more advanced technological service more expensive, but the raw material of AI is not compensated to those who produce it, the general intellect, that is, the collective itself. In AI, there is a self-perpetuating cycle of goods that, by differentiating who can pay more or less, already destroys any notion of democracy at its core. True democracy exists only when technology, products, information, and knowledge are not only made equally available but when the creation of that knowledge is conscious, collective, and accessible to all who contribute to its invention. In the current generative AI model, the opposite occurs. By providing pre-digested information to undergraduate students, for example, these AIs cause them to lose all connection with the production of knowledge and the work it entails. Students are effectively separated and alienated from the processes of scientific investigation—which takes place through the scientific method: choosing an object of study, collecting data, systematizing and understanding that data, constructing syntheses, producing academic papers and essays, and disseminating scientific knowledge.

 **Alfonso Gutiérrez:** It is clear that the incorporation of generative AI into educational institutions is already an almost inevitable reality,

construção de sínteses, da produção de artigos e ensaios acadêmicos, e da divulgação científica.

 **Alfonso Gutiérrez:** Está claro que a incorporação da IA generativa às instituições de ensino já é uma realidade praticamente inevitável, mas a relação da IA com a educação vai além de seu uso para favorecer os processos de ensino-aprendizagem. A Unesco (sigla em inglês para a Organização das Nações Unidas para a Educação, a Ciência e a Cultura), num fórum internacional realizado em Pequim, na China, em dezembro de 2020 — o Fórum Internacional sobre IA e os Futuros da Educação: Desenvolvendo Competências para a Era da IA (no original: *International Forum on AI and the Futures of Education: Developing Competencies for the AI Era*) —, advertia que é importante revisar cuidadosamente os esforços para integrar a IA nos currículos escolares, seja ensinando *sobre* técnicas de IA, seja ensinando *com* ferramentas de IA, ou *preparando os jovens para a vida* em um mundo cada vez mais influenciado pela IA.

A IA generativa, como recurso didático, pode ser de grande utilidade no uso de aplicações especificamente concebidas como assistentes e tutores socráticos personalizados, sempre tendo em conta que ensinar não é tanto transmitir informação, mas favorecer e personalizar as aprendizagens, e que a informação (como já ocorria com meios anteriores) traz consigo vieses, dependendo de suas fontes e da forma como é gerada. Para um verdadeiro aproveitamento das ferramentas de IA no ensino, faz-se necessária a formação de um corpo docente crítico e autônomo, e será também necessário um letramento em IA, tanto por parte de professores quanto por parte de estudantes: um letramento que não se reduza a ensinar sobre técnicas de IA, mas que sensibilize em relação às consequências sociais e econômicas de um desenvolvimento descontrolado da IA. A responsabilidade de articular uma reflexão epistemológica sobre a IA é compartilhada por vários atores: programadores, criadores de conteúdo e a comunidade acadêmica, incluindo docentes e discentes. Nas instituições de ensino, aprende-se e *reflete-se sobre o aprendizado*, e isso deve ser feito com variedade metodológica suficiente para gerar conhecimento *sobre* a IA, *com* a IA e *apesar* da IA.

but links between AI and education go beyond its use to support teaching and learning processes. UNESCO (the United Nations Educational, Scientific and Cultural Organization), in an international forum held in Beijing, China, in December 2020—the International Forum on AI and the Futures of Education: Developing Competencies for the AI Era—warned that it is important to carefully review efforts to integrate AI into school curricula, whether by teaching *about* AI techniques, teaching *with* AI tools, or *preparing young people for life* in a world increasingly influenced by AI.

Generative AI, as a teaching resource, can be extremely useful when applied specifically as personalized Socratic assistants and tutors, always keeping in mind that teaching is less about transmitting information and more about fostering and personalizing learning. Information, as has always been the case with previous media, carries biases depending on its sources and the way it is generated. For a meaningful use of AI tools in education, it is necessary to train a critical and autonomous teaching staff, and AI literacy is also required, both for teachers and students: a form of literacy that should go beyond teaching AI techniques, also raising awareness of the social and economic consequences of uncontrolled AI development. The responsibility for fostering epistemological reflection on AI is shared among several agents: programmers, content creators, and the academic community, including both teachers and students. Within educational institutions, students learn and also *reflect on learning*, and this must be done with sufficient methodological variety to generate knowledge *about* AI, *with* AI, and *despite* of AI.

As in the answer to the previous question, we also need to consider the potential risks of AI, in this case for education. Perhaps the most important is “intellectual laziness” or the systematic delegation of certain cognitive tasks to machines, such as information retrieval and analysis, writing, calculation, diagnostics, and so on, which reduces the ability to perform these functions autonomously. Generative AI is said



Como na pergunta anterior, também faremos referência a possíveis riscos da IA, neste caso para a educação. Talvez o mais importante seja a “preguiça intelectual” ou a delegação sistemática de determinadas tarefas cognitivas às máquinas, como a busca e a análise de informação, a redação, o cálculo, diagnósticos etc., o que reduz a capacidade de executar tais funções de forma autônoma. Diz-se da IA generativa que ela fala e escreve como os humanos; devemos evitar que algum dia possa ser dito que os humanos se expressam como a IA generativa. Essas ferramentas entraram na academia pela porta dos fundos e têm sido utilizadas como substituto, e não como apoio ao aprendizado crítico e reflexivo. Sua grande utilidade para realizar tarefas escolares e acadêmicas despertou grande entusiasmo e vem sendo utilizada sem considerar os riscos de atrofia de competências ou de preguiça intelectual.

**Assim como as IAs generativas se expressam como humanos, corremos o risco de passar a nos expressar como elas — ou de depender demais delas para conseguir fazê-lo**

Cada nova tecnologia representa algo a mais para aqueles que já têm de tudo e algo a menos para aqueles que quase nada possuem. O aumento da brecha digital na educação é outro dos riscos da incorporação da IA. Se as competências para interagir e se beneficiar dela ficarem concentradas nos grupos com maior poder aquisitivo ou capital institucional, acentuar-se-á uma brecha não apenas econômica e digital, mas também cognitiva e produtiva entre quem pode acessar a IA e quem não pode. Isso não apenas reproduz desigualdades preexistentes, mas redefine a distribuição social do conhecimento e do poder de decisão.

to speak and write like humans; we must avoid a future in which it could be said that humans express themselves like generative AI. These tools have entered academia through the back door and have been used as substitutes, rather than as supports for critical and reflective learning. Their great utility for completing academic tasks has generated considerable enthusiasm and has been employed without considering the risks of competency atrophy or intellectual laziness.

**Just as generative AIs express themselves like humans, we risk beginning to express ourselves like them—or relying on them too much to be able to do so**

Every new technology represents something extra for those who already have everything, and something less for those who have almost nothing. The widening of the digital gap in education is another risk associated with the adoption of AI. If the skills to interact with and benefit from AI remain concentrated among groups with greater purchasing power or institutional capital, a gap will emerge not only economically and digitally, but also cognitively and productively, between those who can access AI and those who cannot. This not only reproduces preexisting inequalities but also redefines the social distribution of knowledge and decision-making power.



Em variadas intensidades, as tecnologias digitais (incluindo a Inteligência Artificial) já encontraram caminho para dentro das salas de aula, mas os educadores precisam ir além de ensinar seus estudantes a utilizá-las: é fundamental ensiná-los a refletir sobre essas ferramentas, questionando criticamente seu impacto nas práticas sociais

At varying degrees, digital technologies (including Artificial Intelligence) have already found their way into classrooms, but educators must go beyond teaching students how to use these tools; it is essential to teach them to reflect on these technologies, critically questioning their impact on social practices



# 4


**Blackboxing e o paradoxo de uma ciência que não compreende a si mesma**

**Blackboxing and the paradox of a science that does not understand itself**




Foto/Photo: jackie\_vfx (Adobe Stock)

O conceito de *blackboxing*, ou de criar uma “caixa-preta”, descreve aquelas situações em que um sistema tecnológico funciona, mas baseado em processos internos opacos ou incompreensíveis, em alguns casos até mesmo para os especialistas. Um exemplo emblemático é o AlphaFold, um sistema de IA da Google DeepMind capaz de prever com incrível precisão a estrutura tridimensional de proteínas — o que é muito importante para vários tipos de pesquisa na área da biomedicina e da saúde em geral, já que a forma tridimensional de uma proteína está diretamente relacionada à sua função biológica. Apesar da precisão, no entanto, não é possível entender completamente como essas previsões estão acontecendo; o sistema funciona (pois os resultados podem ser confirmados experimentalmente), mas nós não sabemos *como* e, portanto, não conseguimos replicar os mesmos exercícios preditivos sem o auxílio da IA (o que, de certa forma, ecoa a questão da pergunta anterior, sobre IA e a formação de novos profissionais). Tomando por base que reprodutibilidade e verificabilidade são valores essenciais para a ciência convencional, como é que soluções como o AlphaFold mudam intrinsecamente a nossa forma de conceber e fazer ciência — mesmo que, instrumentalmente, ela funcione?

 **Bruno Aguilar:** No contexto da IA, a opacidade representa um desafio significativo, especialmente em modelos de *Deep Learning* e Redes Neurais Artificiais (como é o caso da chamada arquitetura *Transformer*). Nesses sistemas, a complexidade das múltiplas camadas e a forma como os neurônios artificiais e o modelo matemático geram representações abstratas tornam extremamente difícil interpretar com clareza como o sistema chegou a uma determinada conclusão. Nas ferramentas comerciais (ChatGPT, Gemini, entre outras), esse problema é agravado por outros fatores, como segredos industriais e a proteção da propriedade intelectual, que limitam a transparência dos algoritmos e a arquiteturas dos modelos computacionais utilizados.

Na ciência convencional, a reprodutibilidade (a capacidade de obter os mesmos resultados seguindo os mesmos métodos) e a verificabilidade (a possibilidade de comprovar a verdade ou precisão de uma afirmação ou resultado) são

The concept of blackboxing refers to situations in which a technological system functions effectively, yet operates through internal processes that are opaque or incomprehensible, sometimes even to experts. A striking example is AlphaFold, an AI system developed by Google DeepMind that can predict the three-dimensional structure of proteins with remarkable accuracy—which is highly significant for various types of biomedical and health-related research, since a protein’s 3D structure is directly tied to its biological function. Despite its precision, however, we are unable to fully understand how these predictions are being made; the system does work (since its results can be experimentally verified), but we cannot explain *how* it works, so we cannot replicate the predictive process without the help of AI (which, in many ways, echoes the last question regarding AI and the education of new professionals). Given that reproducibility and verifiability are core values of conventional science, how do solutions like AlphaFold fundamentally reshape our way of understanding and practicing science—even if, instrumentally, it works?


 **Bruno Aguilar:** In the context of AI, opacity represents a significant challenge, especially in Deep Learning and Artificial Neural Network models (such as the so-called Transformer architecture). In these systems, the complexity of multiple layers and the way artificial neurons and the mathematical model generate abstract representations make it extremely difficult for one to clearly interpret how the system arrived at a given conclusion. In commercial tools (such as ChatGPT, Gemini, among others), this issue is further aggravated by factors such as trade secrets and the protection of intellectual property, which both limit the transparency of the algorithms and the architectures of the computational models employed.

In conventional science, reproducibility (the ability to obtain the same results following the same methods) and verifiability (the possibility of confirming the truth or accuracy of a statement or result) are essential pillars. For scientific work

pilares essenciais. Para que o trabalho científico seja legitimado, ele precisa estar inserido no contexto de uma comunidade científica e seguir critérios e normas estabelecidos. Se um sistema de IA como o AlphaFold produz resultados precisos, mas não consegue explicar os passos ou a lógica que levaram a essa previsão, a ciência se depara com um resultado que não pode ser totalmente justificado ou replicado *independentemente* da máquina. A opacidade dificulta a identificação e a mitigação de vieses, o que é crucial em áreas sensíveis como saúde e biomedicina, onde decisões baseadas em IA podem ter impactos significativos na vida das pessoas.

Se não sabemos *como* a IA toma suas decisões, não podemos garantir que elas sejam sempre justas ou imparciais. É fundamental, até mesmo para um melhor aproveitamento e uma melhor utilização da IA generativa na pesquisa, que os sistemas sejam projetados para serem transparentes e explicáveis, documentando a metodologia, os dados e os critérios de decisão. A iniciativa *Explainable Artificial Intelligence* (XAI) busca tornar os processos de tomada de decisão da IA compreensíveis para os humanos, mas é importante ressaltar que essa iniciativa é extremamente desafiadora, uma vez que tenta obter modelos explicáveis com desempenho similar aos modelos atuais de IA generativa. Tais modelos possuem grande complexidade e exigem tempo e recursos financeiros substanciais das grandes *Big Techs*, que se encontram numa intensa corrida de aprimoramento dos seus modelos comerciais proprietários.


O grande desafio, no fim das contas, é encontrar um equilíbrio, de modo a promover a inovação sem perder a essência da investigação científica, baseada na compreensão e na reflexão crítica.

 **Luís Albano:** O AlphaFold é um bom exemplo do dilema contemporâneo entre a promessa e o preço do avanço tecnológico. Ele resolve, com impressionante eficiência, um problema que mobilizou décadas de pesquisa: prever o dobramento de proteínas. No entanto, a forma como chega a essas soluções não é transparente. Não é apenas uma questão de não entendermos todos os passos, mas de não termos, muitas vezes, acesso pleno às bases de treino, às escolhas de modelagem e aos parâmetros que moldam o

to be legitimized, it must be situated within the context of a scientific community and follow established criteria and standards. If an AI system like AlphaFold produces accurate results but cannot explain the steps or logic that led to that prediction, science faces an outcome that cannot be fully justified or replicated independently without the machine. Opacity also hinders the identification and mitigation of biases, which is crucial in sensitive areas such as health and biomedicine, where AI-based decisions can have significant impacts on people’s lives.

If we do not understand how AI makes its decisions, we cannot guarantee that they will always be fair or unbiased. It is essential, even for better utilization of generative AI in research, that systems are designed to be transparent and explainable, documenting methodology, data, and decision criteria. The Explainable Artificial Intelligence (XAI) initiative seeks to make AI decision-making processes understandable to humans, but it is important to emphasize that this initiative is extremely challenging, as it seeks to obtain explainable models with performance comparable to that of current generative AI models. Such models are highly complex and require substantial time and financial resources from major Big Tech companies, which are engaged in an intense race to enhance their proprietary commercial models.

After all, the great challenge lies in finding a balance, promoting innovation without losing the essence of scientific investigation, which is based on understanding and critical reflection.

 **Luís Albano:** AlphaFold is a prime example of the contemporary dilemma between the promise and the cost of technological advancement. It solves, with impressive efficiency, a problem that mobilized decades of research: predicting protein folding. However, the way it arrives at these solutions is not transparent. It is not just a matter of not understanding every step, but often of not having full access to the training datasets, modeling choices, and parameters that shape the

sistema. Isso nos coloca diante de uma mudança epistemológica: passamos de uma ciência que privilegia o entendimento passo a passo para uma ciência que, em alguns casos, aceita o resultado como suficiente, mesmo sem poder reproduzir integralmente o processo.

## A opacidade das IAs deve ser considerada de modo a não comprometer a reprodutibilidade e a verificabilidade da pesquisa científica

Essa mudança não é trivial. Na ciência convencional, a verificabilidade e a reprodutibilidade não são apenas formalidades metodológicas, mas o que permite confiança, debate e avanço coletivo. Modelos como o AlphaFold funcionam mais como caixas-pretas: entregam resultados úteis, mas cuja lógica interna é, em grande parte, inacessível. Isso desloca a fronteira entre “conhecimento” e “instrumento”: começamos a tratar do funcionamento da operação tecnológica como secundário em relação à utilidade prática, o que é eficiente no curto prazo, mas arriscado para a própria estrutura da ciência no longo prazo.

O impacto vai além do laboratório. Quando a confiança nos resultados depende de atores concentrados, empresas ou consórcios com acesso exclusivo à infraestrutura, aos dados e ao código, criamos um modelo de ciência que não é plenamente auditável nem distribuída. E aqui voltamos ao que já foi discutido: trata-se menos de uma “democratização” e mais de uma redistribuição seletiva, em que o acesso ao método se torna um privilégio. Isso não significa que não devamos usar o AlphaFold ou similares, mas que precisamos desenvolver, junto com a aplicação, uma política de abertura, documentação e escrutínio. Caso contrário, estaremos abrindo mão de um pilar central da ciência para ganhar velocidade, e como sabemos de outras áreas da História, ir mais rápido sem saber exatamente para onde nem sempre acaba bem.

system. This presents an epistemological shift: we are moving from a science that values step-by-step understanding to a science that, in some cases, accepts the result as sufficient, even without being able to fully reproduce the process.

## The opacity of AI should be taken into account to avoid compromising the reproducibility and verifiability of scientific research

This shift is not trivial. In conventional science, verifiability and reproducibility are not merely methodological formalities; they are what enable trust, debate, and collective progress. Models like AlphaFold operate more like black boxes: they deliver useful results, but their internal logic is largely inaccessible. This shifts the boundary between “knowledge” and “instrument”: we begin to treat the functioning of the technology as secondary to its practical utility, which is efficient in the short term but risky for the very structure of science in the long term.

The impact goes beyond the realm of science labs. When trust in results depends on concentrated agents, companies or groups with exclusive access to infrastructure, data, and code, we create a model of science that is neither fully auditable nor distributed. And here we return to a point already discussed: this is less about “democratization” and more about selective redistribution, where access to the method becomes a privilege. This does not mean we should not use AlphaFold or similar tools, but that we need to develop, alongside their application, a policy of openness, documentation, and scrutiny. Otherwise, we are sacrificing a central pillar of science for speed, and as history has shown in other fields, moving faster without knowing exactly where we are headed does not always end well.



# 5

A IA deve ser regulada – e, se sim, por quem?

Should AI be regulated—and if so, by whom?



Foto/Photo: jackie\_vfx (Adobe Stock)

O filme *Jurassic Park* (1993), inspirado no livro homônimo de Michael Crichton, utilizou o entretenimento para abordar, de forma alegórica, questões éticas e científicas prementes, sobretudo no contexto de intensa expansão da engenharia genética nas décadas de 1980 e 1990. À época, a ciência avançava rumo ao sequenciamento do genoma humano, à clonagem de mamíferos e ao surgimento dos primeiros organismos geneticamente modificados, ainda sem a devida regulação. *Jurassic Park* traduzia nas telas de cinema esse entusiasmo tecnocientífico ao levantar uma provocação central: que direito têm os cientistas — e especialmente as corporações capitalistas que os financiam — de levar adiante certos experimentos, quando não estão plenamente conscientes das consequências éticas, ambientais e sociais de suas ações?


Uma cena emblemática mostra o matemático Ian Malcolm confrontando John Hammond, dono da empresa de biotecnologia responsável por ressuscitar os dinossauros e comercializá-los como atrações de um parque de diversões. Ele diz: “Você não vê o perigo inerente ao que está fazendo aqui, John? O poder genético é a força mais impressionante que este planeta já viu. Mas você o maneja como uma criança que encontrou a arma do pai. [...] Eu vou te dizer: o problema com o poder científico que vocês usaram é que ele não exigiu nenhuma disciplina para ser alcançado. Vocês leram o que os outros fizeram e deram o próximo passo. Não conquistaram esse conhecimento por si mesmos, então não sentem responsabilidade por ele. Vocês se apoiaram nos ombros de gênios para realizar algo o mais rapidamente possível e, antes mesmo de entenderem o que tinham em mãos, vocês o patentearam, embalaram, estamparam numa lancheira de plástico e agora querem vendê-lo. [...] Seus cientistas estavam tão preocupados em saber se *conseguiriam* fazer o que fizeram, que não pararam para pensar se *deveriam*.”

The film *Jurassic Park* (1993), inspired by Michael Crichton’s novel of the same title, used entertainment as an allegory to address pressing ethical and scientific issues, especially in the context of the rapid expansion of genetic engineering during the 1980s and 1990s. At that time, science was advancing toward sequencing the human genome, cloning mammals, and the emergence of the first genetically modified organisms, all occurring without proper regulation. *Jurassic Park* translated this techno-scientific enthusiasm on the big screen while raising a central provocation: what right do scientists—and especially the capitalist corporations that fund them—have to carry out certain experiments when they are not fully aware of the ethical, environmental, and social consequences of their actions?

A key scene shows the mathematician Ian Malcolm confronting John Hammond, the owner of the biotechnology company responsible for resurrecting dinosaurs and commercializing them as amusement park attractions. He says: “Don’t you see the danger, John, inherent in what you’re doing here? Genetic power is the most awesome force ever seen on this planet. But you wield it like a kid who’s found his dad’s gun. [...] I’ll tell you. The problem with the scientific power you’ve used is it didn’t require any discipline to attain it. You read what others had done and you took the next step. You didn’t earn the knowledge yourselves, so you don’t take the responsibility for it. You stood on the shoulders of geniuses to accomplish something as fast as you could, and before you knew what you had, you patented it, packaged it, slapped it on a plastic lunch box, and now you want to sell it. [...] Your scientists were so preoccupied with whether or not they *could* that they didn’t stop to think if they *should*.”


The analogy with current dilemmas related to AI seems inevitable. Today, once again, we

A analogia com os dilemas atuais relacionados à IA parece inevitável. Hoje, mais uma vez, testemunhamos o avanço vertiginoso de uma tecnologia que impõe dilemas profundos antes que a sociedade possa compreendê-la plenamente ou estabelecer marcos regulatórios adequados. A IA é desenvolvida e implementada em um ambiente marcado por interesses comerciais, disputas geopolíticas e preocupações legítimas com segurança, privacidade, desinformação, desigualdade etc. Sendo assim, você acredita ser possível traçar um paralelo entre os desafios enfrentados pela engenharia genética no final do século XX e aqueles que a IA enfrenta agora, no primeiro quarto do século XXI? Desde então, o que aprendemos em termos de regulação científica? Quando a ciência avança mais rapidamente que os marcos éticos e jurídicos (o que parece irrefreável), quais atores devem liderar os esforços regulatórios: estados, corporações, a comunidade científica, organismos internacionais? E que papel deve ter o usuário, cidadão comum, nesse debate sobre os limites e o uso legítimo da IA?

 **Rodrigo Barchi:** Há dois cenários. O primeiro seria um hipotético-utópico, no qual as democracias realmente existissem — voltamos, então, à questão 3. Nesse primeiro cenário, as sociedades como um todo seriam responsáveis pela produção do saber, pela tecnologia e pelos impactos sociais, econômicos, ambientais e culturais desses processos (bem como estariam devidamente cientes dessa responsabilidade). Poderíamos, nesse panorama, ter tranquilidade em relação à IA. Minha preocupação não é necessariamente em relação à invenção das IAs, nem à sua existência ou pertinência quanto à resolução de problemas científicos; mas sim em relação à sua liberação como produto para as grandes massas, sem o devido processo de formação ética e política para que todas as pessoas possam utilizá-la.

Usar uma tecnologia sem a plena noção de todos os efeitos que ela pode causar é extremamente


witness the rapid advancement of a technology that imposes profound dilemmas before society can fully understand it or establish appropriate regulatory frameworks. AI is developed and implemented in an environment marked by commercial interests, geopolitical disputes, and legitimate concerns about security, privacy, misinformation, inequality, and more. Given all this, do you believe it is possible to draw a parallel between the challenges faced by genetic engineering at the end of the 20<sup>th</sup> century and those that AI faces now, in the first quarter of the 21<sup>st</sup> century? Since then, what have we learned in terms of scientific regulation? When science advances faster than ethical and legal frameworks (which seems inevitable), who should lead regulatory efforts: nations, corporations, the scientific community, or international organizations? And what role should users, as ordinary citizens, play in this debate over the limits and legitimate uses of AI?

 **Rodrigo Barchi:** There are two possible scenarios. The first one would be a hypothetical-utopian one, in which democracies truly existed—and so we return to question #3. In this first scenario, societies as a whole would be responsible for the production of knowledge and technology, as well as for the social, economic, environmental, and cultural impacts of these processes (besides being fully aware of that responsibility). In such a scenario, we could feel at ease with AI. Therefore, my concern is not necessarily about the invention of AIs, nor about their existence or their relevance when it comes to solving scientific problems; rather, it lies in their release as products marketed to the masses, without an adequate process of ethical and political education that would enable people to use them consciously.

Using a technology without full awareness of the effects it can cause is profoundly undemocratic, because it means making something available that, for some, will lead to greater wealth concentration,


antidemocrático, pois você tem a disponibilização de algo que, para alguns, vai promover uma maior concentração de renda e, para outros, processos de dependência, pilhagem, violência e exploração. A regulação, portanto — e aí entramos no outro cenário, o real atual — deveria ocorrer antes da plena disponibilização, depois de passar por um debate coletivo, que decidiria o que e quanto poderia ser disponibilizado. Seriam necessárias discussões em larga escala (nas esferas globais, nacionais, locais) para decidir o passo a passo da liberação dessas tecnologias, pois a própria forma como elas são produzidas e se replicam não é politicamente neutra. Elas servem para atender a quem as produz. Todos os atores deveriam participar dos processos de criação dos limites e das permissões sobre qualquer tecnologia. Isso seria democrático. O que temos hoje não é; é mera imposição de poucas corporações e isso é por demais perigoso.

**A liderança da comunidade científica é essencial para definir princípios éticos; ao mesmo tempo, cidadãos bem informados e uma regulação adequada também são necessários**

 **Renata Lima:** Em ambos os casos — tanto com a engenharia genética no final do século XX quanto com a IA neste início de século XXI —, nós nos deparamos com um dilema central. No século XX, perguntávamo-nos se poderíamos “brincar de Deus” e, neste século, os riscos rondam a desinformação em massa, a perda de empregos, a discriminação algorítmica... O paralelo se evidencia principalmente pela impossibilidade de controle total em ambos os casos. As tecnologias

and for others, to processes of dependency, plunder, violence, and exploitation. Regulation, therefore—and here we enter the second, contemporary and real scenario—should take place before full release, and only after collective debate has determined what and how much should be made available. Broad discussions would be required (at global, national, and local levels) to decide the step-by-step process of introducing such technologies, since the very way they are produced and replicated is not politically neutral. They serve the interests of those who create them. All stakeholders should participate in defining the boundaries and permissions surrounding any technology. That would be democratic. What we have today is not; it is merely the imposition of a few corporations, and that is profoundly dangerous.

**Leadership from the scientific community is essential to define ethical principles; at the same time, well-informed citizens and proper regulation are also required**

 **Renata Lima:** In both cases—genetic engineering at the end of the 20<sup>th</sup> century and AI at the beginning of the 21<sup>st</sup>—we are faced with a central dilemma. In the 20<sup>th</sup> century, the question was whether we could “play God;” in this century, the risks revolve around mass disinformation, job losses, and algorithmic discrimination. The parallel becomes clear above all due to the impossibility of full control in either case. Technologies tend to spread uncontrollably and trigger unpredictable



possuem a tendência de se espalhar de maneira descontrolada, além de suscitar comportamentos imprevisíveis, e, em ambos os casos, a ciência foi mais rápida do que a legislação.

Tanto a engenharia genética quanto a IA apresentam característica dual: a engenharia genética possui o poder de curar doenças, mas também de criar novos patógenos; a IA atual pode democratizar o acesso ao conhecimento, mas também amplificar desigualdades, otimizar processos, mas também eliminar empregos, combater a desinformação, mas também criá-la em escala industrial. Outras semelhanças, além da criação de leis posteriores ao aparecimento dos problemas, podem ser observadas: a falta de coordenação internacional, a influência excessiva das empresas, assim como a falta de conhecimento adequado sobre a tecnologia pelas pessoas que participam de sua regulamentação. As decisões envolvem simultaneamente diversas esferas: técnica, social e política. Temos, mais uma vez, uma situação em que a velocidade da evolução tecnológica supera a capacidade dos reguladores de acompanhar as mudanças. Uma lição importante do século XX diz respeito ao fato de *não* esperar o dano acontecer para depois regulamentar — pois essa pode ser uma estratégia perigosa, uma vez que a tecnologia atual pode se espalhar e impactar a sociedade de forma irreversível.

Surgem, então, as seguintes questões: como construir uma governança tecnológica que seja simultaneamente democrática, tecnicamente competente, globalmente coordenada e adaptável à velocidade da inovação? A autorregulação seria uma resposta capaz de solucionar nossos problemas? Esse debate certamente não será harmonioso, pois existem grupos com objetivos distintos: as empresas tentam liderar o mercado, aumentar o lucro e evitar o que quer que possa restringir sua velocidade de inovação e/ou aumentar seus custos operacionais; os governos têm como meta proteger os cidadãos e garantir a segurança nacional; a comunidade científica e a sociedade civil querem garantir que a IA seja desenvolvida e usada de forma segura, ética, transparente, e que ela beneficie a humanidade como um todo, não apenas alguns poucos.

behavior, and in both instances, science has advanced faster than legislation.

Both genetic engineering and AI share a dual nature: genetic engineering holds the power to cure diseases but also to create new pathogens; today's AI can democratize access to knowledge but also amplify inequalities, optimize processes but also eliminate jobs, fight disinformation but also generate it on an industrial scale. Other similarities—beyond the fact that laws usually emerge only after problems appear—include the lack of international coordination, the excessive influence of private corporations, and the limited understanding of technology among those responsible for regulating it. Decision-making involves overlapping technical, social, and political dimensions. Once again, we find ourselves in a situation where the pace of technological evolution surpasses regulators' ability to keep up with change. An important lesson from the 20<sup>th</sup> century concerns the fact that we should not wait for harm to occur before regulating something—as this can be a dangerous strategy, since today's technology can spread and impact society in an irreversible way.

This raises pressing questions: how can we build technological governance that is simultaneously democratic, technically competent, globally coordinated, and adaptable to the speed of innovation? Could self-regulation be a viable solution? This debate will certainly not be harmonious, since the agents involved have conflicting goals: companies aim to lead markets, maximize profit, and avoid anything that might restrict innovation or increase operational costs; governments seek to protect citizens and ensure national security; and the scientific community and civil society strive to guarantee that AI is developed and used in a safe, ethical, and transparent manner that benefits humanity as a whole, not just a privileged few.

What have we learned in recent years? We've learned that purely reactive regulation is




Foto/Photo: pla2na (Adobe Stock)

Robôs estão à frente da lei: o desenvolvimento tecnológico caminha mais rápido do que nossa capacidade de regulá-lo, então, quando se fala em regulação da Inteligência Artificial por meio da legislação, é preciso ir além da contenção de danos — especialmente depois que eles já estão acontecendo —; em vez disso, é necessário orientar o desenvolvimento dessas tecnologias com metas estabelecidas coletivamente

Robots are ahead of the law: technological development advances faster than our ability to regulate it; therefore, when discussing the regulation of Artificial Intelligence through legislation, we must go beyond merely containing harm—especially once it is already occurring—; instead, it is necessary to guide the development of these technologies according to collectively defined goals


O que aprendemos nesses anos? Aprendemos que a regulação puramente reativa é inadequada para tecnologias de alto impacto. Aprendemos que a liderança da comunidade científica é necessária para definir os princípios éticos, mas que se faz necessário o poder do Estado para que esses princípios sejam de fato implementados. Inevitavelmente, precisamos de um trabalho multidisciplinar. A experiência histórica com a regulação da biotecnologia oferece lições valiosas para a governança da IA, mas também revela a inadequação de abordagens puramente reativas. O desafio contemporâneo reside em desenvolver marcos regulatórios que sejam simultaneamente proativos, tecnicamente informados e democraticamente legitimados. A regulação efetiva da IA demanda uma arquitetura de governança multinível, que combine: princípios éticos globalmente consensuais, marcos regulatórios nacionais adaptáveis, autorregulação responsável da indústria e participação de cidadãos informados. Essa abordagem híbrida deve equilibrar inovação e precaução, competitividade econômica e proteção social.

Quanto ao papel do cidadão comum, ele transcende a mera consulta pública, exigindo letramento tecnológico que permita participação qualificada no debate democrático sobre os rumos da IA. Somente através dessa participação informada será possível construir um consenso social sobre os limites éticos e os usos legítimos dessa tecnologia transformadora. A questão não é mais se a IA deve ser regulada, mas como construir uma governança que preserve tanto o potencial inovador quanto os valores democráticos fundamentais. A resposta a esse desafio definirá não apenas o futuro da tecnologia, mas o próprio futuro da sociedade democrática na era digital.

 **Alfonso Gutiérrez:** A questão sobre a regulação das IAs não admite nenhuma dúvida: dada a sua capacidade de alterar ecossistemas produtivos, relações sociais e dinâmicas de poder, como já apontamos, a ausência de marcos normativos robustos deixaria seu desenvolvimento e aplicação sujeitos quase exclusivamente a interesses corporativos e lógicas de mercado. Consideramos certo que a IA deve ser regulada,

inadequate for high-impact technologies. We’ve learned that scientific leadership is essential when it comes to defining ethical principles, but that state power is equally necessary to ensure those principles are actually implemented. Inevitably, this demands multidisciplinary work. The historical experience of biotechnology regulation offers valuable lessons for AI governance, while also exposing the limits of reactive approaches. The contemporary challenge lies in developing regulatory frameworks that are proactive, technically informed, and democratically legitimate. Effective AI regulation requires a multilevel governance architecture that combines globally agreed ethical principles, adaptable national frameworks, responsible industry self-regulation, and informed citizen participation. This hybrid approach must balance innovation and precaution, economic competitiveness and social protection.

As for the role of ordinary citizens, it goes far beyond mere public consultation: it requires technological literacy that enables meaningful participation in democratic debates about the future of AI. Only through informed participation can a social consensus be built around the ethical limits and legitimate uses of this transformative technology. The question is no longer whether AI should be regulated, but how to design governance that preserves both its innovative potential and the fundamental values of democracy. The answer to this challenge will define not only the future of technology, but the very future of democratic society in the digital age.

 **Alfonso Gutiérrez:** There’s no doubt about the issue of AI regulation: given its capacity to alter production ecosystems, social relations, and power dynamics, as we have already pointed out, the absence of robust normative frameworks would leave its development and application subject almost exclusively to corporate interests and market-driven logic. It is certain that AI must be regulated, and the real question lies in who should exercise that regulation and under what principles.

e o problema se coloca sobre quem deve exercer essa regulação e sob quais princípios.

A questão de fundo é que a IA não pode se autorregular nem ficar nas mãos de quem tem motivos ideológicos e econômicos para diminuir a clareza de seus impactos. Requer uma regulação que não apenas limite danos, mas que também oriente o desenvolvimento tecnológico para metas coletivamente definidas, com critérios de transparência, prestação de contas e equidade no acesso. Do contrário, corre-se o risco de que o marco normativo chegue sempre tarde e incompleto. Essa pode ser a situação da chamada “*EU AI Act*”, a primeira legislação integral sobre Inteligência Artificial no mundo, cujo ritmo de implementação é muito inferior ao do desenvolvimento de novos sistemas e aplicações. Observamos, além disso, que nela se contemplam exceções discutíveis para governos e exércitos, por exemplo.

O paralelismo entre a engenharia genética do final do século XX e a IA neste primeiro quarto do XXI é evidente: em ambos os casos, tratam-se de tecnologias com um potencial profundamente disruptivo, capazes de alterar não apenas processos técnicos, mas também estruturas sociais, marcos regulatórios e concepções éticas fundamentais. Na engenharia genética, marcos como a clonagem da ovelha Dolly, em 1996, ou o desenvolvimento de organismos geneticamente modificados deram lugar a debates sobre segurança, consentimento, impactos ecológicos e limites morais da intervenção humana na natureza. Do mesmo modo, a IA atual coloca questões — já mencionadas — sobre privacidade, vieses, deslocamento do trabalho e a delegação de decisões críticas a sistemas opacos. Em ambos os casos, a tecnologia avança mais rápido que a capacidade das instituições de compreendê-la e regulá-la. Existe, entretanto, uma diferença importante: a coesão geopolítica internacional. A capacidade dos Estados e dos atores internacionais para cooperar e agir de maneira unificada no âmbito global, superando divisões e conflitos, não é hoje a mesma que em 2000 (quando foi assinado o Protocolo de Cartagena sobre biossegurança em genética, destinado a garantir a conservação da biodiversidade e a saúde humana).

The fundamental issue is that AI cannot regulate itself, nor can it be left in the hands of those who have ideological or economic motives to obscure the clarity of its impacts. It requires regulation that not only limits harm but also guides technological development toward collectively defined goals, with transparency, accountability, and equity in access. Otherwise, there is a risk that regulation will always arrive late and incomplete. This may be the case with the so-called “EU AI Act,” the world’s first comprehensive legislation on Artificial Intelligence, whose implementation pace is far slower than that of new systems and applications being developed. Moreover, it includes questionable exceptions for governments and military institutions, for instance.

The parallel between genetic engineering in the late 20<sup>th</sup> century and AI in the first quarter of the 21<sup>st</sup> is evident: in both cases, we are dealing with profoundly disruptive technologies capable of transforming not only technical processes but also social structures, regulatory frameworks, and fundamental ethical conceptions. In genetic engineering, milestones such as the cloning of Dolly the sheep, in 1996, or the development of genetically modified organisms gave rise to debates on safety, consent, ecological impacts, and the moral limits of human intervention in nature. Likewise, current AI raises questions—already mentioned—about privacy, bias, labor displacement, and the delegation of critical decisions to opaque systems. In both cases, technology advances faster than the capacity of institutions to understand and regulate it. There is, however, an important difference: the degree of international geopolitical cohesion. The ability of states and international agents to cooperate and act in a unified way at the global level, overcoming divisions and conflicts, is not the same today as it was in 2000 (when the Cartagena Protocol on Biosafety in Genetic Engineering was signed, aimed at ensuring the conservation of biodiversity and human health).



Não faltam acordos de organismos internacionais, como a “Recomendação sobre a Ética da Inteligência Artificial”, adotada por unanimidade pelos 193 Estados-membros da Unesco em novembro de 2021 — o primeiro marco normativo global sobre ética da IA —, para orientar os países sobre como lidar com seus impactos positivos e negativos. Não faltam recomendações e acordos, mas pode faltar vontade de cumpri-los.

Na medida em que a IA afeta direitos fundamentais, condições de trabalho, acesso ao conhecimento e qualidade da deliberação pública, afetando diretamente o usuário, qualquer marco regulatório que seja desenhado sem a participação dos usuários corre o risco de responder unicamente a interesses corporativos ou a agendas estatais, deixando sem voz aqueles que são seus destinatários finais. A implicação da cidadania não apenas enriquece o debate com perspectivas diversas, mas também aumenta a aceitação social das decisões regulatórias. No caso da IA, isso implica abrir canais para que a cidadania contribua a definir quais aplicações considera aceitáveis, em que condições e com quais salvaguardas. Sem a intervenção dos usuários, o debate sobre os limites e o uso legítimo da IA ficará reduzido a uma negociação entre especialistas e corporações, deixando de fora a pluralidade de interesses e visões que conformam a vida democrática.

Para que essa participação seja efetiva, é indispensável ter um letramento midiático e informacional que permita compreender as implicações técnicas, éticas e sociais da IA e de qualquer outro meio. Sem uma formação básica em Tecnologias da Informação e Comunicação (TICs), a intervenção cidadã corre o risco de ser simbólica, capturada por discursos simplistas ou dominada por atores com maior poder de influência.

International bodies have not been lacking in producing agreements, such as the “Recommendation on the Ethics of Artificial Intelligence,” unanimously adopted by UNESCO’s 193 Member States in November 2021—the first global normative framework on AI ethics—, intended to guide countries in addressing its positive and negative impacts. What is lacking is not recommendations or agreements, but rather the political will to comply with them.

As AI affects fundamental rights, working conditions, access to knowledge, and the quality of public deliberation, directly impacting users, any regulatory framework designed without user participation risks responding solely to corporate interests or state agendas, leaving voiceless those who are its ultimate recipients. Citizen involvement not only enriches the debate with diverse perspectives but also increases social acceptance of regulatory decisions. In the case of AI, this means opening channels for citizens to help define which applications they consider acceptable, under what conditions, and with what safeguards. Without user involvement, the debate on the limits and legitimate uses of AI will be reduced to a negotiation between experts and corporations, excluding the plurality of interests and perspectives that shape democratic life.

For this participation to be effective, it is essential to have a form of media and information literacy that allows for an understanding of the technical, ethical, and social implications of AI, as well as any other medium. Without a basic education in Information and Communication Technologies (ICTs), civic engagement risks being merely symbolic, co-opted by simplistic narratives, or dominated by external agents with greater influence.



Debates sobre os usos e limites da Inteligência Artificial devem incluir a perspectiva do usuário final — mas isso precisa ser feito de forma crítica, o que exige formas específicas de letramento para permitir que as pessoas compreendam adequadamente a epistemologia dessas ferramentas (ou seja, a maneira como elas constroem o conhecimento que posteriormente irão disseminar)

Debates on the uses and limitations of Artificial Intelligence must include the perspective of the end user—but they need to be conducted critically, which requires specific forms of literacy in order to enable people to properly understand the epistemology of these tools (that is, the way they construct the knowledge they will subsequently disseminate)