



SEÇÃO: ÉTICA E FILOSOFIA POLÍTICA

## Sobre a eficiência da ética como ferramenta de governança da inteligência artificial

*On the efficiency of ethics as a governing tool for artificial intelligence*

*Sobre la eficacia de la ética como herramienta rectora de la inteligencia artificial*

**Nicholas Kluge Corrêa<sup>1</sup>**

[0000-0002-5633-6094](mailto:0000-0002-5633-6094)

[nicholas.correa@acad.pucrs.br](mailto:nicholas.correa@acad.pucrs.br)

**Nythamar Fernandes de Oliveira<sup>1</sup>**

[0000-0001-9241-1031](mailto:0000-0001-9241-1031)

[nythamardeoliveira@gmail.com](mailto:nythamardeoliveira@gmail.com)

**Diogo Fernando**

**Massmann<sup>1</sup>**

[0000-0001-5944-7139](mailto:0000-0001-5944-7139)

[diogofm1818@gmail.com](mailto:diogofm1818@gmail.com)

**Recebido em:** 10 jan. 2022.

**Aprovado em:** 28 abr. 2022.

**Publicado em:** 11 ago. 2022.

**Resumo:** A 4ª Revolução Industrial é o culminar da era digital. Atualmente, tecnologias como robótica, nanotecnologia, genética e inteligência artificial prometem transformar nosso mundo e a maneira como vivemos. O campo da Segurança e da Ética da Inteligência Artificial (IA) são áreas de pesquisa emergentes que vêm ganhando popularidade nos últimos anos. Diversas organizações de cunho privado, público e não governamentais têm publicado diretrizes propondo princípios éticos para a regulamentação do uso e desenvolvimento de sistemas inteligentes autônomos. Meta-análises do campo de pesquisa em Ética da IA apontam uma convergência sobre certos princípios éticos que, supostamente, governam a indústria da IA. Entretanto, pouco se sabe sobre a eficiência desta forma de "Ética". Neste estudo, gostaríamos de realizar uma análise crítica do atual estado da Ética da IA, e sugerir que essa forma de governança baseada em diretrizes éticas principialista não é suficiente para normatizar a indústria da IA e seus desenvolvedores. Acreditamos que drásticas mudanças sejam necessárias, tanto nos processos de formação de profissionais das áreas ligadas ao desenvolvimento de *software* e sistemas inteligentes quanto no aumento da regulamentação desses profissionais e sua indústria. Para tanto, sugerimos que o Direito se beneficie das contribuições recentes da Bioética, de forma a explicitar em termos legais as contribuições da Ética da IA para a governança.

**Palavras-chave:** Diretrizes Éticas. Ética da Inteligência Artificial. Governança. Bioética. Regulamentação.

**Abstract:** The 4th Industrial Revolution is the culmination of the digital age. Nowadays, technologies such as robotics, nanotechnology, genetics, and artificial intelligence promise to transform our world and the way we live. The field of Artificial Intelligence Safety and Ethics are emerging research areas that have been gaining popularity in recent years. Several private, public and non-governmental organizations have published guidelines proposing ethical principles for regulating the use and development of autonomous intelligent systems. Meta-analyses of the AI Ethics research field point to convergence on certain ethical principles that supposedly govern the AI industry. However, little is known about the effectiveness of this form of "Ethics." In this paper, we would like to conduct a critical analysis of the current state of AI Ethics and suggest that this form of governance based on principled ethical guidelines is not sufficient to norm the AI industry and its developers. We believe that drastic changes are necessary, both in the training processes of professionals in the fields related to the development of software and intelligent systems and in the increased regulation of these professionals and their industry. To this end, we suggest that law should benefit from recent contributions from bioethics, to make the contributions of AI ethics to governance explicit in legal terms.

**Keywords:** AI Ethics. Ethical Guidelines. Governance. Bioethics. Regulation.

**Resumen:** La Cuarta Revolución Industrial es la culminación de la era digital. Hoy en día, tecnologías como la robótica, la nanotecnología, la genética y la inte-



Artigo está licenciado sob forma de uma licença  
[Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

<sup>1</sup> Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, RS, Brasil.

ligencia artificial prometem transformar nuestro mundo y la forma en que vivimos. El campo de la seguridad y la ética de la inteligencia artificial son áreas de investigación emergentes que han ido ganando popularidad en los últimos años. Varias organizaciones privadas, públicas y no gubernamentales han publicado lineamientos que proponen principios éticos para regular el uso y desarrollo de sistemas inteligentes autónomos. Los metanálisis del campo de investigación Ética de la IA apuntan a la convergencia en ciertos principios éticos que supuestamente rigen la industria de la IA. Sin embargo, poco se sabe sobre la eficacia de esta forma de "ética". En este artículo, nos gustaría realizar un análisis crítico del estado actual de la ética de la IA y sugerir que esta forma de gobierno basada en pautas éticas basadas en principios no es suficiente para normar la industria de la IA y sus desarrolladores. Creemos que son necesarios cambios drásticos, tanto en los procesos de formación de profesionales en los campos relacionados con el desarrollo de software y sistemas inteligentes como en el aumento de la regulación de estos profesionales y su industria. Para ello, sugerimos que el derecho se beneficie de los aportes recientes de la bioética, para hacer explícitos en términos jurídicos los aportes de la ética de la IA a la gobernanza.

**Palabras clave:** Ética de la IA. Guías éticas. Gobernanza. Bioética. Regulación.

## Introdução

Sistemas autônomos inteligentes estão se tornando cada vez mais parte de nossa sociedade. Tarefas e trabalhos anteriormente realizados por nós são agora delegados a sistemas incorporados com inteligência artificial (IA). A chamada 4ª Revolução Industrial é o culminar da era digital, onde tecnologias como robótica, nanotecnologia, genética e inteligência artificial prometem transformar o nosso mundo e a maneira como vivemos. E, atualmente, a mais acessível e massivamente utilizada dessas tecnologias é a IA. Antes de mais nada, convém observar que evita-se uma definição conclusiva de IA, como recomenda a UNESCO, justamente porque tal definição estaria mudando periodicamente na medida em que novas tecnologias e recursos técnicos permitam inovações no próprio conceito de IA.<sup>2</sup> Todavia se postularmos que a inteligência humana é o que viabiliza a execução de tarefas tais como a percepção visual, o reconhecimento de fala, a

tomada de decisão e a tradução entre idiomas, a IA seria compreendida como um complexo técnico-teórico de sistemas computacionais e tecnologias de processamento de informações integrando modelos e algoritmos que produzam uma capacidade de aprender e realizar tarefas cognitivas que levam a resultados como a previsão e a tomada de decisão em ambientes materiais e virtuais, incluindo o aprendizado de máquina, o aprendizado profundo e o aprendizado por reforço – *machine learning, deep learning, and reinforcement learning* (geralmente os dois segundos são incluídos em uma visão abrangente do primeiro) (RUSSELL; NORVIG, 2022, p. 669-738). Pode-se, ainda, distinguir, seguindo John Searle (1980), entre IA fraca (*weak AI*) – a ideia de que máquinas de IA poderiam agir como se fossem inteligentes – e IA forte (*strong AI*) – a afirmação de que as máquinas que fazem isso estão realmente pensando conscientemente e não apenas simulando o pensamento. Com o tempo, a definição de IA forte passou a se referir ao que também tem sido chamado de "IA de nível humano" ou "IA geral" (*general AI*) – programas que podem resolver uma gama arbitrariamente ampla de tarefas, incluindo tarefas novas, e podem fazê-lo tão bem quanto um ser humano (RUSSELL; NORVIG, 2022, p. 1032).

A fim de ilustrar tais sistemas autônomos inteligentes, podemos citar os exemplos abaixo, como alguns tipos de utilização da IA pelo poder público brasileiro:<sup>3</sup>

- a) a *Agência Nacional de Telecomunicações* utiliza modelos de linguagem natural para identificar o comportamento padrão dos consumidores;
- b) a *Agência Nacional de Transportes Terrestres* utiliza IA para prever o fluxo médio diário de tráfego nas rodovias federais;
- c) o *Banco do Brasil* utiliza modelos de linguagem natural (*chatbots*) para atendimento ao consumidor, CNNs para reconhecimento facial;
- d) a *Caixa Econômica Federal* utiliza IA para

<sup>2</sup> UNESCO. Recommendation on the ethics of artificial intelligence. 41<sup>st</sup> session of the General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Paris, from 9 to 24 nov. 2021. Disponível em: <https://en.unesco.org/artificial-intelligence/ethics>. Acesso em: 9 jan. 2022

<sup>3</sup> GOVERNANCE recommendations – Use of Artificial Intelligence by Public Authorities. In: *Transparência Brasil*. [S. l.], 2021. Disponível em: [https://www.transparencia.org.br/downloads/publicacoes/Governance\\_Recommendations.pdf](https://www.transparencia.org.br/downloads/publicacoes/Governance_Recommendations.pdf). Acesso em: 9 jan. 2022.

predição de transações eletrônicas fraudulentas;

e) a *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* utiliza IA para confirmar a autoria de publicações e projetos acadêmicos;

f) o *Departamento de Polícia Federal* utiliza IA para reconhecimento facial e modelos de linguagem natural para previsão de riscos (e.g., detecção de fraudes);

g) a *Empresa Brasileira de Pesquisa Agropecuária* utiliza IA para prever o melhor vegetal a ser cultivado em uma lavoura e classificação de imagem para detectar doenças nas culturas agrícolas;

h) o *Instituto Nacional do Seguro Social* utiliza IA para prever irregularidades na emissão de benefícios sociais;

i) o *Supremo Tribunal Federal* utiliza IA para categorizar os procedimentos legais sob repercussão geral;

j) o *Supremo Tribunal de Justiça* utiliza IA para realizar varreduras automáticas para cada recurso apresentado ao tribunal, cruzar recursos com decisões anteriores (busca de precedentes legais), e geração de recomendações para os ministros.

E existem muitos mais casos de aplicação desse tipo de tecnologia, não apenas pelo setor público, mas por todos os setores que, de alguma forma, interagem com a nossa sociedade.

Dado o poder transformador da IA, seria importante que tivéssemos um consenso sobre as normas e as diretrizes que regulam tais tecnologias. O campo da segurança e da ética da IA (*AI Safety & AI Ethics*) são áreas de pesquisa emergentes que ganharam popularidade nos últimos anos. Diversas organizações privadas, públicas e não governamentais têm publicado diretrizes que propõem princípios éticos para melhorar a regulamentação de sistemas inteligentes autônomos (RUSSELL; DEWEY; TEGMARK, 2015; BODDINGTON, 2017; GOLDSMITH; BURTON, 2017; GREENE *et al.*, 2019).

No presente artigo, focaremos nos desafios normativos e nos problemas inerentes à Ética da IA. Por mais que este seja um campo relativamente novo da Ética, a Ética da IA já possui literatura o bastante para que as meta-análises do campo

tenham sido realizadas (JOBIN; IENCA; VAYENA, 2020; HAGENDORFF, 2020). Um dos resultados que podemos observar nessas meta-análises é que aparentemente existe uma convergência para onze princípios éticos genéricos (principalmente quando a Ética da IA é abordada de uma forma principiológica, *i.e.*, através de guias e diretrizes éticas): Transparência, Justiça/Equidade, Não maleficência, Responsabilidade, Privacidade, Caridade, Liberdade/Autonomia, Confiança, Dignidade, Sustentabilidade e Solidariedade. Destes onze princípios, os mais citados pela literatura são:

a) responsabilização (Como tornar a indústria de IA responsável por suas tecnologias?);

b) transparência (Como entender e interpretar os processos decisórios de sistemas de IA?);

c) privacidade (Como preservar a privacidade dos indivíduos em uma era em que "dados" são o combustível que impulsiona a indústria tecnológica?).

Entretanto, um criticismo que vem sendo levantado contra o estado atual da Ética da IA é sua aparente negligência com certas questões (CORRÊA; DE OLIVEIRA, 2021). Das 22 principais diretrizes éticas publicadas nos últimos cinco anos (HAGENDORFF, 2020), apenas nove mencionam direitos trabalhistas e desemprego tecnológico; oito mencionam a militarização da IA e a nova corrida armamentista; enquanto apenas duas diretrizes mencionam a falta de diversidade no desenvolvimento e governança da IA.

Ao mesmo tempo, críticas também têm sido feitas (RESSÉQUIER; RODRIGUES, 2020) com relação a este aparente uso da "Ética" como uma ferramenta de regulamentação e governança. Afinal de contas:

a) qual seria o verdadeiro poder normativo de diretrizes éticas para a indústria de IA e seus desenvolvedores?

b) seria a Ética da IA uma ferramenta eficiente de governança?

Essas serão as perguntas que tentaremos responder neste artigo. Ademais, buscaremos traçar um paralelo entre a Bioética/Ética Médica e a Ética da IA, para assim tentar buscar uma saída para este modelo de governança principiológica.

## 1 Bioética versus Ética da IA

Desde o início do século XX a Ética, enquanto estudo filosófico da moral, tem sido tradicionalmente dividida em três grandes domínios, a saber: Metaética, Ética Normativa e Ética Aplicada. Teorias éticas, amplamente interpretadas, procuram justificar os argumentos e o raciocínio moral ao lidar com problemas morais comuns, tais como: "por que matar é moralmente errado?" ou "como devemos levar uma vida moralmente boa?" (SMITH, 1994). A Ética Aplicada é uma disciplina da filosofia que se propõe a aplicar teorias éticas a procedimentos e processos reais de tomada de decisão. Podemos utilizar a Bioética como um "caso de sucesso" da Ética Aplicada, pois debates éticos em relação às questões da ética médica e da saúde tais como aborto, eutanásia, assistência à saúde, pesquisa com células-tronco, clonagem, eugenia, pesquisa genética, lentamente vêm ganhando projeção e representação em nossa esfera legislativa. Observamos, *en passant*, que a Bioética abrange a Ética Médica, mas não se reduz a ela. Essa é, outrossim, uma forma exemplar de como os debates éticos podem, por intervenção do Direito, serem transcompilados em ferramentas de governança (*e.g.*, leis). Ao utilizar o conceito de governança, estamos nos referindo a:

[...] processos empreendidos por um governo, ou qualquer forma de organização, sobre outras organizações, formais ou informais, na forma de leis, normas ou poder, com o intuito de determinar um padrão de conduta previamente estabelecido (BEVIR, 2012, p. 2).

Com essa definição em mente, gostaríamos de levantar a seguinte questão: "Seriam diretrizes éticas ferramentas de governança eficientes para a indústria da IA?"

Para Calo (2017), diretrizes éticas acabam servindo apenas como uma estratégia de *marketing*, não mitigando os riscos nem impondo consequências reais quando princípios éticos são transgredidos ou danos são causados aos impactados:

[...] Vários esforços estão em andamento, dentro da indústria, do meio acadêmico, e de outras organizações, para resolver a ética de IA. Mas

esses esforços provavelmente não podem substituir a elaboração de políticas (CALO, 2017, p. 407-408).

Críticas como a de Calo têm se tornado cada vez mais comuns na literatura. Floridi *et al.* (2018) realizaram uma comparação entre os princípios éticos levantados dentro da Ética da IA e aqueles defendidos pela Bioética enquanto Ética Médica. Os autores mostraram que quatro princípios clássicos da Ética Médica (*i.e.*, respeito pela autonomia humana, prevenção de danos/não maleficência, beneficência e justiça/equidade) são reinterpretados dentro do contexto da Ética da IA.

Contudo, não podemos pressupor que tal abordagem principiológica/fundacionista (dentro da própria Ética Médica e Bioética) possui diversas críticas dentro da literatura. Podemos citar Norman Daniels (1985), que buscou transpor a ideia rawlsiana do equilíbrio reflexivo para o contexto da Bioética em sua obra *Just Health Care*. Nela, Daniels critica o problema do fundacionismo inerente ao principialismo, que deveria ser agora repensado em termos narrativos coerentes como uma justificação plausível (em contraste com uma explicação da verdade ou como um sistema de crenças básicas enquanto princípios irreduzíveis).

Segundo Daniels (1985), uma abordagem reflexiva (*i.e.*, equilíbrio reflexivo), em contraponto com uma abordagem principialista em bioética, proporcionaria uma forma de "atualização moral", onde novos contextos causam uma mudança na interpretação dos próprios princípios/fundamentos que baseiam a nossa ética, e por consequência, as leis que devem normatizar a nossa sociedade.

Um outro aspecto propriamente ético-normativo e justificatório de toda a pesquisa em Bioética diz respeito à demarcação da pesquisa multidisciplinar nas interfaces da filosofia moral e das ciências jurídicas com a medicina, a biologia e as ciências da saúde em geral. Assim, cada vez mais notamos que a Ética da IA, assim como no caso da Bioética, não pode prescindir sem o mesmo tipo de articulação multidisciplinar que a Bioética promove. Ademais, pode-se ainda pensar na articulação teórico-conceitual e na inovação

tecnológica proporcionadas pela mediação da Ética da IA entre a Bioética e a Neuroética, tomando essa, segundo a formulação lapidar de Adina Roskies (2021), enquanto "ética da neurociência" dentro do vasto "campo interdisciplinar enfocando questões éticas levantadas por nossa compreensão crescente e constantemente aprimorada do cérebro e de nossa capacidade de monitorá-lo e influenciá-lo". Como Walter Glannon (2006, p. 2) argumentou apropriadamente, a capacidade de mapear, intervir e alterar os correlatos neurais da mente levanta decerto importantes questões éticas. Mas, para além disso, tais questões são indiscutivelmente mais instigantes e importantes do que qualquer outro conjunto de questões normativas em qualquer outra área da Bioética, na medida em que as técnicas que visam o cérebro podem revelar e modificar a fonte da mente e afetar a identidade pessoal, a vontade e outros aspectos de nós mesmos.

Agora que traçamos esse estreito paralelo da Bioética/Ética Médica com a Ética da IA e a situamos com relação à emergente pesquisa interdisciplinar em Neuroética e as suas novas interfaces tecnológicas (neuroimagem, neuro-modulação e redes neurais são apenas as mais conhecidas), a pergunta que levantamos é a seguinte: "Seria uma comparação entre a Bioética/Ética Médica, que no ocidente pode ser rastreada até os tempos de Hipócrates e Sócrates, com o muito mais recente campo da Ética da IA, uma comparação justa?".

Para Mittelstadt (2019), uma comparação entre a Ética da IA e Ética Médica não pode ser feita no estado atual em que a Ética da IA se encontra, principalmente, por certos déficits que a indústria da IA e, mais especificamente, os seus profissionais, possuem com relação à regulamentação, códigos de conduta e diretrizes éticas "vigentes". Mittelstadt (2019) aponta que a prática do desenvolvimento de sistemas inteligentes (e de todo desenvolvimento de *software* na verdade) carece de alguns pontos. Pontos que a tornam totalmente diferente da prática da Medicina e outros tipos de área de atuação (*e.g.*, Direito, Psicologia, Fisioterapia):

*a) objetivos e deveres comuns:* enquanto a prática da Medicina é norteadas por um objetivo comum (*i. e.*, a restauração e a manutenção da saúde), a indústria da IA não possui nenhum tipo de objetivo em comum. Os objetivos da indústria da IA são ditados pela demanda do mercado e não por um objetivo moral comum;

*b) tradição:* a Medicina possui uma longa tradição de princípios éticos que ajudaram a nortear a sua regulamentação (*e. g.*, juramento Hipocrático, Formula Comititis Archiatrorum, Declaração de Genebra, Declaração de Helsinki), e definir o que queremos dizer com uma conduta ética na área de Medicina (BODENHEIMER; JONSEN, 2001). Enquanto isso, a Ética da IA ainda não possui uma tradição de conduta robusta;

*c) responsabilização legal:* existe todo um aparato social que garante que se a prática da medicina for realizada de forma antiética ou insegura, tais praticantes serão punidos com consequências reais. Além disso, praticantes da Medicina possuem uma intensa formação acadêmica onde devem corresponder a altos padrões de qualidade. Após o término de sua formação, tais profissionais ainda estão sujeitos a perder sua licença, sendo regulados por vigilantes comitês de ética. Esse mecanismo fornece uma maneira de ligar o "mau comportamento" com sanções legais. Já no caso da engenharia de *software* nenhum tipo de licença é necessário para praticar a profissão. Institutos e associações regulamentadoras da área, como a IEEE (Institute of Electrical and Electronic Engineers) e a ACM (Association for Computing Machinery), não possuem poder para afetar a prática desses profissionais (GOERTZEL, 2016). Profissões relacionadas ao desenvolvimento de *software* ainda não são reconhecidas legalmente como profissões regulamentadas em diversos países, não havendo um padrão de licenciamento geral (ZOLLERS *et al.*, 2005).

Mittelstadt (2019) também argumenta que profissões caracterizadas por possuírem uma robusta estrutura ética embutida em sua prática têm:

a) educação e treinamento especializados para lidar com dilemas éticos;



b) compromisso com o serviço público e objetivos pró-sociais;

c) um grande padrão de cuidado;

d) autogovernança instituída por organizações de profissionais da área que também licenciam a prática da profissão.

Contudo, como a própria profissão de "programador" ou "desenvolvedor de *software*" ainda não é regulamentada em muitos países (como o Brasil), acreditamos que não seja possível dizer que os profissionais que desenvolvem sistemas autônomos inteligentes contemplem essa série de critérios. Assim, como podemos traduzir os princípios éticos levantados pela literatura da Ética da IA (como a Bioética realiza) para mecanismos robustos de governança, se não temos "sobre quem" impor as leis/normas que criarmos?

## 2 Ética no desenvolvimento de *software*

O desenvolvimento de sistemas inteligentes ainda é, em grande parte, engenharia de *software*, onde Aprendizagem de Máquina (*Machine Learning*) representa apenas uma pequena porção desse desenvolvimento (SCULLEY *et al.*, 2015). Com aproximadamente 30 anos de história, a Ética no Desenvolvimento de *Software* ainda é um campo de pesquisa muito recente. Dentro dessa literatura, encontramos autores como Friedman (1996), Bynum (2006), Davis e Nathan (2015), que argumentam a favor da ideia de "*Value Sensitive Design*", uma forma de projetar tecnologia levando em conta valores humanos durante o processo. Contudo, a ideia de *design* baseado em valores humanos, ou "*Human-Centeredness*" (*i.e.*, a ideia de se desenvolver tecnologias centradas nos seres humanos), ainda é algo pouco incorporado pela literatura padrão de Engenharia de *Software*, e muito menos nas práticas de formação de seus profissionais.

Para apoiar tal proposta, podemos citar um estudo controlado conduzido por McNamara *et al.* (2018). O único objetivo desse estudo consistia em investigar se as diretrizes éticas têm um efeito normativo na tomada de decisão de

desenvolvedores de *software*. Em sua pesquisa, os autores avaliaram 63 estudantes de engenharia de *software* e 105 desenvolvedores profissionais de *software*, analisando se as diretrizes éticas da ACM<sup>4</sup> teriam alguma influência sobre a decisão dos participantes ao lidar com dilemas morais relacionados ao desenvolvimento de *software*. O resultado encontrado pelos autores foi o seguinte: "Apesar de nosso objetivo declarado, não encontramos evidências de que o código de ética da ACM influencie a tomada de decisões éticas" (MCNAMARA; SMITH; MURPHY-HILL, 2018, p. 4).

Por mais que exista um grande número de diretrizes éticas publicadas, ainda há pouca pesquisa voltada para como podemos implementar princípios éticos na prática do Desenvolvimento de *Software* (MORLEY *et al.*, 2019). Vakkuri *et al.* (2020) argumentam que alguns dos motivos pelos quais a literatura da Ética da IA têm recebido pouca atenção por parte das áreas relacionadas ao Desenvolvimento de *Software* são:

a) pesquisas em Ética da IA são predominantemente filosóficas;

b) a área de Desenvolvimento de *Software* não vê sentido em abordar preocupações filosóficas;

c) dessa forma, a formação em Ética da IA não faz parte do sistema educacional.

No estudo de Vakkuri *et al.* (2020), os autores buscaram explorar como questões éticas são abordadas em ambientes semelhantes a *startups*, onde sistemas inteligentes são comumente produzidos. Os autores realizaram um estudo de caso qualitativo com três *startups* finlandesas da área da saúde, utilizando entrevistas semiestruturadas com seus desenvolvedores, supervisores e gerentes. A área de saúde foi escolhida pelo pressuposto de que considerações éticas "iriam receber maior atenção".

Nas entrevistas, os autores buscaram saber quais medidas concretas, sejam por ferramentas de programação, normas de conduta, verificação de *software* ou outra não citada, as empresas estavam tomando para tornar os seus produtos alinhados com três princípios éticos já bem sa-

<sup>4</sup> ACM Code of Ethics and Professional Conduct. Disponível em: <http://www.acm.org/binaries/content/assets/membership/images2/fac-stu-poster-code.pdf>. Acesso em: 9 jan. 2022.

lietados pela literatura:

- a) responsabilidade (Qual a responsabilidade dos desenvolvedores?);
- b) transparência (Como podemos explicar e entender a decisão de um algoritmo);
- c) responsabilização (Em caso de mau uso ou efeitos colaterais imprevistos, como os desenvolvedores devem lidar?).

Os autores do estudo argumentam que por mais que durante as entrevistas os participantes demonstrassem (às vezes) preocupações com problemas éticos, não havia nenhuma maneira clara de como os desenvolvedores estavam atendendo a tais preocupações no desenvolvimento dos seus produtos. Alguns trechos das entrevistas que gostaríamos de citar, relacionados aos três princípios éticos abordados são (VAKKURI *et al.*, 2020, p. 202-205):

- a) responsabilidade: "Nossa responsabilidade é manter o projeto dentro do prazo".
- b) transparência: "Os desenvolvedores reconhecem a transparência como um objetivo, mas ela não é formalmente perseguida".
- c) responsabilização: "Desenvolvedores não têm planos para lidar com o comportamento inesperado do sistema resultante, por exemplo, da Aprendizagem da Máquina ou da futura expansão do contexto de uso do sistema".

Desses achados, acreditamos que seja possível argumentar que:

- a) existe uma desconsideração por parte de certas empresas de desenvolvimento de *software* em relação à Ética profissional;
- b) certas empresas desconhecem a importância de seu papel em mitigar os efeitos colaterais de seus produtos;
- c) existe uma lacuna na formação de profissionais da área de Desenvolvimento de *Software*;
- d) o Desenvolvimento de *Software* ainda carece de ferramentas para aplicar princípios éticos em seu desenvolvimento prático.

Vários outros estudos corroboram com a ideia de que apenas diretrizes éticas não possuem efeito na tomada de decisão de profissionais (BRIEF *et al.*, 1996; CLEEK; LEONARD, 1998; LERE;

GUAMNITZ, 2003; OSBORN *et al.*, 2009). E tal ideia ressoa com várias críticas levantadas contra o atual estado da Ética da IA. De acordo com Jobin *et al.* (2020, p. 38g):

O envolvimento do setor privado no campo da ética da IA tem sido questionado por potencialmente usar políticas brandas de alto nível como uma forma de tornar um problema social, em algo técnico, ou para evitar completamente a regulamentação.

Ademais, segundo Hagendorff (2020, p. 99), "a Ética da IA — ou a Ética em geral — não tem mecanismos para reforçar suas próprias reivindicações normativas". Também poderíamos dizer, assim como Rességuier e Rodrigues (2020, p. 1), que a "Ética tem grandes dentes poderosos. Infelizmente, mal estamos usando-os na Ética da IA — não é de se admirar então que a Ética da IA seja chamada de *desdentada (toothless)*". De forma mais aplicada a contextos institucionais da vida social e em resposta aos desafios normativos da governança, Corrêa e De Oliveira (2021, p. 13) argumentam:

Como a teoria crítica, a Ética da IA deve se concentrar em destacar os aspectos negligenciados de nossa sociedade e sua relação com a indústria tecnológica, desafiando suas estruturas de poder para que a promessa de uma IA benéfica para todos possa ser cumprida. Não apenas como um ideal para o futuro da humanidade, mas também para as pessoas de hoje.

O ponto que queremos reforçar é que o papel social da Ética não é o de ser uma versão branda da lei, pois não é aí que a normatividade ética encontra seu poder justificatório. E em um sentido prático, o que precisamos hoje é de regulamentação, e não apenas de sugestões, opiniões e palpites.

### 3 A necessidade de regulamentação

Um dos estudos mais recentes a tratar sobre questões éticas relacionadas às noções de direitos humanos na governança da IA e Big Data foi o levantamento SHERPA (SANTIAGO, 2020), um estudo Delphi<sup>5</sup> comissionado pela União

<sup>5</sup> O método Delphi é um processo utilizado para chegar a uma opinião ou decisão de grupo através da consulta de um painel de

Europeia. O estudo, através de um questionário contendo as perguntas a seguir, obteve os seguintes resultados:

a) *Quais você acha que são as três questões éticas, ou de direitos humanos, mais importantes levantadas pela IA e / ou Big Data?* – As cinco questões éticas que receberam maior número de menções foram “falta de transparência” (19), “falta de privacidade” (17), “Viés (*Bias*) e discriminação algorítmica” (17), “perda de tomada de decisão humana” (12), “controle e uso malicioso de dados” (10). Já as questões menos mencionadas foram “concentração de Poder (centralização na governança da IA)” (5), “violação de direitos humanos” (4), “desaparecimento de empregos” (3), “despertar de uma Singularidade” (2), “impactos adversários não previstos” (1).

b) *Quais abordagens, métodos ou ferramentas atuais para abordar tais questões você conhece?* – A abordagem mais mencionada foi regulamentos através de leis e de legislações. Contudo, nenhum respondente identificou um instrumento de regulamentação internacional, enquanto alguns respondentes apontaram as limitações práticas de se criar e implementar uma abordagem internacional.

c) *Quais você acha que são os prós e os contras dessas abordagens, métodos ou ferramentas atuais?* – Críticas levantadas aos regulamentos foram as mais frequentemente citadas, argumentando que a regulamentação tende a “limitar a inovação”. Outra crítica feita fora que as medidas atuais são demasiadamente “abstratas e difíceis de medir”, e implementá-las seria muito custoso para a indústria. Outro ponto citado foi que a regulamentação ainda ocorre de forma muito lenta, incapaz de acompanhar o rápido avanço da indústria tecnológica.

d) *Quais devem ser os três principais problemas para que a sociedade selecione e priorize?* – Os problemas mais mencionados foram “impacto social” (12), “transparência” (10), “respeito aos direitos humanos” (5) e “impacto às minorias” (4).

e) O que você proporia para abordar melhor tais questões? – Novamente, a regulamentação fora

a proposta mais mencionada. Contudo, acordos internacionais ou regionais foram mencionados poucas vezes, talvez dado pela evidente dificuldade em se criar tais acordos. Um respondente propôs a criação de “jurados de cidadãos” como um meio de incentivar o diálogo entre a indústria tecnológica e aqueles que serão afetados por suas criações, uma ideia que vem sendo sugerida recentemente pela literatura (*i.e.*, a democratização da governança da IA) (CORRÊA; DE OLIVEIRA, 2021).

Como podemos ver, existe um certo consenso de que o que necessitamos não são novas “diretrizes éticas”, mas sim regulamentação. Novamente, gostaríamos de ressaltar que o “sucesso” da Bioética pode (principalmente) ser atribuído ao fato de que o debate gerado por sua temática pode ser mais facilmente implementado em processos legislativos, pois os profissionais e entidades atuantes das áreas da saúde já são fortemente regulamentadas por leis, conselhos de ética, conselhos de profissionais etc.

Contudo, não podemos dizer o mesmo pelo ainda “faroeste digital” que governa de forma desregrada a indústria de desenvolvimento de *software* e os sistemas inteligentes. Desenvolvedores não são regulamentados ou treinados sobre um robusto código de conduta ética. Nós tememos que a regulamentação “enrijeça” o nosso avanço tecnológico. Enquanto fatores como estes ainda são uma realidade, a Ética não pode ser usada como uma ferramenta de regulamentação para a indústria da IA. Nós precisamos transpor do principalismo para a regulamentação, e talvez o primeiro passo para isso seja a própria regulamentação daqueles que desenvolvem as tecnologias que prometem transformar a nossa sociedade. Em última análise, como argumentaram de forma independente Fabrício Polido (FRAZÃO; MULHOLLAND, 2020, p. 204) e Juarez Freitas e Thomas Freitas (2020), as justificativas para os projetos de lei em nosso País seriam várias, oscilando desde a preocupação emergente quanto à convivência humana com entes autônomos e inteligentes até questões de segurança



jurídica e da proteção dos direitos humanos, mas deveriam incluir também e de forma normativa diretrizes como a indelegabilidade da decisão intrinsecamente humana visando assegurar a responsabilidade moral e a própria sustentabilidade da vida humana. Ora, lembramos aqui que em 29 de setembro de 2021, a Câmara de Deputados da República Federativa do Brasil aprovou o Projeto de Lei n. 21/2020, estabelecendo fundamentos e princípios para o desenvolvimento e a aplicação da inteligência artificial em nosso País. O PL n. 21/2020 foi originalmente escrito pelo Deputado Eduardo Bismarck, reformulado pela atual presidente da Frente Digital, Deputada Luísa Canziani, e representa a louvável iniciativa de: a) sinalizar para o mundo que o Brasil está atento às preocupações pertinentes à IA; b) proporcionar maior segurança jurídica no uso destas tecnologias; e c) promover a disseminação da IA no Brasil. Atualmente, o projeto aguarda tramitação no Senado Federal, que o discutirá juntamente com os projetos de lei n. 872/2021, n. 5691/2019 e n. 5051/2019 e, ainda, pode sofrer modificações.

### Considerações finais

Neste artigo buscamos introduzir o leitor ao emergente campo de pesquisa em Ética da IA, onde uma de suas preocupações se encontra em estabelecer os princípios e os valores que devem guiar (e posteriormente normatizar) o desenvolvimento tecnológico deste tipo de tecnologia.

Buscamos também mostrar que isso não é nada novo (*i.e.*, o uso da ética para tentar regulamentar o avanço tecnológico de alguma área específica do saber). A Bioética é, como vimos, um excelente e instrutivo exemplo. Atualmente (mas não como se tomássemos tais questões como resolvidas), nossas preocupações morais em relações a questões como clonagem, engenharia genética, eutanásia, foram transportadas do debate ético para a área do saber responsável por criar, interpretar e defender as normas por quais concordamos viver em sociedade, a saber, o Direito.

Dependendo de onde vivemos (pois tais normas variam de cultura para cultura), profissionais

da área da saúde sabem os limites e deveres de sua atuação. E isso se dá por uma série de fatores que foram expostos no texto desse artigo (*i.e.*, objetivos e deveres comuns, tradição, responsabilização legal). Contudo, um paralelo entre a Bioética/Ética Médica não pode ser traçado sem que vejamos disparidades entre ambas as áreas.

Cada vez mais surgem críticas para a abordagem principiológica, pois (apenas) princípios abstratos não podem regulamentar a indústria da IA. Princípios éticos e valores humanos podem ser interpretados de inúmeras formas. A forma como ocidentais entendem o conceito de "Privacidade" talvez seja um pouco diferente da forma do que orientais entendem o mesmo princípio. E não há nada de errado nisso. Um dos pontos que gostaríamos de frisar novamente nesta conclusão é que, no estado atual da Ética da IA, constata-se que:

a) suas pesquisas são ainda predominantemente filosóficas, e as ferramentas que traduzam princípios abstratos para a prática do desenvolvimento de sistemas inteligentes ainda são poucas;

b) existe ainda pouca adesão por parte da comunidade geral para com as preocupações levantadas pela literatura da Ética da IA;

c) desenvolvedores não são regulamentados ou supervisionados por nenhum tipo de órgão ou conselho profissional, e a ética aplicada geralmente não faz parte do currículo de treinamento destes atores.

Portanto, não é possível dizer que a Ética da IA, por si só, seja capaz de normatizar a indústria da IA. Assim, ela não pode ser considerada uma ferramenta "eficiente" de governança. Se queremos que a Ética da IA seja algo mais do que um mero "teatro de ética" (*ethics washing*), precisamos, assim como a Bioética, encontrar formas de tornar nossas preocupações em legislações capazes de sancionar e guiar a indústria tecnológica. E, para isso, talvez o primeiro passo que necessitamos dar seja formalizar e regulamentar a atuação dos desenvolvedores desta indústria.

Se queremos viver em um futuro em que a IA seja criada com base em leis, e não em promessas e votos de confiança, talvez seja necessário

a própria indústria da IA, finalmente, *se profissionalizar*. E instituir o estudo da Ética (aplicada ao seu contexto) como parte fundamental de sua formação. Talvez, assim, o Direito tenha as ferramentas necessárias, como já é feito no caso da Bioética e da Ética Médica, para normatizar o uso e o desenvolvimento da inteligência artificial.

## Referências

- BEVIR, M. *Governance: A Very Short Introduction*. UK, Oxford: OUP Oxford, 2012.
- BODDINGTON, P. Towards a code of ethics for artificial intelligence. *Springer International Publishing*, Cham, Switzerland 2017. <https://doi.org/10.1007/978-3-319-60648-4>.
- BODENHEIMER, T.; JONSEN, A. A Short History of Medical Ethics. *Journal of Public Health Policy*, Basingstoke, UK, v. 22, n. 2, p. 247-249, 2001. <https://doi.org/10.2307/3343467>.
- BRIEF, A. P.; DUKERICH, J. M.; BROWN, P. R.; BRETT, J. F. What's wrong with the treadway commission report? Experimental analyses of the effects of personal values and codes of conduct on fraudulent financial reporting. *Journal of Business Ethics*, Charlottesville, Virginia, v. 15, n. 2, p. 183-198, 1996. <https://doi.org/10.1007/BF00705586>.
- BYNUM, T. Flourishing ethics. *Ethics Inf. Technol. [S. l.]*, v. 8, n. 4, p. 157-173, 2006.
- CALO, R. Artificial intelligence policy: a primer and roadmap. *SSRN Journal*, [S. l.], v. 51, p. 399-435, 2017. <https://doi.org/10.2139/ssrn.3015350>.
- CLEEK, M. A.; LEONARD, S. L. Can corporate codes of ethics influence behavior? *Journal of Business Ethics*, [S. l.], v. 17, n. 6, p. 619-630, 1998. <https://doi.org/10.1023/A:1017969921581>.
- CORRÊA, N. K.; DE OLIVEIRA, N. F. Good AI for the Present of Humanity Democratizing AI Governance. *AI Ethics Journal*, [S. l.], v. 2, n. 2, p. 1-16, 2021. <https://doi.org/10.47289/AIEJ20210716-2>.
- DANIELS, N. *Just Health Care*. New York: Cambridge University Press, 1985.
- DAVIS, J.; NATHAN, L.P. Value sensitive design: applications, adaptations, and critiques. In: VAN DEN HOVEN, J.; VERMAAS, P. E.; VAN DE POEL, I. (ed.). *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*. Dordrecht: Springer, 2015. p. 1-40.
- FLORIDI, L.; COWLS, J.; BELTRAMETTI, M. et al. AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, Dordrecht, NE, v. 28, p. 689-707, 2018. <https://doi.org/10.1007/s11023-018-9482-5>.
- FRAZÃO, A.; MULHOLLAND, C. (org.). *Inteligência Artificial e Direito*. 2. ed. São Paulo: Thomson Reuters Brasil, 2020.
- FREITAS, J.; FREITAS, T. *Direito e inteligência artificial: Em defesa do humano*. Belo Horizonte: Editora Fórum, 2020.
- FRIEDMAN, B.; KAHN, P.H.; BORNING, A.; HULDTGREN, A. Value sensitive design and information systems. In: *Early engagement and new technologies: opening up the laboratory*. Berlin: Springer, 2013. p. 55-95.
- GLANNON, W. Neuroethics. *Bioethics*, Oxford, UK, v. 20, n. 1, p. 37-52, 2006.
- GOERTZEL, K. M. Legal liability for bad software. *Cross-Talk*, Force Base, Utah v. 29, n. 5, p. 23-28, 2016.
- GOLDSMITH, J.; BURTON, A. E. Why teaching ethics to AI practitioners is important. In: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 31., 2017. *Proceedings [..]*. [S. l.]: AAAI-17, 2017. p. 4863-4840, 2017. Disponível em: <https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14271/13992>. Acesso em: 9 jan. 2022.
- GREENE, D.; HOFMANR, A. L.; STARK, L. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 52., 2019. *Proceedings [..]*. [S. l.]: HICCL, 2019. p. 2122-2131. <https://doi.org/10.24251/HICSS.2019.258>.
- HAGENDORFF, T. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, Dordrecht, NE, v. 30, p. 99-120, 2020. <https://doi.org/10.1007/s11023-020-09526-7>.
- JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of AI ethics guidelines. *Nat Mach Intell. [S. l.]*, n. 1, p. 389-399, 2019. <https://doi.org/10.1038/s42256-019-0088-2>.
- LERE, J. C.; GAUMNITZ, B.R. The Impact of Codes of Ethics on Decision Making: Some Insights from Information Economics. *Journal of Business Ethics*, [S. l.], n. 48, p. 365-379, 2003. <https://doi.org/10.1023/B:BUSI.000000574737500.c8>.
- MCMANARA, A.; SMITH, J.; MURPHY-HILL, E. Does ACM's code of ethics change ethical decision making in software development? In: ESEC/FSE 2018: ACM JOINT MEETING ON EUROPEAN SOFTWARE ENGINEERING CONFERENCE AND SYMPOSIUM ON THE FOUNDATIONS OF SOFTWARE ENGINEERING, 26., 2018. *Proceedings [..]*. [S. l.], ESEC/FSE, 2018. p. 729-733. <https://doi.org/10.1145/3236024.3264833>.
- MITTELSTADT, B. Principles alone cannot guarantee ethical AI. *Nat Mach Intell. [S. l.]*, n. 1, p. 501-507, 2019. <https://doi.org/10.1038/s42256-019-0114-4>.
- MORLEY, J.; et al. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. In: ArXiv. [S. l.], 2019. Disponível em: <https://arxiv.org/abs/1703.04741>. Acesso em: 9 jan. 2022.
- OSBORN, M.; DAY, R.; KOMESAROFF, P.; MANT, A. Do ethical Guidelines make a difference to decision-making? *Internal medicine journal*, [S. l.], v. 39, n. 12, p. 800-805, 2009. <https://doi.org/10.1111/j.1445-5994.2009.01954.x>.

RESSÉGUIER, A.; RODRIGUES, R. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, Waterford, Ireland, v. 7, n. 2, p. 1-5, 2020. <https://doi.org/10.1177/2053951720942541>.

ROSKIES, A. Neuroethics. In: ZALTA, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy*. [S. l.], 2021. (Spring 2021 Edition) Disponível em: <https://plato.stanford.edu/archives/spr2021/entries/neuroethics>. Acesso em: 9 Jan 2022

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 4. ed. London: Pearson, 2022.

RUSSELL, S.; DEWEY, D.; TEGMARK, M. An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. Open Letter. Signed by 8,600 people. In: *Future Online*. [S. l.], 2015. Disponível em: [https://futureoflife.org/data/documents/research\\_priorities.pdf](https://futureoflife.org/data/documents/research_priorities.pdf). Acesso em: 9 jan. 2022.

SANTIAGO, N. SHERPA Delphi Study – Round 1 Results [Project Deliverable]. SHERPA project, 2020. Disponível em: <https://www.project-sherpa.eu/wp-content/uploads/2020/03/sherpa-delphi-study-round-1-summary-17.03.2020.docx.pdf>. Acesso em: 9 jan. 2022.

SCULLEY, D. *et al.* Hidden technical debt in machine learning systems. *Adv. Neural Inf. Process. Syst.*, [S. l.], n. 2, p. 2503-2511, 2015.

SEARLE, J. R. Minds, brains, and programs. *The Behavioral and Brain Sciences*, [S. l.], n. 3, p. 417-457, 1980.

SMITH, M. *The Moral Problem*. Oxford: Blackwell, 1994.

UNESCO. Recommendation on the ethics of artificial intelligence. 41st session of the General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO). Paris, from 9 to 24 November, 2021. Disponível em: <https://en.unesco.org/artificial-intelligence/ethics>. Acesso em: 9 Jan 2022

IVAKKURI, V.; KEMELL, K. K.; JANTUNEN, M.; ABRAHAMSON, P. "This is Just a Prototype": How Ethics Are Ignored in Software Startup-Like Environments. In: STRAY, V.; HODA, R.; PAASIVAARA, M.; KRUCHTEN, P. (ed.). *Agile Processes in Software Engineering and Extreme Programming*. XP 2020. Lecture Notes in Business Information Processing, [S. l.], v. 383, p. 195-210, Springer: Cham, 2020. [https://doi.org/10.1007/978-3-030-49392-9\\_13](https://doi.org/10.1007/978-3-030-49392-9_13).

ZOLLERS, F. E.; MCMULLIN, A.; HURD, S. N.; SHEARS, P. No More Soft Landings for Software: Liability for Defects in an Industry That Has Come of Age. *Santa Clara High Tech. L. J.*, [S. l.], v. 21, p. 745-782, 2005. Disponível em: <https://digitalcommons.law.scu.edu/chtlj/vol21/iss4/4>. Acesso em: 9 jan. 2022.

---

## Nicholas Kluge Corrêa

Doutorando em Filosofia no PPGF da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), em Porto Alegre, RS, Brasil.

---

## Nythamar Fernandes de Oliveira

Doutor e professor no PPG em Filosofia da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), em Porto Alegre, RS, Brasil.

---

## Diogo Fernando Massmann

Mestre em Filosofia e Ciências Cognitivas pela Universidade Estadual Paulista (UNESP), em Marília, SP, Brasil. Doutorando em Filosofia pelo PPGF da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), em Porto Alegre, RS, Brasil.

---

## Nicholas Kluge Corrêa; Nythamar Fernandes de Oliveira; Diogo Fernando Massmann

Pontifícia Universidade Católica do Rio Grande do Sul  
Programa de Pós-Graduação em Filosofia  
Av. Ipiranga, 6681, Prédio 8, 4º andar, Sala 403  
Partenon, 90.619-900  
Porto Alegre, RS, Brasil

*Os textos deste artigo foram revisados pela Poá Comunicação e submetidos para validação do(s) autor(es) antes da publicação.*