



XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-5– Política e Economia da Informação

ÉTICA DA INFORMAÇÃO E ÉTICA EM INTELIGÊNCIA ARTIFICIAL: APROXIMAÇÃO NECESSÁRIA, MAS INCIPIENTE

INFORMATION ETHICS AND ETHICS IN ARTIFICIAL INTELLIGENCE: A NECESSARY, BUT INCIPIENT APPROACH.

Elizabeth Maria Freire de Jesus. UFRJ.

Modalidade: Trabalho Completo

Resumo: A discussão sobre inteligência artificial ética e confiável é sobretudo a discussão sobre responsabilidades. A construção de sistemas de Inteligência artificial envolve uma extensa rede de atores com conhecimentos, hábitos, interesses e prioridades possivelmente diferentes e conflituosas, levando, portanto, a existência de diferenças e percepções importantes sobre a dimensão ética da inteligência artificial. Considerando que necessitam dispor de um equipamento teórico-prático e de um conhecimento autorreflexivo sobre suas próprias práticas, alinhando-os aos conhecimentos disciplinares necessários para a sua atuação profissional, aponta-se a importância do desenvolvimento de uma competência crítica alicerçada e revigorada pelo indagar e refletir própria da ética, enquanto ramo da filosofia prática e, em especial, da ética da informação. O objetivo desse trabalho é investigar em documentos que tratam de ética em Inteligência artificial a existência de remissivas ao estímulo do desenvolvimento de competências, para além das competências técnica e instrumental, que possibilitem um contínuo pensamento crítico-reflexivo fundamentado na teoria e na prática de forma a orientar ações responsáveis, justificáveis e, sobretudo, transformadoras. Trata-se de uma pesquisa bibliográfica e documental. A partir da análise do *corpus* dos documentos, é possível perceber que o debate sobre inteligência artificial ética em diálogo com a ética da informação é incipiente.

Palavras-Chave: Inteligência Artificial Ética. Ética da Informação. Diretrizes sobre Inteligência Artificial.

Abstract: The debate over ethical and trustworthy artificial intelligence is primarily a discussion of responsibilities. The construction of artificial intelligence systems involves an extensive network of actors with possibly different and conflicting knowledge, habits, interests and priorities, thus leading to the existence of important differences and perceptions about the ethical dimension of artificial intelligence. Considering that they need to have theoretical-practical equipment and self-reflexive knowledge about their own practices, aligning them with the disciplinary knowledge necessary for their professional performance, the importance of developing a critical competence based and reinvigorated by asking and reflection proper to ethics, as a branch of practical philosophy and, in particular, of information ethics. The objective of this work is to investigate, in documents dealing with ethics in Artificial Intelligence, the existence of references to the stimulation of the development of competences, in addition to technical and instrumental competences, which allow for a continuous critical-reflective thinking based on theory and practice in a to guide responsible, justifiable and, above all, transformative actions. This is a bibliographic and documentary research. From the analysis of the corpus of documents, it is possible to perceive that the debate on ethical artificial intelligence in dialogue with the ethics of information is incipient.

Keywords: Ethical Artificial Intelligence. Information Ethics. Guidelines on Artificial Intelligence.



1 INTRODUÇÃO

Quaisquer discussões em torno do desenvolvimento e uso de Inteligência Artificial (IA) ética e, sobretudo confiável devem considerar como ponto de partida que as tecnologias não são moralmente neutras. Sistemas inteligentes são projetados, desenvolvidos, testados, aceitos e usados por pessoas, sendo, portanto, inescapavelmente carregados de valores humanos, culturais e sociais. As tecnologias “transformam a natureza mesma da relação entre o sujeito e o mundo e, por conseguinte, a própria autocompreensão do agente moral humano” (CAPURRO, 2012, p.39). Essas considerações apontam-nos para a necessidade de concentrar a atenção na perspectiva ética que exige o questionamento e a reflexão de como as tecnologias digitais afetam os modos de agir, de conhecer e produzir conhecimento, e também como reproduzem, reforçam ou criam formas de ser e (con)viver em sociedade.

A definição de IA é diversificada. Um entendimento abrangente que nos permite pensar, a partir de um olhar mais aguçado e crítico sobre a questão da intencionalidade ou do(s) propósito(s) “explícito(s)” e codificado(s) em sistemas de IA em sua relação com as responsabilidades *ex ante* que recaem nos atores envolvidos no *design*, desenvolvimento e processamento desses sistemas, e as responsabilidades *ex post* em relação aos resultados, sejam esses moralmente positivos, valiosos ou não.

Os sistemas de inteligência artificial (IA) são sistemas de software (e eventualmente também de hardware) concebidos por seres humanos, que, tendo recebido um objetivo complexo, atuam na dimensão física ou digital percebendo o seu ambiente mediante a aquisição de dados, interpretando os dados estruturados ou não estruturados recolhidos, raciocinando sobre o conhecimento ou processando as informações resultantes desses dados e decidindo as melhores ações a adotar para atingir o objetivo estabelecido. Os sistemas de IA podem adaptar o seu comportamento mediante a análise do modo como o ambiente foi afetado pelas suas ações anteriores (EUROPEAN COMMISSION, 2019, p.6).

A quase onipresença de sistemas inteligentes intensamente informados capazes de executar ações moralmente relevantes tem intensificado esforços no sentido de enfrentar os desafios para garantir um “comportamento ético”¹ desses agentes, haja vista as vastas evidências de impactos moralmente negativos desses sistemas de IA, como a criação ou reforço de *bias* em cenários como na previsão de atividades criminosas em áreas urbanas

¹ A expressão “comportamento ético” será utilizada para caracterizar ações de agentes de inteligência artificial que seriam consideradas éticas do ponto de vista da ação humana, segundo um conjunto de princípios e valores que tonalizam a moral vigente ancorada em determinado tempo e local.



O'NEIL, 2020); em avaliações de risco criminal (LARSON et al., 2016); na determinação do que passa por “filtros” de visibilidade que confinam cidadãos em “bolhas” (PARISER, 2012), interferindo naquilo que pode ser objeto de conhecimento e decisão coletiva; na manipulação do emocional como acontece pela mobilização da memória e da nostalgia a partir do processamento de sofisticadas técnicas de inteligência artificial que conseguem tridimensionalizar, colorizar, animar fotos antigas e dar “vida” ao passado (BEIGUELMAN, 2021, p.144) ou na transferência de estados emocionais pela manipulação de posts positivos (animados!) ou negativos (chateado!) (O'NEIL, 2020); na produção de conteúdo falso de forma automatizada com a intenção de desinformar (KNIGHT, 2019); nas novas e obscuras formas de controle, gerenciamento e exploração do trabalho que beneficiam ou punem os trabalhadores com base em ranqueamentos e participação vigilante dos consumidores, onde os riscos, custos e responsabilidades são transferidos para o trabalhador, ao mesmo tempo em que são eliminados proteções, direitos e garantias (ABÍLIO, 2020).

Os problemas éticos atrelados às tecnologias de IA se acumulam e se agravam, não sendo incomum localizar a fonte dos mesmos em grandes corporações, a exemplo da G-MAFIA (Google, Microsoft, Amazon, Facebook, IBM e Apple) que detém grande poder de processamento e controle de gigantescos *datacenters* que são *inputs* para a construção de complexos sistemas de IA.

Os sistemas de IA têm mudado a existência e as experiências humanas, massificando-as ou atomizando-as à medida em que o indivíduo se torna parte de uma ou mais amostras, sendo a sua relação com os outros definida por simulações de identidades e padrões estimativos que antecipam potencialidades, sejam de consumo, econômicas, comportamentais entre outras (MACHADO, 2018, p.53) ou quando o indivíduo tem seu comportamento conhecido com base em seus comportamentos anteriores (MACHADO, 2018, p.52). Ambas as situações expressam modos de atuação da modulação algorítmica que utilizam sofisticadas técnicas de IA. O debate sobre IA ética se desdobra para além das questões técnicas, como os modelos matemáticos e os volumosos dados utilizados para o treinamento de sistemas de IA, ao lançar o foco para os ambientes onde esses artefatos são construídos e para as pessoas que participam de sua construção (INGRAM, 2021), estendendo a atenção para os seus usos.



O contexto de desenvolvimento de sistemas de IA é inerentemente heterogêneo e plural. Envolve uma diversidade de atores com diferentes interesses, diferentes atuações e poder de decisão. A constituição de espaços de reflexão, a pluralidade e representatividade de seus participantes bem como os modos e meios de participação e condução das discussões são elementos que afetam as condições de possibilidade de os atores envolvidos alcançarem uma possível compreensão compartilhada e justificada de que princípios e valores estão em jogo, quais serão considerados, como se chegou a uma interpretação consensuada e como estes códigos morais serão implementados, ou seja, como princípios e valores morais serão “traduzidos” em códigos computacionais. Esta dinâmica aplicada nos ambientes de construção de IA, que supostamente busca e conduz para uma IA ética, embora pareça simples e suficiente, não o é, pois a decisão do que é uma IA ética pode refletir a perspectiva de poucos indivíduos ou grupo, como desenvolvedores, comunidades de cientistas ou pesquisadores em laboratórios ou corporações, como alertou Arbix (2020, 409). Por isso, é fundamental mobilizar conhecimentos e práticas plurais para se estabelecer diretrizes para IA ética centrada no ser humano e alinhadas com valores culturais e sociais, tendo-se em conta diferenciações na estrutura social, a diversidade cultural e a consideração de que valores e princípios morais variam no tempo e no espaço.

Em face aos impactos em vários domínios e a preocupação com os riscos e consequências negativas de sistemas de IA, como viés ou discriminação algorítmica, principalmente a partir dos últimos cinco anos, organismos nacionais, internacionais, empresas privadas, legisladores, setor público e organizações da sociedade civil têm debatido e elaborado diretrizes para orientar o desenvolvimento e uso ético daqueles, conforme apresentado nos estudos de Floridi et al. (2018), Jobin, Ienca e Vayena (2019), Ryan e Stahl (2021) e Rudschies, Schneider e Simon (2021).

O debate em torno da IA ética envolve uma diversidade de princípios e valores, os quais têm alto grau de abstração, deixando, portanto, espaço para amplas e distintas interpretações. Jobin, Ienca e Vayena (2019, p.7) evidenciaram uma convergência emergente em torno dos princípios como transparência, justiça e equidade, não maleficência, responsabilidade e privacidade, como também evidenciaram divergências semânticas e conceituais significativas tanto na forma como esses princípios éticos são interpretados quanto nas recomendações específicas ou áreas de preocupação derivadas de cada um.



No entanto, para que essas orientações/recomendações destinadas aos atores envolvidos em qualquer estágio do ciclo de vida de sistemas de IA se convertam em ações concretas efetivamente realizadas de forma ética, é necessário mais do que o conhecimento de uma lista de princípios e valores a serem considerados. Requer uma disposição do sujeito para um exercício de reflexão que leve em conta os princípios e valores da “lista” em seu possível confronto ou alinhamento com outras concepções atribuídas para esses mesmos valores e princípios. Esse exercício, por sua vez, requer uma disposição do sujeito para buscar e aprimorar continuamente a reflexão sistemática e informada com base no pensamento filosófico, particularmente, o da ética aplicada no domínio informacional, tal como a ética da informação.

O presente trabalho tem como objetivo investigar em documentos que tratam de ética em IA, a existência de remissivas ao estímulo do desenvolvimento de competências, para além das competências técnica e instrumental, que estimulem e possibilitem um contínuo pensamento crítico-reflexivo fundamentado na teoria e na prática de forma a orientar ações responsáveis, justificáveis e, sobretudo, transformadoras.

2 METODOLOGIA

Trata-se de uma pesquisa bibliográfica e documental, de natureza qualitativa. A pesquisa bibliográfica orbitou em torno de temas/assuntos como inteligência artificial, ética da informação, competência crítica. A pesquisa documental teve como foco documentos normativos voltados para IA ética. Sem a pretensão de constituir uma amostra densa, ao menos nesse momento, foram selecionados documentos primários, de domínio público, produzidos por atores representativos no debate de ética no contexto da IA.

A fim de identificar a existência (ou não) de remissivas as questões de interesse da investigação apontadas na seção anterior, bem como os contextos em que aparecem, foram analisados o *corpus* de cada documento selecionado. Utilizou-se o software Voyant Tools, que é online, gratuito e de código aberto, como ferramenta auxiliar na análise de dados, em face das várias funcionalidades que dispõe, a exemplo, daquela que exibe cada ocorrência de uma palavra-chave com um trecho circundante (contexto), sendo tal funcionalidade útil para analisar como os termos são utilizados em diferentes contextos. Os termos de interesse utilizados nos critérios de busca no VoyanTools foram: *literacy*/literacia, *education*/educação, *training*/treinamento, *critical thinking*/pensamento crítico. A amostra considerada é pequena



se considerado o quantitativo de 167 *guidelines* apresentado pelo *AI Ethics Guidelines Global Inventory*². No entanto, os documentos da amostra são referenciais para muitos outros. Foram selecionados documentos de natureza deliberativa de abrangência nacional, internacional, do setor privado, de *multistakeholders* e de natureza legal, como o Projeto de Lei nº 21/20, aprovado pela Câmara de Deputados e em tramitação no Senado Federal. Os documentos analisados, bem como os excertos referentes às remissivas envolvendo os termos do critério de busca, encontram-se no Quadro 1.

3 ÉTICA DA INFORMAÇÃO E ÉTICA NA IA

Apreendendo de forma geral o entendimento de Sánchez Vásquez (2020), pode-se dizer que ética diz respeito à prática, à ação. Diz respeito a como se deve agir ou comportar-se diante de determinados problemas ou situações concretas que, em geral, podem afetar direta ou indiretamente uma ou muitas pessoas. A ética, portanto, diz respeito à reflexão crítica sobre a ação moral. Busca refletir sobre as normas, princípios e valores que regulam o comportamento humano em certa cultura e época, sobre seus fundamentos, se se sustentam ou não, porque merecem ou não serem defendidos, rechaçados ou censurados. A ética diz respeito ao comportamento de um agente dotado de consciência, de certo grau de liberdade de escolha e racionalidade, que são condições fundamentais para o agente justificar e ser responsável pelos seus atos, bem como ser responsabilizado pelos os mesmos. No entanto, a aplicação prática desse entendimento basilar de ética não é trivial, sobretudo, quando colocado no âmbito de sistemas de IA onde há expectativas com relação ao “comportamento ético” desses artefatos.

A ética da informação (EI), “como ética aplicada, investiga as questões éticas que surgem do ciclo de vida da informação, incluindo a geração, coleta, organização, recuperação, distribuição e uso da informação” (BRITZ, 2013, p.3). Tem como uma de suas tarefas precípuas, refletir sobre as mudanças na relação entre informação e ação moral no contexto das tecnologias digitais. Para Capurro (2014, p.9), a EI pode ser descritiva ou emancipatória.

Como reflexão descritiva, explora as estruturas de poder que determinam atitudes e costumes informacionais e comunicacionais em diferentes culturas e épocas. A ética emancipatória que inclui uma reflexão crítica normativa, avalia o desenvolvimento de valores no campo informacional e comunicacional, a criação de novas estruturas de poder, os mitos

² Disponível em: <https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/>. Acesso em: 1 abr. 2022.



informacionais e comunicacionais, as contradições ocultas e as intencionalidades das teorias e práticas informacionais e comunicações, bem como o desenvolvimento de conflitos éticos neste domínio (CAPURRO, 2014, p.9).

Para Capurro (2014), pensar eticamente sobre como as tecnologias digitais interativas moldam e afetam a nossa forma de ser-no-mundo com referências aos códigos morais, econômicos, políticos e técnicos que nos regem até então, implica, em última instância, pensar em quem somos nós mesmos nesse mundo digitalmente globalizado. O nosso ser-no-mundo, no entanto, é um estar-no-mundo provisório, na medida em que participamos de um jogo social que envolve justamente a nossa liberdade de revelar ou esconder quem somos em diferentes contextos e com diferentes objetivos, que se torna possível porque somos representados por sequências de bits. Essa liberdade de esconder e revelar quem somos como indivíduos, grupos ou como sociedade, própria do jogo social que hibridiza um “quem” e um “quê”, cada vez mais se fragiliza à medida que essa liberdade é de certa forma apropriada por grandes empresas digitais e seus modelos extrativistas de dados (MOROZOV, 2018, p.166), que ao capturarem nossos dados fornecidos, em grande medida, voluntariamente como também de forma sub-reptícia, sobredeterminam a dinâmica do binômio revelar-esconder, quando por exemplo, sistemas de IA determinam quem terá visibilidade ou não nas redes sociais, ou o que se torna visível ou não para quem, de acordo com as suas prerrogativas opacas de ordenamento e seus critérios difusos (MOROZOV, 2018, p.48).

Por outro lado, a mesma liberdade basilar de revelar ou esconder quem somos mediados pelas tecnologias digitais, nos coloca no radar dos modelos de vigilância algorítmica, cuja ênfase, agora recai preferencialmente no indivíduo e na relação entre indivíduos do que na vigilância centralizada. De uma perspectiva da ética da informação, o controle e a vigilância de sofisticados sistemas de IA não só violam a esfera privada como também põem em perigo a liberdade de pensamento e opinião (CAPURRO, 2014).

Para Leslie (2019), o campo da ética da IA surgiu em resposta à variedade de consequências negativas que os sistemas de IA podem causar. Define ética da IA como “um conjunto de valores, princípios e técnicas que empregam padrões amplamente aceitos do que é certo ou errado para guiar a conduta no desenvolvimento e uso de tecnologias de IA” (LESLIE, 2019, p.3). Para Bester e Fisher (2021, p.10), o debate sobre ética em IA não é um novo debate, mas uma continuação aprofundada das discussões de tópicos já abordados na



ética da informação (EI), agora reformulados em razão dos impactos das tecnologias de IA sobre as sociedades e o meio ambiente possibilitados pelo grande poder de computação e os volumosos *datasets* e algoritmos que, em seu conjunto, constituem complexos sistemas sociotécnicos.

No diálogo entre EI e ética da IA, alguns dos tópicos que requerem escrutínio na perspectiva ética, são os efeitos potenciais da IA em relação aos direitos humanos fundamentais e dignidade das pessoas; questões relacionadas à privacidade; aos riscos de a IA perpetuar, reforçar ou gerar preconceitos e discriminações pela incorporação de vieses voluntários ou não, tanto nos algoritmos como na coleção de dados, incluindo os impactos ambientais da IA à medida em que demandam significativo aumento do uso de recursos naturais como energia e água, além do descarte de resíduos; os impactos nos sistemas legais; a falta de ou a limitada conscientização e compreensão do cidadão sobre as consequências de suas escolhas e uso de sistemas de IA, levando-os à decisões mal informadas e danos subsequentes, entre tantas outras.

As diretrizes (*guidelines*) para a IA ética, enquanto instrumentos normativos de caráter deliberativo são formulados com a expectativa de serem adotados por atores envolvidos no desenvolvimento e uso de sistemas de IA e deveriam, em princípio, adotar uma abordagem global, pluralista, multidisciplinar, multicultural e multissetorial. Por outro lado, os atores destinatários dessas diretrizes, em especial, aqueles que participam da construção de sistemas de IA, muito provavelmente, possuem diferentes conhecimentos, experiências, capacidades sociocognitivas, preferências e idiosincrasias, como também diferentes disposições para o pensamento crítico-reflexivo, que os permitam considerar, entender, ponderar e julgar quais valores e princípios estão em jogo, quais serão priorizados, que razões justificam a sua incorporação (ou não) e, sobretudo, quais os possíveis efeitos ou consequências das escolhas e tomadas de decisão que ao fim e ao cabo sobredeterminam o “comportamento moral” dos sistemas de IA. Esta última questão é a que direciona a discussão para o desenvolvimento de uma competência crítica no contexto de construção de sistemas de IA, que notadamente é intensamente informacional.

O contexto da IA é caracterizado por três componentes: grande volume de dados, poder computacional e poder de análise a partir do processamento de algoritmos para reordenar, reinterpretar e representar os dados disponíveis como informação (BESTER;



FISHER, 2021, p.7). Dados e informação são conceitos polissêmicos, com rica bibliografia no campo da Ciência da Informação a respeito (RIBEIRO; SANTOS, 2020). Em contextos de *big data*, dados possuem grande, em especial, quando tratados, contextualizados, categorizados, calculados, acondicionados, corrigidos e resumidos, sendo esses alguns dos processos de transformação de dados em informação. Dados são *inputs* necessários para os sistemas de IA. O pensamento crítico-reflexivo, próprio da ética, sobre como são adquiridos, agregados, analisados e utilizados por complexas técnicas de IA, a exemplo de aprendizado de máquina, sobre os impactos reais ou potenciais daqueles processos é fundamental para a condução de práticas responsáveis e compromissadas, expressando, por sua vez, um tipo de competência, uma competência crítica necessária e esperada para aqueles envolvidos na construção de sistemas de IA.

Um tipo de competência que juntamente com a competência instrumental e técnica, os possibilite re(conhecer) e evitar instancias de violência social, como por exemplo, o racismo estrutural da sociedade, agora, manifestado em contornos datificados e algoritmizados. Como Noble (2021, p.8) pontuou, “parte do desafio de compreender a opressão algorítmica é perceber que as formulações matemáticas que guiam as decisões automatizadas são feitas por humanos”. E, justamente por isso, não exclusivamente, mas em particular, os envolvidos na construção de IA necessitam dispor de um equipamento teórico-prático-reflexivo que lhes permita o desenvolvimento e aprimoramento do conhecimento autorreflexivo sobre suas próprias práticas, alinhando-o e não se esgotando nos conhecimentos disciplinares necessários para a sua atuação profissional.

Bester e Fisher (2021, p.10), defendem que os formadores de desenvolvedores de algoritmos devem receber intensivo treinamento em EI devido às vastas implicações desses artefatos em diversos domínios. Também pontuam a importância de serem promovidas pesquisas sobre assuntos relacionados à diversidade cultural na formação de aprendizes interessados em codificação de algoritmos.

4 ANÁLISE DOS DOCUMENTOS NORMATIVOS VOLTADOS PARA IA ÉTICA

Considerando-se o quão distribuída e heterogênea é a rede de atores que participam da construção de sistemas de IA, de que todos os envolvidos têm responsabilidades com relação às potenciais consequências negativas desses artefatos, dentro do escopo e abrangência de suas atividades ou papéis, e a importância de que a palavra ética não seja



utilizada nos contextos de construção de IA apenas como um palavra curinga e vazia de compreensão substantiva, buscou-se identificar nos documentos selecionados possíveis remissivas à importância de um contínuo pensamento crítico, de formação e conhecimento em ética, os quais, em seu conjunto, contribuiria para o desenvolvimento de uma competência crítica fundamentalmente necessária para uma IA ética e confiável. Os documentos normativos selecionados, a natureza do ator(es) envolvidos na elaboração dos mesmos, bem como os excertos destacados, encontram-se no Quadro 1.

Quadro 1 – Excertos de remissivas relacionadas à competência crítica no domínio da IA

DOCUMENTO (ANO) NATUREZA DO ATOR	TERMO(S)	EXCERTO DA(S) REMISSIVA(S)
UNESCO/COMEST (2019) ³ Internacional	<i>Literacy</i>	[...] <i>demanding not only 'data literacy' that allows students to read, analyse and efficiently manage this information but also 'AI literacy' to enable critical reflection on how intelligent computer systems have been involved in the recognition of information needs, selection, interpretation, storage and representation of data.</i> <i>AI requires that education fosters AI literacy, critical thinking, resilience on the labour market, and educating ethics to engineers.</i>
UNESCO (2021) ⁴ Internacional	<i>Literacy/ critical thinking</i>	[...] <i>should promote the acquisition of "prerequisite skills" for AI education, such as basic literacy, numeracy, coding and digital skills, and media and information literacy, as well as critical and creative thinking, teamwork, communication, socio-emotional and AI ethics skills.</i>
UNIVERSITY OF MONTREAL (2018) ⁵ Multistakeholders	<i>Literacy/ critical thinking</i>	<i>It is crucial to empower citizens regarding digital technologies by ensuring access to the relevant forms of knowledge, promoting the learning of fundamental skills (digital and media literacy), and fostering the development of critical thinking.</i>
IEEE (2019) Multistakeholders ⁶	<i>literacy</i>	<i>Therefore, technology awareness and understanding of social and ethical issues of A/IS are new literacy skills society must embrace if A/IS applications are to be accepted and trusted as an integral part of modern living.</i>
MICROSOFT (2018) ⁷ Setor privado	<i>ethical skill</i>	[...], <i>the social sciences and humanities will become even more important.</i> [...] <i>courses can teach critical, philosophical and ethics-based skills that will be instrumental in the development and management of AI solutions.</i>
EUROPEAN COMMISSION (2019) ⁸ Intergovernamental	<i>literacy</i>	Basic AI literacy should be fostered across society . A prerequisite for educating the public is to ensure the proper skills and training of ethicists in this space .

³ Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000367823>

⁴ Disponível em: [Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library](#)

⁵ Disponível em: <https://www.montrealdeclaration-responsibleai.com/the-declaration>

⁶ Disponível em: https://standards-qa21.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

⁷ Disponível em: https://news.microsoft.com/wp-content/uploads/2018/02/The-Future-Computed_2.8.18.pdf

⁸ Disponível em: [Ethics guidelines for trustworthy AI - Publications Office of the EU \(europa.eu\)](#)



BRASIL (2020) ⁹ Legislativo (Câmara dos Deputados)	capacitação	[...] V – estímulo à capacitação e à preparação das pessoas para a reestruturação do mercado de trabalho; VI – Estímulo a práticas pedagógicas inovadoras, com visão multidisciplinar, e ênfase da importância de ressignificação dos processos de formação de professores para lidar com os desafios decorrentes da inserção da inteligência artificial como ferramenta pedagógica em sala de aula.
BRASIL (2021) ¹⁰ Governo	Educação/ competência	[...] a Base Nacional Comum Curricular (2017) inclui, entre as competências gerais da educação básica , a de “ compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas sociais. Em relação ao mercado de trabalho, a importância da literacia digital é ainda mais latente. Muitos empregadores, [...] exigem o domínio de habilidades digitais, [...]. Ainda se inclui neste cenário a capacidade de usar a informação de forma ética no contexto social/virtual em que vivemos, respeitando indivíduos e grupos sociais. Os cursos de idiomas, arte, história, economia, ética, filosofia, psicologia e desenvolvimento humano podem ensinar habilidades críticas, filosóficas e éticas que serão fundamentais para o desenvolvimento e gerenciamento de soluções de IA. A promoção de literacia digital passa a ser fator chave para o desenvolvimento de uma nova massa de profissionais preparado para os desafios do próximo século.

Fonte: Elaborado pela autora (2022)

Em UNESCO/COMEST (2019), os termos “*data literacy*” e “*AI literacy*” são colocados no âmbito da educação formal com destaque para a formação dos profissionais das engenharias, apontando, de um lado, as limitações do primeiro, que se aproxima da noção clássica de *information literacy*, como a apresentada pela *International Federation of Library Associations and Institutions*, salientando uma competência instrumental e gerencial em dados e informação e, de outro lado, ressaltando a importância da *AI literacy* para uma reflexão crítica sobre os sistemas de IA. Nessa direção, interessante notar a atenção dada para a reflexão sobre o envolvimento desses sistemas no reconhecimento das necessidades de informação, seleção, interpretação, armazenamento e representação de dados, remetendo a proposta de uma *AI literacy* no bojo das discussões e ações de conscientização e combate aos usos não éticos dos dados e informação, como por exemplo, na produção e disseminação de desinformação. A Declaração de Montreal para IA é fortemente principiológica, mas também oferece uma gama de linhas políticas de ação. Ressalta a importância de uma *digital and*

⁹ Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/151547>.

¹⁰ Disponível em: [ebia-digramacao_4-979_2021.pdf \(www.gov.br\)](https://www.gov.br/ebia-digramacao/4-979_2021.pdf)



media literacy referindo-se para o desenvolvimento da capacidade de acessar, gerenciar, entender, integrar, comunicar, avaliar e criar informações de forma segura e apropriada por meio de ferramentas digitais e de tecnologias em redes, de forma a permitir a sua participação na vida econômica e social. O documento IEEE (2019), associa *news literacy skills* às habilidades fundamentais para a conscientização e entendimento das questões sociais e éticas dos sistemas inteligentes como sendo fundamentais com reflexos na adoção e a confiabilidade desses sistemas. O documento da Comissão Europeia, aponta a promoção da *AI literacy* para toda a sociedade, bem como o papel dos eticistas na educação. Em BRASIL (2021), recorrendo ao trecho presente em MICROSOFT (2018), é ressaltado a importância dos conhecimentos próprios do campo das Humanidade para a formação de um pensamento filosófico e ético com foco, no entanto, na preparação de uma massa de profissionais para o desenvolvimento de soluções de IA. Já em BRASIL (2020), a ênfase recai na atualização de práticas pedagógicas e na capacitação para o mercado de trabalho.

A elaboração de diretrizes é uma importante expressão das (re)ações por parte de diversos atores frente às preocupações e aos desafios para uma IA ética e confiável. Também é possível perceber nos documentos analisados, a ausência de referências à uma gama de questões cruciais discutidas na ética da informação, que não só considera as questões éticas que surgem no ciclo de vida da informação, mas também explora os sistemas morais no campo informacional, tendo-se em conta os aspectos históricos em meio as diferentes tradições culturais, as estruturas de poder, as contradições e intencionalidades veladas que, em grande medida, sobredeterminam os comportamentos informacionais, em especial, mas não exclusivamente, daqueles envolvidos no desenvolvimento de sistemas de IA. Entende-se, portanto que as questões éticas no domínio da IA seria melhor abordada com maior diálogo com a EI.

4 CONSIDERAÇÕES FINAIS

A discussão sobre IA ética é sobretudo a discussão sobre responsabilidades. A construção de sistemas de IA envolve uma extensa rede de atores envolvidos /interessados com diferenças e percepções importantes sobre a dimensão ética da IA. A construção desses artefatos envolve, portanto, uma extensa rede de responsabilidade distribuída.

Em um tempo histórico onde dataísmo, colonialismo de dados e superinteligência artificial estão presentes nos discursos e preocupações éticas (ainda que haja preocupações



em outras dimensões), indubitavelmente são os dados os objetos de interesse (e de cobiça) que quaisquer diagnósticos dessas realidades ou de futuros desejáveis (ou não) irão apontar. Recorrendo à Kitchin (2017), Ingram (2021) apresenta a ideia de uma “*assemblage* de dados” que inclui “ideias, técnicas, tecnologias, sistemas, pessoas e contextos” que “evoluem e mudam ao longo do tempo, contribuindo para a reflexão a questão da responsabilidade distribuída própria da construção de sistemas de IA, uma vez que remete à indagações em torno da responsabilidade em duas direções distintas e entrelaçadas: a que considera a perspectiva individual e a que considera a perspectiva da governança, se adotado o pensamento de Simon (2014), quando discorre sobre responsabilidade epistêmica distribuída.

Entendendo que os envolvidos na construção de IA ética e confiável necessitam dispor de um equipamento teórico-prático e de um conhecimento autorreflexivo sobre suas próprias práticas, alinhando-os aos conhecimentos disciplinares necessários para a sua atuação profissional, sem se restringir aos mesmos, aponta-se a importância do desenvolvimento de uma competência crítica alicerçada e revigorada pelo indagar e refletir própria da ética, enquanto ramo da filosofia prática e, em especial, da ética da informação. A partir da análise do *corpus* documental analisados, é possível perceber que o debate sobre IA ética será muito mais rico e frutífero em diálogo com a EI.

REFERÊNCIAS

- ABÍLIO, Ludmila Costhek. Plataformas digitais e uberização: Globalização de um Sul administrado?. **Contracampo**, Niterói, v. 39, n. 1, p. 12-26, abr./jul. 2020. Disponível em: <https://periodicos.uff.br/contracampo/article/view/38579/pdf>. Acesso em: 04 jun. 2022.
- ARBIX, Glauco. A transparência no centro da construção de uma IA ética. **Novos estud. - CEBRAP**, v.39, n.3, 2020. Disponível em: <https://www.scielo.br/j/nec/a/pD9k5gtHpXwsgFcsMC5gbJg/?lang=pt>. Acesso em: 11 jun. 2022.
- BESTER, C.; FISCHER, R. The Essential Relationship between Information Ethics and Artificial Intelligence. **The International Review of Information Ethics**, Edmonton, Canada, v. 29, 2021. Disponível em: <https://informationethics.ca/index.php/irie/article/view/428>. Acesso em: 11 jun. 2022.
- BRITZ; Johannes. Understanding Information Ethics. In: OCHOLLA, Dennis; BRITZ, Johannes; CAPURRO, Rafael; BESTER, Coetzee (Eds.). **Information ethics in Africa: cross-cutting Themes**. Pretoria: Groep 7 Drukkers, 2013.
- BEIGUELMAN, Giselle. **Políticas da imagem: vigilância e resistência na dadosfera**. São Paulo: Ubu Editora, 2021. 224p.



CAPURRO, Rafael. Informação e ação moral no contexto das tecnologias de comunicação. In: GONZALEZ, Maria Eunice Quilici; Broens, Mariana Claudia; Martins, Clélia Ap. (Org.).

Informação, conhecimento e ação ética. Marília, São Paulo: Cultura Acadêmica, 2012. 222p.

CAPURRO, Rafael. La libertad en la era digital. Informatio. **Revista del Instituto de Información de La Facultad de Información Y Comunicación**, v. 19, n.1, p. 5-23. Disponível em: <https://informatio.fic.edu.uy/index.php/informatio/article/view/150>. Acesso em: 08 jun. 2022.

EUROPEAN COMMISSION. **Ethics guidelines for trustworthy AI**, 2019. Disponível em: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Acesso em: 08 jun. 2022.

FJELD, Jessica; ACHTEN, Nele; HILLIGOSS, Hannah; NAGY, Adam; SRIKUMAR, Madhulika. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. **Berkman Klein Center Research Publication**, n. 2020-1, 2020 Disponível em: <https://ssrn.com/abstract=3518482>. Acesso em 04 jun. 2022.

FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica; CHATILA, Raja; CHAZERAND, Patrice; DIGNUM, Virginia; LUETGE, Christoph. AI4People - Uma estrutura ética para uma boa sociedade de IA: oportunidades, riscos, princípios e recomendações. **Minds and Machines**, n.28, 2018, p.689–707. Disponível em: <https://link.springer.com/content/pdf/10.1007/s11023-018-9482-5.pdf>. Acesso em: 06 jun. 2022.

INGRAM, Katrina. Constructing AI: examining how ai is shaped by data, models and people. **The International Review of Information Ethics**, Edmonton, Canada, v. 29, 2021. Disponível em: <https://informationethics.ca/index.php/irrie/article/view/415>. Acesso em: 7 jun. 2022.

JOBIM, Anna; IENCA, Marcello; VAYENA, Effy. Artificial Intelligence: the global landscape of ethics guidelines. **Nature Machine Intelligence**, v.1, 2019. Disponível em: https://www.researchgate.net/publication/334082218_Artificial_Intelligence_the_global_landscape_of_ethics_guidelines. Acesso em: 4 jun. 2022.

KNIGHT,Will. An AI that writes convincing prose risks mass-producing fake news. **MIT Technology Review**. Disponível em: <https://www.technologyreview.com/2019/02/14/137426/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish/>. Acesso em: 05 jun. 2022.

KITCHIN, Rob. **The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences**. UK: Sage Publications Ltd, 2017.

LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren; ANGWIN, Julia. How we analyzed the COMPAS recidivism algorithm. ProPublica, 2016. Disponível em:

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 04 jun. 2022.

LESLIE, David. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. Disponível em: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf. Acesso em: 5 jun. 2022.



MACHADO, Débora. A modulação de comportamento nas plataformas de mídias sociais. In: SOUZA, Joyce; AVELINO, Rodolfo; SILVEIRA, Sérgio Amadeu da (Org.). **A Sociedade de Controle**: manipulação e modulação nas redes sociais. São Paulo: Editora Hedra Ltda, 2018.

MOROZOV, Evgeny. **Big Tech**: a ascensão dos dados e a morte da política. São Paulo: Ubu Editora. 2018. 192 p.

NOBLE, Safiya Umoja. **Algoritmos de opressão**: como o Google fomenta e lucra com o racismo. Tradução: Felipe Damorim. Santo André, SP: Editora Rua do Sabão, 2021. 390 p.

O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução: Rafael Abraham. 1.ed. Santo André, SP: Editora Rua do Sabão, 2020. 339 p.

PARISER, Eli. **O filtro invisível**: O que a internet está escondendo de você. Tradução: Diego Alfaro. 1. ed. Rio de Janeiro, RJ: Editora Zahar, 2012. 291 p.

RIBEIRO, Anna Carolina M. L.; SANTOS, Carlos Denner dos. Isso não é uma pirâmide: revisando o modelo clássico de dado, informação, conhecimento e sabedoria. *Ciência da Informação*, [S. l.], v. 49, n. 2, 2020. Disponível em: <http://revista.ibict.br/ciinf/article/view/5066>. Acesso em: 5 jun. 2022.

RUDSCHIES, Catharina; SCHNEIDER, Ingrid; SIMON, Judith. Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities. **The International Review of Information Ethics**, Edmonton, Canada, v. 29, 2021. DOI: 10.29173/irrie419. Disponível em: <https://informationethics.ca/index.php/irrie/article/view/419>. Acesso em: 5 jun. 2022.

RYAN, Mark; STAHL, Bernd Carsten. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. **Journal of Information, Communication & Ethics in Society**, v. 19, n.1, 2021. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JICES-12-2019-0138/full/pdf?title=artificial-intelligence-ethics-guidelines-for-developers-and-users-clarifying-their-content-and-normative-implications>.

SIMON, Judith. Distributed Epistemic Responsibility in a Hyperconnected Era. In: FLORIDI, Luciano (Eds). **The Onlife Manifesto**. Springer, Cham. p.145-149, 2014. Disponível em: <https://link.springer.com/content/pdf/10.1007/978-3-319-04093-6.pdf>. Acesso em: 12 jun. 2022.