

Inteligência artificial e tomada de decisão ética no setor público

Artificial intelligence and ethical decision- making in the public sector

Sandro Trescastro Bergue ¹

¹Escola Superior de Gestão e Controle do Tribunal de Contas do Estado do RS,
Porto Alegre, Rio Grande do Sul, Brasil.

Resumo

Objetivo: Investigar a ética como exercício do pensamento crítico-reflexivo no uso da inteligência artificial (IA), em particular nas tomadas de decisão de gestores do serviço público.

Método/abordagem: Ensaio teórico que examina os conceitos de ética e de conduta expressos na literatura especializada que relaciona inteligência artificial e ética na administração pública.

Contribuições teóricas/práticas/sociais: O trabalho demonstra a necessidade de repensar o conceito dominante de ética, dando ênfase às políticas de educação; a eficácia limitada da abordagem da codificação de condutas e princípios balizadores, seja aos profissionais desenvolvedores destas tecnologias, seja nos algoritmos.

Originalidade/relevância: Residem na proposição do entendimento da ética para além de um conjunto de princípios e valores a orientar a conduta dos agentes, passando a destacar o conceito como o exercício de um juízo crítico-reflexivo, radical e situado incidente sobre os fundamentos conformadores dos parâmetros morais compartilhados envolvidos nos processos de tomada de decisão com destaque para o complexo campo da administração pública.

Palavras-chave: Ética. Pensamento crítico-reflexivo. Códigos de conduta. Tomada de decisão. Inteligência artificial.

Abstract

Purpose: Investigate ethics as an exercise in critical-reflective thinking in the use of artificial intelligence (AI), particularly in decision-making by public service managers..

Design/methodology/approach: Theoretical essay that examines the concepts of ethics and conduct expressed in specialized literature that relates artificial intelligence and ethics in public administration.

Research, Practical & Social implications: The work demonstrates the need to rethink the dominant concept of ethics, placing emphasis on education policies; the limited effectiveness of the approach to coding conduct and guiding principles, whether for professionals developing these technologies or in algorithms.

Originality/value: They reside in the understanding of ethics in addition to a set of principles and values to guide the conduct of agents, highlighting the concept as the exercise of a critical-reflective, situated and radical judgment incident on the foundations of the shared moral parameters involved in the processes decision-making process with emphasis on the complex field of public administration.

Keywords: Ethics. Critical-reflective thinking. Codes of conduct. Decision making. Artificial intelligence.

Introdução

Este ensaio tem como principal objetivo examinar o lugar da ética como exercício do pensamento crítico-reflexivo, situado e de extensão radical necessário aos agentes tomadores de decisão no ainda emergente e fluido terreno de uso da inteligência artificial (IA), em particular no espaço do serviço público. E o faz a partir do que a filosofia tem como objeto, a saber, segundo Deleuze e Guattari (1992), a produção ou reelaboração de conceitos em resposta a problematizações contextualizadas. Eis um desafio para o qual um texto desta natureza se presta, pois não somente pela amplitude e alcance potencial do campo de incidência, mas em termos de necessário assentamento conceitual, os contornos da IA, tanto quanto os da ética – que tem a deliberação justa como um de seus escopos essenciais –, podem ser considerados ainda em franca construção (Garcia, 2020; Wang, 2019).

O impulso inicial de produção deste artigo de natureza ensaística reside na percepção de que o tema da ética vem sendo abordado no campo da inteligência artificial (IA) nos mesmos estreitos moldes (Munn, 2023; Waelen, 2022; Gal, Hansen & Lee, 2022) segundo os quais tem sido tratado no campo da administração pública de modo geral, a saber, reduzido à instrumentalidade fundada em códigos de conduta. A propósito, investigar a ética no contexto da IA também oferece uma oportunidade de repensar esse mesmo

conceito no seu convencional espaço de produção e aplicação na administração pública, a saber, inserto nos programas de integridade que a tomam como um de seus elementos (Brasil, 2023). Isso porque remete a um necessário reexame acerca da compreensão que se tem sobre ética, códigos de conduta e subsequentes processamentos de natureza disciplinar associados ao tema (Meyer-Sahling & Mikkelsen, 2022; Bergue, 2022b; Downe, Cowell & Morgan, 2016). E a ética pressupõe tudo isso, mas os transcende.

Assume-se, assim, que a capacidade de elaboração de juízos éticos tem se concentrado mais no território da sanção e menos no da ação refletida e justificada. Ora, o risco da manutenção desta trajetória sabidamente limitada ganha destaque ao se reconhecer tanto o impacto desta tecnologia (IA), quanto a amplitude e repercussões potenciais do seu emprego no serviço público. Daí a relevância de reexaminar o entendimento sobre ética e as possibilidades alternativas de sua aplicação efetiva e transformadora no espaço de produção e de utilização de tecnologias de inteligência artificial de suporte à tomada de decisão justificada no contexto da administração pública brasileira.

Observa-se que tal como adotado na literatura, o termo IA ética (Munn, 2023; Bankins & Formosa, 2023; Waelen, 2022; Hagendorff, 2022a) poderia sugerir a possibilidade de tecnologias de IA tomarem decisões

éticas, o que seria algo, no mínimo, duvidoso. Então, por que a expressão vigora? Propõe-se que o deslinde está subordinado ao necessário esforço de pensar em maior profundidade teórica sobre que conceito ética tem iluminado o debate. O ponto essencial aqui reside no reconhecimento de que os juízos éticos constituem uma capacidade fundamentalmente humana, ainda distantes do alcance dos algoritmos de IA, a começar pelo fato de que, observado o rigor conceitual, juízos éticos exigem o sopesamento consciente e a reflexão sobre valores conforme as circunstâncias e não somente a observância de padrões prescritos de conduta.

Nessa linha, para melhor posicionar o tema ainda em caráter introdutório, é importante reconhecer dois entendimentos dominantes e entrelaçados sobre ética: a ética como um conjunto de princípios e valores constituídos com o intuito de orientar a ação das pessoas em uma determinada coletividade; e a ética como uma capacidade crítico-reflexiva situada, de extensão radical incidente sobre os parâmetros valorativos e normativos morais socialmente instaurados. No primeiro caso o senso de ética se aproxima do conceito de moral e da concepção e proposições formais comumente disseminadas nas políticas institucionais, ao passo que no segundo, a ética mais se alinha a uma atitude filosófica incidente sobre os fundamentos daqueles parâmetros morais de conduta, tal como definido

no campo da filosofia (Cortina & Martinez, 2005). A deliberação ética constitui, portanto, ação fundamentada em uma moralidade examinada e devidamente refletida. Em face do objetivo deste texto, o senso de ética é tomado com maior ênfase nos termos do segundo entendimento.

Assim, esta investigação teórica é também impulsionada pela compreensão sobre o percurso e lugar convencional de aplicação do conceito de ética no contexto da tomada de decisão baseada também em produtos da IA, a saber, essencialmente assentado na definição e codificação de diretrizes de conduta moral aplicadas aos desenvolvedores das soluções de tecnologia. Pretende-se, então, pôr em relevo a possibilidade do deslocamento da incidência da ética para além do momento em que se empreendem os esforços de produção das IA's (princípios, diretrizes e códigos de conduta de desenvolvimento), e estendê-la como atitude crítico-reflexiva e situada para o momento de seu uso subsequente no processo de tomada de decisão.

Isso porque reduzir a ética a parâmetros morais de conduta positivados parece análogo, para fins estritamente ilustrativos, a circunscrever o amplo e complexo senso de justiça ao estreito conceito de legalidade. Ou seja, códigos de conduta inspiram a ação ética refletida e situada; ao passo que a legalidade estrita – expressão de diminuta amplitude do fenômeno jurídico – indica a

conformidade da ação do sujeito ao texto normativo aderente ao contexto, ainda que em ambos os casos incida o exercício de interpretação, mas não esgotando o mais denso significado de justiça. Para iluminar ainda mais essa distinção no plano conceitual e subsidiar sua melhor compreensão tem-se as figuras do equânime e do prudente, assinaladas por Hobuss (2009) ao examinar o entendimento de justiça na abordagem aristotélica das virtudes (Aristóteles, 2014), respectivamente nos territórios jurídico e da ética. Ainda, sob uma perspectiva jurídica mais estrita, também para ilustrar, a partir do contido no Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal, instituído pelo Decreto nº 1.171/1994 (Brasil, 1994), tem-se que o significado de ação moral examinada e situada – a atitude ética – extrapola a órbita da legalidade para visar a necessária busca do “*equilíbrio entre a legalidade e a finalidade*” (Brasil, 1994, Cap. I, Seção I, incisos II e III).

Reitere-se, desde já, que não se advoga uma IA capaz de tomar decisões éticas, senão: a) que a abordagem da codificação de condutas e princípios balizadores, seja dos profissionais desenvolvedores destas tecnologias, seja incorporadas ao próprio algoritmo, são de eficácia limitada; e b) o imperativo de que as pessoas – usuários – mantenham e ampliem suas capacidades de tomar decisões éticas tendo como subsídio os produtos gerados por IA’s.

Assume-se, também, que a IA é um espaço de produção de conhecimento e tecnologias de natureza tipicamente interdisciplinar, constituído de fronteiras fluidas e interdependentes em relação a outras dimensões do próprio campo das ciências da computação (Kaufman, 2016). Desse modo, são necessárias interseções de natureza transdisciplinar, assim entendidas aquelas que se propõem avançar em relação à interdisciplinaridade e produzir novos esquemas cognitivos (Japiassú, 2006). Aqui se pretende, portanto, um exame transversal em relação à filosofia, particularmente com foco no ramo da ética, e neste, uma análise orientada para a sua expressão aplicada, mas transcendendo a perspectiva convencional focada na coerção e sancionamento.

Também não se tem em vista identificar desafios éticos que se impõem ante o emprego da IA, tal como o fazem Rochel e Evéquoz (2021), senão o objetivo de estabelecer uma relação entre IA e ética assinalando o imperativo de superar os limites da prática recorrente de definição de diretrizes e da codificação de princípios e valores para a concepção de tecnologias de IA referida por Hagendorff (2022b), igualmente denunciada por Wernaart (2022) e Waelen (2022). Munn (2023), a propósito, no particular da criação de códigos de conduta, refere não somente a profusão destas publicações como a sua inutilidade por gerarem princípios

vagos e imprecisos. Diante disso, o texto propõe uma extensão do escopo de incidência da ética neste espaço dinâmico da produção de conhecimento, avançando para a proposição da necessária concepção e constituição de competências que habilitem pessoas usuárias dos produtos gerados pelas tecnologias de IA como ingredientes de exame dos temas promovendo uma inserção qualificada do agente, devidamente situada, esclarecida e consciente a subsidiar a tomada de decisão.

Ainda, a opção por abordar o serviço público prende-se à escala de impacto potencial dos resultados das decisões neste campo – a sociedade (ou segmentos por vezes sensíveis desta), atores específicos da cena política, entre outros, como é o caso das atividades desenvolvidas pelos agentes e órgãos de controle, por exemplo (Charalabidis, Medaglia & Noordt, 2024; Valle-Cruz, Gil-Garcia & Sandoval-Almazan, 2024; Straub, Morgan, Brighth & Margetts, 2023; García Rodríguez, Rodríguez-Montequín, Ballesteros-Pérez, Love & Signor, 2022; Steden, 2020; Sousa, Melo, Bermejo, Farias & Gomes, 2019).

Sobrevêm a esta introdução uma seção que aborda o entrelaçamento conceitual entre IA e ética em processo de franca elaboração. A seguir, examina-se a produção científica envolvendo o que a literatura tem definido como IA ética, explorando o que se entende por ética neste contexto. Adiante tem-se uma abordagem da ética como capacidade crítico-reflexiva

essencialmente humana, explorando-se os limites das codificações e definições de diretrizes para a concepção de “IA’s éticas”. Por fim, considerações que sinalizam, além dos limites da abordagem, pontos que podem constituir objeto de pesquisas futuras em maior amplitude e profundidade.

Ética na inteligência artificial: temas em processo de elaboração conceitual

A interseção entre ética e IA vem sendo estudada a partir de diferentes perspectivas, desde as abordagens de Garcia-Murillo e MacInnes (2024), de Bankins e Formosa (2023) e de Bankins (2021) envolvendo as implicações éticas da aplicação de IA no trabalho e no seu significado, até um exame mais crítico sobre o que a literatura vem denominando “IA ética” (Munn, 2023; Waelen, 2022; Hagendorff, 2022a; Rochel & Evéquoz, 2021; Franzke, Muis & Schäfer, 2021; Siau & Wang, 2020; Hagendorff, 2020). Trata-se, no entanto, de um terreno fértil para a investigação em ainda maior extensão e profundidade à luz da contribuição que pode a filosofia emprestar, mormente se reconhecidas as distintas vertentes teóricas da ética (Hagendorff, 2022b; Wernaart, 2022; Hanna & Kazim, 2021; Buchanan, 2020).

Conforme a legislação da União Europeia sobre IA (Parlamento Europeu, 2023, p. 137), IA é entendida como “um sistema baseado em máquina projetado para operar com vários níveis de

autonomia e que pode, para objetivos explícitos ou implícitos, gerar resultados como previsões, recomendações, ou decisões que influenciam ambientes físicos ou virtuais". De modo semelhante, conforme referem Giarmoleo, Ferrero, Rocchi e Pellegrini (2024), a OCDE estabelece os contornos da definição de IA como "um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia." (OCDE, 2024, p. 7). Assim, o senso de IA pode ser associado entre outros aspectos, a um esforço orientado para produzir a capacidade de gerar artificialmente operações lógicas e aprender com a experiência, também para realizar decisões ou subsidiar a deliberação (Banking & Formosa, 2023; Rochel & Evéquo, 2021; Wang, 2019; Buchanan, 2005).

Contudo, a capacidade de criar, que no ser humano envolve um conjunto complexo de interações ativadas também pela sensibilidade e outros traços distintivos da espécie, no caso das máquinas requer a prévia concepção de uma ordem de operações com este propósito, todas estas atravessadas por valores estabelecidos consciente ou inconscientemente pelo profissional que concebe os algoritmos. Assim, além da transmissão desses elementos valorativos próprios do desenvolvedor da tecnologia, há que se considerar os pressupostos implícitos

nos recortes de dados com os quais a IA é alimentada, que combinados podem produzir vieses na operação e nos resultados gerados.

Deste modo, é preciso reconhecer que os conceitos de neutralidade e objetividade, por vezes considerados qualidades destas tecnologias, podem estar comprometidos desde a origem do processo de concepção. E a explicitação desta dimensão da subjetividade é relevante para este esforço de produção científica na medida em que aqui também se procura deslocar, no processo de tomada de decisão, o foco convencional de análise da ética do desenvolvedor dos algoritmos para o usuário dos resultados que as tecnologias de IA produzem. Esse ponto, a propósito, remete a uma fértil discussão, aqui desbordada, acerca da natureza essencialmente política da administração pública e da subsequente dimensão subjetiva das decisões neste campo.

Segundo uma perspectiva predominantemente técnica (Rochel & Evéquo, 2021), as tecnologias de IA são capazes de oferecer respostas ágeis e seguras (do ponto de vista da lógica de programação adotada, as quais implicam escolhas) em relação à operação de elevados volumes de dados e informações, o que repercute em diferentes expressões de economia de recursos pela racionalização do tempo, de operações, etc. Entre as questões que se impõem, reside a referente ao uso que se faz das

ferramentas baseadas nestas tecnologias, e especialmente aqui: qual é o momento de inserção da ética? Seria por ocasião da concepção (desenvolvimento das tecnologias), estágio este em que se tem pensado largamente em diretrizes de conduta a regular o trabalho do profissional desenvolvedor dos algoritmos (Hagendorff, 2022b)? Ou seria pensar no atravessamento da ética por parte do usuário dos resultados produzidos pela IA para fins de tomada de decisão? Ou em ambas as fases? Mas antes de tudo, o que se tem entendido como ética neste campo?

Nesse cenário, Giarmoleo *et al.* (2024), após amplo esforço de revisão da literatura relacionando ética e IA, assinalam convergência orientada para dois aspectos essenciais a fundamentar a relevância do tema, a saber, a tecnologia é sempre determinada pelo ser humano, sendo social e culturalmente situada, encerrando, portanto, uma carga moral a ser, necessariamente submetida a escrutínio ético; e os múltiplos níveis de integração da IA e da ética: a “*ética desde a concepção*” (*ethics by design*), a “*ética na concepção*” (*ethics in AI design*), e a “*ética para o design de IA*” (*ethics for AI design*), esta última referindo-se aos códigos de conduta, princípios, padrões, processos e diretrizes a orientarem a ação dos desenvolvedores e usuários de sistemas de IA. Perceba-se, aqui, um senso de “*ética*” *ex ante*, ou seja, um conjunto de princípios e valores orientadores daqueles que concebem o instrumento

de IA. No exame que ora se propõe, contudo, ventila-se a ideia de que a despeito da sua relevância para a orientação dos desenvolvedores, tais parâmetros são insuficientes, também (talvez mais) importando a capacidade dos usuários de fazerem um uso ético (examinado, refletido e justificado) dos resultados gerados pela IA.

De modo geral, as investigações que abordam a ética nas aplicações de IA têm considerado que os algoritmos herdam não somente pressupostos valorativos por ocasião da sua concepção, mas também que as características das massas de dados utilizadas podem gerar padrões enviesados de análise (Waelen, 2022), especialmente em razão de como as máquinas “*aprendem*” a partir da influência do conteúdo das bases de dados e informações com as quais são alimentadas e treinadas, e da “*inteligência*” decorrente (Garcia, 2020; Piteira, Aparicio & Costa, 2019). Mas há, ainda, estudos mais críticos que alertam especificamente para as repercussões éticas do uso da inteligência artificial em determinados campos (Hagendorff, 2022a; 2020b; Siau & Wang, 2020).

Posto isso, para os fins desta análise também se faz necessário estabelecer duas categorias de emprego da IA no contexto da administração pública, a saber, a primeira para a oferta de serviços públicos mais simples e diretamente ao cidadão, e a segunda na análise e interpretação de grandes volumes de dados, e subsequente

produção de textos, indicadores, recomendações, relatórios em geral, entre outros, todos estes subsídios à tomada de decisão, cuja ênfase das análises sobre o uso da IA tem recaído em sua expressão instrumental e orientada para resultados, preponderantemente sob o ponto de vista técnico.

Pretende-se aqui, então, também abordar uma dimensão ainda não alcançada pela IA, e que é própria do ser humano: a capacidade – e o imperativo – de realizar juízos refletidos e situados, ou seja, sensível às circunstâncias, na esteira da vertente aristotélica da ética, ou ética das virtudes (Aristóteles, 2014; Hobuss, 2009). Faz-se isso a partir da diferenciação inicial entre o resultado entregue pelas tecnologias de inteligência artificial e o uso que as pessoas fazem com estes produtos do tratamento de dados para subsidiar a tomada de decisão. Mais precisamente, intenta-se alertar para a necessidade de não somente reafirmar a complexidade do processo de tomada de decisão, em especial no serviço público por suas repercussões de diferentes ordens, mas fazer incidir, por parte do usuário, um profundo exame ético entendido em seu significado substantivo, inclusive para fins de atuação dos agentes públicos nos espaços de ação discricionária (Boer & Raaphorst, 2023; Thomann, Engen & Tummers, 2018; Lotta & Santiago, 2017; Hupe, 2013).

Uma inteligência artificial ética? O que de fato se entende por ética?

A complexidade e as repercussões das ações no setor público exigem que as deliberações nas diferentes etapas do processo de consecução dos atos administrativos não somente considerem uma pauta difusa de elementos do contexto, como também se deem de modo consistente, entre outros, com os valores do interesse público, da dignidade e da justiça devidamente sopesados conforme o caso. É neste ponto que se busca compreender o atravessamento do tema da IA com a ética. Sendo assim, a ética exige ser tomada aqui em sua essência conceitual de fundo filosófico como exame situado de natureza crítico-reflexiva e de extensão radical incidente sobre os valores fundantes que presidem o pensamento e a ação do sujeito acerca do que seja o justo, o certo e o bom, com o intuito de justificar escolhas e deliberações de forma consciente e esclarecida (Bergue, 2022a; 2022b).

Antes de avançar é preciso referir outro elemento que afeta o caráter ético da ação dos agentes no serviço público a ser considerado para fins de apreciação da relação entre ética e IA, a saber, a diferença entre legalidade e eticidade (expressão da moralidade refletida). Tanto no que respeita aos conceitos de ato vinculado e ato discricionário, quanto à centralidade da legalidade como

princípio fundamental a orientar os atos da administração pública, impõe-se observar que o senso de ação moral examinada – ética – transcende a esfera da legalidade estrita. E nesses termos, para ilustrar, sinaliza o Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal, instituído pelo Decreto nº 1.171/1994 (Brasil, 1994), tal como definido no Capítulo I, Seção I, incisos II e III do Anexo ao referido Decreto.

II - O servidor público não poderá jamais desprezar o elemento ético de sua conduta. Assim, não terá que decidir somente entre o legal e o ilegal, o justo e o injusto, o conveniente e o inconveniente, o oportuno e o inoportuno, mas principalmente entre o honesto e o desonesto, consoante as regras contidas no art. 37, caput, e § 4º, da Constituição Federal.

III - A moralidade da Administração Pública não se limita à distinção entre o bem e o mal, devendo ser acrescida da idéia de que o fim é sempre o bem comum. O equilíbrio entre a legalidade e a finalidade, na conduta do servidor público, é que poderá consolidar a moralidade do ato administrativo (Brasil, 1994).

Assinala-se do excerto, indicando a não somente possível, mas necessária atuação do agente nos espaços de discricionariedade (Bovens

& Zouridis, 2002), além do destaque à dimensão ética da ação administrativa, especialmente o imperativo da busca do “*equilíbrio entre a legalidade e a finalidade*” (Brasil, 1994, Cap. I, Seção I, incisos II e III), que converge para o conceito de exame situado como condição para a deliberação e a justa consecução do ato administrativo. A ação do agente encerra, portanto, uma dimensão teleológica, ou seja, que considera uma finalidade coerente, tanto com o conteúdo ético do interesse público, como também com o significado substantivo de ser servidor público.

De que se trata, então? De ética na IA? De uma IA ética? Ou de um exame ético – crítico, reflexivo, radical e sensível às circunstâncias (situado) – incidente sobre os produtos das tecnologias de IA para a tomada de decisão? Uma revisão dos estudos integrando ética e IA demonstra tanto a fecundidade quanto o potencial de expansão da investigação sobre o tema enfatizando uma perspectiva crítica acerca dos limites que se impõem à constituição do que se possa entender como uma IA ética (Bankins & Formosa, 2023; Waelen, 2022; Hallamaa & Kalliokoski, 2022; Hagedorff, 2022a; 2022b; Hanna & Kazim, 2021; Buchanan, 2020; Garcia, 2020; Siau & Wang, 2020; Wang, 2019; Kaufman, 2016).

Se de um lado Kaufman (2016) faz alusão a uma *machine ethics* como sistemas conscientes dotados de valores humanos tais como a intuição,

Hagendorff (2022a) aponta lacunas da abordagem convencional da ética em tecnologias de IA, denominados pontos cegos de investigação (*blind spots*), e o faz assinalando a concentração de esforços na prescrição de princípios e valores, em geral identificados como explicabilidade, justiça, privacidade, segurança e responsabilidade, presentes na quase totalidade dos assim denominados códigos de ética atinentes ao tema (Hagendorff, 2022b). Nesses termos, as linhas de investigação convencionais parecem alimentar a expectativa de constituir uma IA ética subordinada a tais princípios, o que também se assenta no pressuposto tecnicista da inevitabilidade da evolução da tecnologia a que aludem Hanna e Kazim (2021, p. 407) como “*determinismo tecnológico*” e, nesse contexto, no imperativo de aperfeiçoar a capacidade técnica dos algoritmos de IA de produzirem resultados éticos. Será que uma IA é capaz de produzir resultados eticamente sustentados? Ou de tomar decisões conscientes e eticamente justificadas? Assume-se que não; e que os contornos de moralidade da deliberação de um instrumento baseado em IA decorrerá tanto do arranjo das massas de dados com as quais é alimentada, quanto dos parâmetros morais herdados e subjacentes à sua programação. Portanto, haverá, sim, uma deliberação conforme padrões morais dados em maior ou menor nível de explicitação, mas não haverá uma decisão propriamente ética (reflexivamente

escrutinada). A partir destas questões tem-se trajetórias relevantes de investigação que se seguem, sendo que a contribuição ora pretendida se orienta mais precisamente para o usuário dos resultados gerados pela IA nos processos de tomada de decisão (Rochel & Evéquoz, 2021).

Ainda, alternativamente, Hagendorff (2022a) propõe que se desloque o foco de atenção dos estudos e práticas subsequentes para alcançar o necessário exame acerca de que campos os usos da IA seriam eticamente defensáveis. Assim, destacando a importância de um exame do tema em maior radicalidade, o autor mostra-se crítico à aplicação indiscriminada da IA em todas as áreas da sociedade.

Como se disse, neste artigo propõe-se um passo lateral que busque explicitar a diferença entre os produtos gerados por uma IA e o uso que deles se faz, especialmente para fins de tomada de decisão, o que exige a capacidade de levar a efeito exames éticos por parte dos usuários (Bergue, 2022a; Rochel & Evéquoz, 2021). Está pressuposto aqui o fato de que a ética se dá e se exige, fundamentalmente, no processo de tomada de decisão como elemento substantivo de justificação da deliberação por uma alternativa dentre outras igualmente possíveis, o que Wernaart (2022, p. 13) define como “*problema moral*”. Desloca-se, assim, a ênfase da incidência da ética do estágio de concepção das ferramentas de IA (que pressupõe a possibilidade de uma “*ética*” instalada na IA, entendida como

um conjunto de parâmetros valorativos balizadores da sua concepção previamente definidos) para também, especialmente por parte dos agentes públicos, o momento de uso dos resultados que essas tecnologias produzem, o que exige a competência de realizar juízos éticos; capacidade esta que é própria do ser humano (por enquanto).

E, assinale-se, ainda que em horizonte temporal qualquer se venha a desenvolver uma tecnologia capaz de realizar artificialmente esses juízos éticos, a consecução deste intento não afastaria, até lá, o risco e as implicações de eventuais decisões tomadas sem o devido exame crítico e refletido levado a efeito por pessoas. Ainda, rigorosamente, dotar tecnologias de IA da capacidade de produzir estes juízos justificados a partir de esforços de crítica e reflexão situada exigirá das pessoas envolvidas nessa suposta elaboração de algoritmos a prévia compreensão acerca do que seja a deliberação ética em seu conteúdo substantivo, a saber, para além da enunciação ou codificação de princípios, valores ou diretrizes de programação dos algoritmos.

Note-se, sinteticamente, que, em parte, os desafios de pensar uma IA ética a partir da definição de parâmetros valorativos a informar a concepção dos correspondentes algoritmos se assemelha aos limites sabidamente enfrentados pela assimilação do conceito de uma gestão ética quando essencialmente

assentados em políticas concentradas na instituição de códigos de conduta e na orientação predominantemente disciplinar de sua aplicação, conforme alertado por Bergue (2022b). Nesta mesma linha, para Hagendorff (2022a, p. 852),

a ética da IA reduz frequentemente o seu escrutínio às decisões de concepção e frequentemente ignora a situação econômica ou política dos sistemas tecnológicos em questão. O problema é que mesmo sob condições em que todos os principais requisitos dos princípios éticos da IA são cumpridos (o que significa que os sistemas de IA produzem resultados justos, são explicáveis, preservam a privacidade, são robustos e responsáveis), as aplicações de IA podem ser utilizadas para exacerbar os danos ambientais, fomentar a opressão social, apoiar modelos de negócios antiéticos e outros desenvolvimentos problemáticos. (Hagendorff, 2022a, p. 852).

No excerto, além do estreito entendimento acerca do que possam ser “resultados justos”, está destacada a distância entre parâmetros morais *ex ante* (balizadores do desenvolvimento), e o imperativo de exames éticos posteriores, no uso dos resultados para a tomada de decisão. Portanto, o desafio de uma deliberação eticamente consciente e justificada não reside ou se

esgota no desenvolvimento dos algoritmos de IA aderentes a parâmetros morais pré-definidos, mas na subsequente capacidade do agente de tomar decisões éticas fazendo uso da tecnologia e seus produtos como subsídios para a formação do juízo.

O deslinde desse problema requer, então, e antes, um assentamento conceitual fundamental, a saber: a diferença entre ética e moral. Quando se toma a ética como um conjunto de princípios e valores operacionalizados em regras para informar a concepção dos algoritmos de IA, não se está tratando efetiva e conceitualmente de ética em sua plenitude e potência conceitual, mas de diretrizes de conduta moral, que são estáticas, porque programadas no caso das IA's (e socialmente constituídas e reproduzidas no caso das coletividades em geral). São critérios ou regras balizadoras da operação do algoritmo. Assim, a prescrição de parâmetros de comportamento, seja para as pessoas, seja para a operação das máquinas, não esgota o domínio da ética, senão mais se alinha ao conceito de moralidade, que pode ser traduzido em normas de conduta, tais como as codificadas, tanto nas organizações, quanto nas profissões (Bergue, 2022b).

Nessa mesma linha, para Cortina e Martinez (2005, p. 20),

Essa distinção é útil, pois se trata de dois níveis de reflexão diferentes, dois níveis de pensamento e de linguagem acerca da ação moral, e

por isso se torna necessário utilizar dois termos diferentes se não queremos cair em confusões. Assim, chamamos de “moral” esse conjunto de princípios, normas e valores que cada geração transmite à geração seguinte na confiança de que se trata de um bom legado de orientações sobre o modo de se comportar para viver uma vida boa e justa. E chamamos de “Ética” essa disciplina filosófica que constitui uma reflexão de segunda ordem sobre os problemas morais. A pergunta básica da moral seria, então: “O que devemos fazer?”, ao passo que a questão central da Ética seria antes: “Por que devemos?”, ou seja, “Que argumentos corroboram e sustentam o código moral que estamos aceitando como guia de conduta?” (Cortina & Martínez, 2005, p. 20).

A ética constitui, nestes termos, o exame rigoroso, percuciente e em profundidade sobre os parâmetros morais; não se reduz, portanto, a disciplina, a códigos de conduta, ou à atuação de colegiados de ética (Bergue, 2022b). Em geral, age-se moralmente até que dilemas se manifestem, oportunidades em que os exames éticos se impõem a fim de esclarecê-los e solvê-los. Ética, então, em muito transcende aqueles conceitos.

Diferentemente, contudo, percebe-se a predominância de estudos que afirmam se preocupar com a ética

no espaço da IA fazendo referências à enunciação e codificação de valores e princípios éticos (Hagendorff, 2020) a orientar a ação dos desenvolvedores destas tecnologias. Hanna e Kazim (2021, p. 407), bem evidenciam essa perspectiva no território da IA, ao indicarem que

a ética da IA é a tentativa de orientar a conduta humana na concepção e utilização de autômatos artificiais ou máquinas artificiais, também conhecidos como computadores, em particular, formulando e seguindo racionalmente princípios ou regras que reflitam os nossos compromissos individuais e sociais básicos e os nossos principais ideais e valores. (Hanna & Kazim, 2021, p. 407).

Aqui se observa a tendência recorrente de apropriação da ética como codificação de conduta a orientar a ação dos sujeitos no desenvolvimento e emprego das tecnologias de IA, tanto quanto explicita que estas soluções devam repercutir os principais compromissos, ideais e valores. Atente-se para o fato que não se advoga a impropriedade, tampouco o afastamento destas definições, senão que são insuficientes, especialmente no que se refere aos fins últimos de sua utilização. Uma deliberação ética requer a elaboração de juízos situados e refletidos em seus parâmetros morais subjacentes ao pensamento do próprio

sujeito que decide. A codificação de princípios de conduta pode, por certo, operar como instrumento de educação e orientação, inclusive como referencial para eventual sancionamento, mas está longe de esgotar, como a realidade demonstra, o tratamento possível e consistente do tema. Concorrendo, ainda, para a explicitação das lacunas desta abordagem, Mittelstadt (2019) assinala os limites de uma tendência à convergência da ética da IA direcionada para a ética médica a partir de um conjunto de princípios clássicos já consubstanciados nas codificações de conduta daquele campo.

Também nesta perspectiva da ética aplicada, assentada na codificação de princípios, Hallamaa e Kalliokoski (2022) referem as Diretrizes de ética para inteligência artificial confiável produzida por especialistas em IA a partir de demanda da Comissão Europeia em 2019, constituindo-se, segundo as autoras como *“a tentativa mais extensa até agora de criar uma ferramenta ética para orientar o desenvolvimento e design de IA.”* (Hallamaa & Kalliokoski, 2022, p. 2).

Veja-se que a crença subjacente reside na geração de um desenho de IA que seja ético, ou seja, *“uma IA justa, não discriminatória, transparente, segura e protegida”* (Hallamaa & Kalliokoski, 2022, p. 2) como resultado da prescrição de padrões normativos de conduta e diretrizes de concepção das tecnologias. Neste percurso de

entendimento acerca do que se possa entender por uma IA ética, tem-se, essencialmente diretrizes que sinalizam balizadores de responsabilização das pessoas que constroem essas ferramentas (Maas, 2022), que mais se alinham à prática geral da instituição de códigos de conduta na administração pública, a saber, a instituição de um comportamento disciplinar e persecutório movido por atitudes inquisidoras, policiaesca, sancionadoras e punitivas que pouco ou nada se relacionam com ética em seu preciso significado.

O corpus da ética europeia da IA compreende um sistema de valores e princípios, onde a aplicação das normas a casos e questões concretas se traduzirá numa ação eticamente sólida, ou seja, num mecanismo que faz a mediação entre os ideais morais na realidade dos cidadãos e das sociedades. O nível normativo primário do documento compreende ideais éticos que apresentam uma base para uma vida boa e virtuosa. Estes ideais fornecem uma estrutura para a compreensão de características moralmente relevantes e para a conceptualização dos constituintes morais de uma determinada atividade dentro do seu contexto.

O segundo nível apresenta diretrizes para deliberação e raciocínio moral para ajudar desenvolvedores, designers e usuários a pensar antes de agir e a tomar diversas decisões

durante os processos de desenvolvimento e manutenção. Outro trunfo que serve à deliberação ética é o critério de escolha de metas e ações. O objetivo destas ferramentas é garantir que os objetivos escolhidos (e as ações destinadas a alcançá-los) sejam moralmente aceitáveis. O documento também contém listas de verificação e regras práticas para apoiar os profissionais a tornar um procedimento de rotina a verificação de falhas morais em seus produtos, aplicações e serviços. O último nível de avaliação normativa ocorre *ex post facto*, quando as escolhas feitas (e os resultados que delas resultaram) são escrutinadas e julgadas. (Hallamaa & Kalliokoski, 2022, p. 3).

Reconhecida a importância de cada nível descrito, é a partir deste último nível referido – o de “*de avaliação normativa ocorre ex post facto, quando as escolhas feitas são escrutinadas e julgadas*” – que se assenta a atenção deste artigo. Assim, a opção pelo percurso de elaborar declarações que descrevem princípios e valores a pautar as ações no âmbito da IA (Hagendorff, 2020; Mittelstadt, 2019) parece orientar-se para o mesmo desafio que se encontra na implementação dos códigos de conduta (Hallamaa & Kalliokoski, 2022). De se destacar que neste mesmo

sentido avança Munn (2023, p. 871), que apesar de empreender uma crítica à fértil produção de códigos de conduta, mantém uma perspectiva de sancionamento ao referir-se a princípios éticos “*desdentados*” (*toothless*), pela ausência de “*consequências*”. Perceba-se como subjacente ao entendimento do autor um senso de imperativo de punição ou sancionamento daquele que porventura transgrida os parâmetros de conduta codificados. Reitere-se, são ações necessárias, especialmente para os casos de aplicações mais limitadas, mas não suficientes.

É fundamental, então, diferenciar uma IA que se pretenda ética (pela ação ética dos seus programadores), geradora de resultados que observem padrões morais estabelecidos; do produto destas ferramentas de IA a serem utilizados pelas pessoas como subsídio para análises subsequentes e tomada de decisão eticamente justificadas. Tomando-se os casos em que os resultados gerados pelas tecnologias de IA constituem subsídios para a tomada de decisão, a apropriação do conceito de ética passa a ser distinto (não restrito às codificações de diretrizes antes aludidas e processamentos disciplinares correspondentes), a saber, constituindo-se como a capacidade de examinar crítica e reflexivamente, de forma situada este ingrediente (o produto da IA: relatórios, informações etc.) entre os demais componentes de

deliberação, sem abdicar da autonomia de pensamento e ação.

Ética como capacidade crítico-reflexiva (ainda) essencialmente humana: um resgate do homem parentético e da autonomia do sujeito

Assentou-se que a ética, diferentemente (ainda que não dissociada) da moral, envolve tanto um esforço crítico-reflexivo de extensão radical – além de corajoso, porque expõe as entranhas do próprio pensamento –, quanto um componente situacional que particularizam as condições do juízo. Se a moral é acrítica e irrefletida, de certo modo operando na superfície dos fenômenos; a ética requer um exame não só agudo e percuciente, mas sensível às circunstâncias e devidamente contextualizado dos atos e fatos em apreciação. Essa diferenciação – não significando que estejam dissociados os conceitos, reitere-se – é essencial, seja para se esclarecer e alcançar o pretendido significado da ética como atitude assentada em um pensamento examinado, seja para bem demarcar o lugar do ser humano no processo de tomada de decisão ética no campo de aplicação da IA.

No esforço de definir os contornos do conceito de ética, merece destaque, por oportuno, o conceito de

reflexão. Refletir é uma forma particular de pensar. A despeito do recorrente uso comum intercambiável entre os verbos pensar e refletir, com sensível tendência recente inclinando-se para este último, inclusive como modismo de discurso, tratam-se de ações distintas em sua substância.

Mais precisamente, a reflexão pode ser entendida como uma forma particular de atitude intelectual, pois implica uma ação de pensar levada a efeito pelo sujeito que se volta para si próprio interrogando-se sobre os fundamentos valorativos delineadores do seu pensar. Segundo Japiassú e Marcondes (2006, p. 236) o esforço reflexivo constitui uma *“introspecção pela qual o pensamento volta-se para si mesmo, investiga a si mesmo [...] estabelecendo os princípios que a fundamentam. [...] Consciência crítica [...] que examina sua própria constituição, seus próprios pressupostos”*. Nessa mesma linha, Abbagnano (2012, p. 986) destaca três expressões para a reflexão, a saber: *“1ª como conhecimento que o intelecto tem de si mesmo; 2ª como consciência; 3ª como abstração.”* De modo geral tende-se a reduzir a reflexão à abstração, ou ao simples pensar, levando o conceito a perder sua potência.

Refletir é, assim, requisito para a autonomia, ou seja, a capacidade de deliberar e agir sem se deixar conduzir pelo outro, implicando autodeterminação, inclusive como condição do sujeito ante o produto gerado pelas IA's. Ou, ainda, alinha-se

ao que o filósofo iluminista Immanuel Kant denominou de esclarecimento (Kant, 2012, p. 63-64), a saber, a saída do sujeito de uma condição de menoridade, definida como *“a incapacidade de fazer uso do seu entendimento sem a direção de outro indivíduo”* (aqui, o algoritmo de IA), da qual ele próprio é responsável *“se a causa dela não se encontra na falta de entendimento, mas na falta de decisão e coragem de servir-se de si mesmo sem a direção de outrem. Sapere aude!”*. E essa independência que se funda no senso de autonomia, mormente em relação à tecnologia – que é recurso, portanto, algo do que o sujeito se serve, e não algo ao que se subordina –, constitui atributo esperado no processo de tomada de decisão em todas as áreas e níveis institucionais no contexto da administração pública.

Também é importante assinalar, antes, as diferenças entre perguntar, criticar e refletir. Por certo os algoritmos podem ser capazes de propor perguntas (na sua expressão mais elementar, pela aposição de um ponto de interrogação em uma frase), mas seria a IA capaz de criticar? Criticar requer, em apertada síntese, que se produza um questionamento contundente e devidamente fundamentado com o intuito de promover um estremeamento na estrutura de argumentativa ou de entendimento do interlocutor a fim de fazê-lo pensar melhor. Mas a crítica, em situações particulares, também pode ser enraizada a ponto de

posicionar-se rigidamente em um lugar de pensamento e ação, limitando a sua potência. Especialmente se for entendida assim (em sua expressão parcial), pode-se até conceber que um algoritmo possa vir a ser capaz de tais intervenções de questionamento mais agudo, ainda que a partir de uma perspectiva localizada. No entanto, o conceito fundamental quando se pensa em ética parece ser o de reflexividade, a saber, a capacidade de empreender um exame autodirigido em busca de consciência sobre os fundamentos que presidem o próprio pensamento; logo, a capacidade de colocar em perspectiva e sob exame percuciente em radicalidade a própria posição de pensamento e ação. Este estado de consciência em relação aos elementos de decisão, seus fundamentos valorativos e as circunstâncias de situação é que constituem o diferencial da capacidade humana de deliberar eticamente. Assim, os algoritmos de IA podem ser concebidos de modo a perguntar, mesmo a produzir uma crítica (especialmente se esta configurar-se como parcial por engastada em um lugar de percepção do mundo), mas ainda não a refletir, o que exige consciência situada e autodirigida. Há que se considerar, ainda, o recorrente desafio de alinhamento conceitual e de expectativas de resultados a ser estabelecido entre os segmentos de desenvolvimento de algoritmos de IA e os ditos “*de negócio*”, mormente quando se pensa nas diferentes áreas e

peculiaridades atinentes a cada tema e seus correspondentes arranjos de variáveis e elementos valorativos fundantes.

Reitere-se, as tecnologias baseadas em recursos de IA, notadamente as que operam no nível do aprendizado de máquina, produzem (e depois reproduzem) sua “*inteligência*” – capacidade de entregar respostas adequadas a problemas emergentes – também a partir das bases de conteúdo com as quais são alimentadas, processo que potencialmente gera padrões referenciais que se caracterizam por vieses (Garcia, 2020). Tal qual a aprendizagem humana em sua expressão mais elementar, aquela na qual o sujeito se apropria de elementos de seu contexto, conforme uma matriz cultural e de conceitos através dos quais interpreta seu entorno, a inteligência artificial procede de modo semelhante em face daquilo com que se alimenta. Assim, o produto da inteligência reflete não somente a qualidade, mas os contornos de recorte do conteúdo de dados que absorve (Garcia, 2020).

Desse modo, tanto quanto o ser humano depende da competência crítico-reflexiva para posicionar-se de modo qualificado ante os temas e desafios que se lhe apresentam, também o produto da inteligência artificial reclama do sujeito um exame crítico e um escrutínio ético refletido estendido às raízes valorativas do pensamento vigente. Se no caso das

As pessoas têm-se conceitos e preconceitos, a inteligência artificial tem potencial de produzir vieses de geração de resultados. Assim, a exemplo dos preconceitos de modo geral, os vieses gerados pela IA nem sempre são explícitos, impondo-se a incidência de um exame que se estenda à radicalidade dos seus pressupostos constituintes e demais elementos circunstanciais que afetam o fenômeno sob análise.

Neste ponto exsurge a importância da capacidade crítica plena e reflexiva incidente sobre os resultados produzidos pelas IA's. Se de um lado é possível estabelecer parâmetros morais de programação dos algoritmos, de outro é preciso ter em mente que nem sempre se tem a possibilidade de "controlar" a qualidade dos dados utilizados e seus vieses encobertos. Daí a relevância de constituir competências tais que, para além da concepção, habilitem e qualifiquem os agentes para, então, interagir, examinar e depurar os resultados das IA's, tomando-os como insumos, entre outros, para o processo de deliberação ética. A capacidade de identificar valores, pressupostos e pré-conceitos intrínsecos – ocultos, portanto – é algo que a apreciação crítica permite. O pensamento crítico-reflexivo não só produz conceitos em resposta aos esforços de problematização, mas os reexamina e redefine. E assim o faz com os preconceitos, que têm como uma de suas expressões os referidos vieses.

Reitere-se que há aplicações de IA em que externalidades negativas têm menor potencial ofensivo, mas há outras com implicações de maior vulto, tais como aquelas em que os resultados destas tecnologias integram o elenco de subsídios para a tomada de decisão, sobretudo que afetam a vida, a saúde, a segurança, a imagem e a liberdade das pessoas. Para ilustrar, tome-se o caso em que uma ferramenta baseada em IA analisa um determinado conjunto de dados e emite um juízo dito técnico (supostamente "neutro" e "objetivo") sobre a responsabilidade de uma pessoa, ou o caso em que uma tecnologia de IA produz uma sentença ou despacho interlocutório qualquer, ou ainda um parecer técnico que orienta para uma diretriz de deliberação ou de ação, entre outras situações potencialmente impactantes. Todas estas são respostas de uma IA que repercutirão na vida em sentido amplo de pessoas e comunidades, mas que antes de serem reconhecidas e aceitas como resultados finalísticos, devem ser tomados como insumos para os processos de exame e deliberação que, submetidas a profundo escrutínio crítico-reflexivo, informem, efetivamente, juízos éticos. Alerta-se, portanto, para o potencial danoso de situações em que as pessoas passem a abdicar daquilo que é próprio do ser humano, neste caso a capacidade de realizar exames que envolvam um conjunto de ações crítico-reflexivas que perscrutem os fundamentos valorativos do próprio

pensar, e de forma situada (o que significa em atenção a elementos subjetivos e próprios do contexto).

Assinala-se, desse modo, o imperativo de atentar para que o avanço do potencial das tecnologias de IA não transforme as pessoas em autômatos naturais de consciência limitada. Também, que em nome da produtividade, da racionalização ou da ingênua busca de objetividade não se subtraia a substância do que seja ser humano e o significado e as repercussões das suas ações. Isso parece sugerir que deva haver limites para a aplicação da IA, se não em relação a áreas da sociedade tal como advoga Hagendorff (2022a), então em momentos críticos do processo de tomada de decisão que tomem seus produtos como elementos de deliberação tendo em vista sobretudo aquilo que é o objeto essencial da ética – a justiça.

Nesta eventual condição subordinada, as pessoas transformam-se em meios no processo de decisão, e não em protagonistas quanto aos fins em si. Neste caso, vale retomar lição da filosofia de Immanuel Kant, que na formulação da humanidade dada ao conceito de imperativo categórico – referencial básico a partir do qual se pode considerar como ética uma ação – assim refere:

Age de tal maneira que uses a humanidade, tanto na tua pessoa como na de qualquer outro, sempre e simultaneamente como fim e

nunca simplesmente como meio. (Kant, 2019, p. 73).

Alinhado à inspiração da deontologia kantiana, Hanna e Kazim (2021, p. 406), ao se referirem à dignidade humana como fundamento da ética na IA destacam:

Primeiro, a dignidade humana realmente existe porque (i) ninguém, nem mesmo um cético em relação à dignidade, poderia dar o seu consentimento racional, real ou possível, para ser tratado como um mero meio para os fins de alguém (isto é, seus desejos ou objetivos) ou como uma mera coisa, e (ii) ser absolutamente inadmissível tratar qualquer pessoa (incluindo a si mesmo) como um mero meio ou como uma mera coisa é uma propriedade essencial da dignidade humana. (Hanna & Kazim, 2021, p. 406).

Deve-se reconhecer, em caráter de alerta, que a exemplo da racionalização que o modelo de produção de inspiração taylorista-fordista imprimiu nas organizações e estendeu-se às demais dimensões da vida das pessoas em sociedade até a contemporaneidade, a IA é apenas mais uma tecnologia. Impactante e relevante sim, mas é antes um instrumento – um meio – e não a substituição do sujeito particularmente naquilo que é próprio do ser humano,

notadamente a elaboração de juízos éticos. Pensar a ética na IA implica, portanto, considerar o pensamento ético como exercício humano no contexto do uso dos resultados gerados pela tecnologia.

Neste aspecto, o pensamento ético, como expressão da prática do fazer filosófico em busca da deliberação justa, exige também o que Guerreiro Ramos denominava atitude parentética, a saber aquela levada a efeito pelo homem parentético (em contraste com o que denominou homem operacional). Essa atitude de inserção social do sujeito – a parentética – corresponde à capacidade de colocar o fenômeno entre parêntesis, suspendê-lo, a fim de capturar a sua substância essencial (Santos, Leal, Serafim & Moraes, 2018). Nesta condição o sujeito exercita um exame crítico de consciência em relação aos valores envolvidos (Ramos, 1984; Bergue, 2022c; 2022d) e eventuais conflitos entre estes (Oldenhof, Postma & Putters, 2014).

Pretende-se com isso assinalar o necessário cuidado tanto com o uso que se faz dos resultados da IA, quanto com a eventual tendência de delegação para uma ferramenta de IA de atividades que requerem exames éticos incidentes sobre um conjunto de elementos situacionais de decisão outros, para além daqueles com os quais a ferramenta foi alimentada e o algoritmo parametrizado para realizar.

É preciso, portanto, diferenciar o produto das ferramentas de IA –

relatórios, recomendações, pareceres, análises, gráficos, indicadores, etc. – do uso e, especialmente, dos juízos que se faz a partir destes resultados tendo-se em vista as implicações das decisões. Reitere-se e reconheça-se que tecnologias baseadas em IA na administração pública são capazes de oferecer serviços ou suporte ao cidadão de forma mais ágil e econômica, mesmo de modo mais eficaz; mas outro uso da tecnologia, que inspira mais cuidado, reside na produção de informações para a tomada de decisão, e a ética atravessa estas duas situações de formas distintas.

Quando o produto das tecnologias de IA – relatórios, gráficos, imagens, indicadores – são assumidos como subsídios para o processo de tomada de decisão, o juízo ético reafirma-se em seu lugar natural – o ser humano. Isso porque são as pessoas que deliberam de modo contextualizado sobre o uso que fazem dos produtos das tecnologias de IA. Nesta vertente de uso da IA o risco reside em delegar a elas (IA's) aquele esforço deliberativo (ainda que concebidas sob parâmetros de conduta especificados e assumidos como “éticos”) que exige a sensibilidade humana em busca do senso mais profundo e orgânico acerca do que seja o certo e o justo. Isso porque as ferramentas de IA proporcionam importantes análises e sínteses de dados e informações, mas não alcançam a capacidade de produzir um

juízo dinâmico, consciente e circunstanciado do caso.

Note-se, ainda, que a proposta de restrição dos campos de aplicação advogada por Hagendorff (2022a) constitui, de fato, um limite real de difícil implementação, mormente pelas razões de fundo econômico e múltiplos outros interesses incidentes. Sendo assim, senão alternativamente, então adicionalmente, fazem-se necessários esforços de educação que capacitem agentes para o exercício de um pensamento ético especialmente no curso do processo de tomada de decisão no qual os produtos da IA constituam recursos de apreciação. Reafirme-se que pensar a ação ética não se restringe a orientar pessoas para a aplicação de regras, princípios e valores prescritos em códigos formais (Hagendorff, 2020; Mittelstadt, 2019), mas educá-las para a realização de um exame crítico-reflexivo de extensão radical e situado que alcance os fundamentos de sustentação da ação em busca da justificação de sua pertinência.

Há, pois, uma questão de fundo aqui: a distância entre o saber e o fazer. Ferraz (2014, p. 176), assinala que saber o que seja o certo difere sensivelmente de fazer o certo. Nesse contexto, e particularmente sobre a crença na ação racional humana, merece destaque a lição de Kant (2006, p.205), para quem a capacidade de aperfeiçoar-se conforma o sujeito como um ser “dotado da faculdade da razão (*animal rationabile*)” que pode “fazer de si um

animal racional (animal rationale)” (Kant, 2006, p. 205). Nesses termos, como assinala Ferraz (2014), em alusão a Kant (2006), o sujeito não é um ser racional, mas um ser capaz de racionalidade. Então, o será sempre que se propuser a fazer uso de sua razão para agir.

Sendo assim, a norma e seus consectários de sancionamento, a despeito de importantes, são insuficientes, de modo que emerge a necessidade também de outro aspecto essencial e precedente: um arranjo valorativo intrínseco ao sujeito e conformador de uma ideologia ética. Este, sabe-se, constituído a partir de um longo processo formativo consistente com a transformação cultural pretendida. Encontram-se em disputa, portanto, neste caso, o entendimento que valoriza a educação ante outro que recorre à mudança pela coerção assentada na explicitação do risco de punição dos agentes.

Para isso, entende-se que ao agente público, em especial pela extensão das repercussões que as suas decisões podem alcançar, é crescentemente necessária a capacidade crítico-reflexiva de extensão radical e situada. Essa competência refere-se àquela em que o sujeito corajosamente pratica o percuciente esforço de voltar-se para si e examina os fundamentos valorativos e pressupostos do próprio pensamento (Bergue, 2022a). Exercício este, a propósito, condição para um juízo esclarecido e contextualizado, e que

fundamenta e habilita a atuação firme, consciente e justificada especialmente no espaço da ação discricionária. Sem isso, eleva-se o risco da desumanização e perda do significado substantivo do trabalho, mormente a substância ética de ser servidor público, conceito aqui tomado em sentido amplo.

Considerações finais

Pretendeu-se uma breve contribuição para o pensamento e a reflexão sobre conceito de ética no curso dos esforços de seu tratamento no contexto da IA como subsídio à tomada de decisão. Não se buscou abordar a IA propriamente, senão o imperativo da capacidade de pensar eticamente em seu mais substantivo significado e como uma ação humana. A IA aqui opera, então, quase como uma metáfora.

Assentou-se, em suma, que o debate envolvendo a ética na IA ou o que espera uma ética da IA mantém-se ainda preso à reduzida interpretação da ética como composição de um conjunto de valores e princípios capazes de nortear a conduta humana. E é isso também; mas não só. Estes elementos são insumos para o exame ético. A eles somam-se, além da situação, a capacidade de nela sopesar valores e princípios em conflito, solvendo o que se denomina de dilemas morais.

Não é suficiente esperar que a proposição de um código de conduta, com a maior complexidade e

completude em que possa assentar sua concepção e aplicação, dê conta dos desafios da ética, não somente neste terreno da inteligência artificial, mas especialmente nele. Mesmo os textos que se propõem a uma abordagem “crítica” de uma IA ética orientam-se por este percurso, como é o caso de Waelen (2022), que refere princípios balizadores da programação destas ferramentas de IA.

E em se tratando de administração pública, particularmente no que se refere à ética, esse pensar de alcance radical precisa incidir sobre os fundamentos (valores e pressupostos) dos atos e fatos administrativos com vistas ao que seja o justo, o bom e o certo. Ética implica – independentemente da corrente teórica que se tome como referência de análise – sempre uma capacidade de examinar não somente o objeto, mas antes a si mesmo em relação aos próprios alicerces valorativos do pensar. Já se disse que a ética – o pensamento crítico-reflexivo de alcance radical e situado – examina a moral entendida como padrões de conduta socialmente construídos e aceitos; e é neste ponto que se impõe a questão sobre quem examina não somente os resultados gerados por uma IA, mas as implicações éticas do conteúdo que esta tecnologia produz.

Tem-se, assim, nestes tempos de tecnologia em ascensão, não somente a necessidade, mas antes a excepcional oportunidade de efetivamente nos voltarmos àquilo que define o humano;

reconhecendo não somente a centralidade, mas a riqueza advinda da assunção da subjetividade, dado que tudo passa pelo juízo do sujeito, e pouco de fato se sustenta na frieza do objeto que por vezes alimenta a crença na tão aventada quanto ingênua expectativa de objetividade nas ações e decisões, e com especial ênfase na administração pública, seja pela sua amplitude e complexidade, seja pela sua sabida substância política e implicações todas que deste traço advém.

É sempre importante frisar que os limites normalmente atribuídos à IA são, de fato, mais propriamente limites humanos, mormente de fundo sociocultural, refletidos na concepção e operação destas tecnologias, seja pelas dificuldades de programação, seja pelas deficiências associadas a como são alimentadas (fonte e qualidade dos dados, que são sempre recortes da realidade). Retorne-se à questão da ética ser apropriada no transcurso formativo das pessoas ao longo da sua vida. A ética é antes um problema de educação que tão-somente se reproduz no mundo do trabalho (Munn, 2023).

Em síntese, em termos de limitações deste esforço de exame teórico tem-se não somente a revisão da literatura, que sempre pode ser ampliada e melhor especificada com o que se produz doravante, mas o aprofundamento dos enlaces entre ética e IA para além das codificações de conduta, e o exame de pontos outros transversais como o caso da

responsabilidade, da liberdade das pessoas. Também entre as possibilidades de ampliação do escopo do estudo está o aprofundamento empírico do tema e das repercussões destacadas não somente em órgãos de controle da administração pública, mas em políticas públicas de modo geral.

Referências

- Abbagnano, N. (2012). *Dicionário de Filosofia*. São Paulo: Editora WMF Martins Fontes.
- Aristóteles. (2014). *Ética a Nicômaco*. 4ª Edição, São Paulo: EDIPRO.
- Bankins, S. & Formosa, P. (2023). The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work. *Journal of Business Ethics*, 185, 725–740. <https://doi.org/10.1007/s10551-023-05339-7>
- Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: a decision-making framework. *Ethics and Information Technology*, 23, 841–854. <https://doi.org/10.1007/s10676-021-09619-6>
- Bergue, S. T. (2022a). Ética como competência: interseções entre a administração e a filosofia. *Revista Gestão & Planejamento*, 23, 73-87. DOI: 10.53706/gep.v.23.7297
- Bergue, S. T. (2022b). Ética, códigos de conduta e integridade na administração pública brasileira. *Administração Pública e Gestão Social*, 14(4).

- <https://doi.org/10.21118/apgs.v14i4.13459>
- Bergue, S. T. (2022c). Contribuições da problematização filosófica para o estudo da administração pública. *Revista Eletrônica de Administração*, 28(1). DOI: 10.1590/1413-2311.344.118517
- Bergue, S. T. (2022d). *Pensamento filosófico na administração pública*. Caxias do Sul: EDUCS. Acessível em: <https://www.ucs.br/educs/livro/pensamento-filosofico-na-administracao-publica/>
- Boer, N. & Raaphorst, N. (2023). Automation and discretion: explaining the effect of automation on how street-level bureaucrats enforce. *Public Management Review*, 25(1), 42-62. <https://doi.org/10.1080/14719037.2021.1937684>
- Bovens, M. & Zouridis, S. (2002). From street-level to systems-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174-184. <https://doi.org/10.1111/0033-3352.00168>
- Brasil. (2023). Decreto nº 11.529, de 16 de maio de 2023. *Institui o Sistema de Integridade, Transparência e Acesso à Informação da Administração Pública Federal e a Política de Transparência e Acesso à Informação da Administração Pública Federal*. Diário Oficial da União, Brasília, p. 4, 2023.
- Brasil. (1994). Decreto nº 1.171, de 22 de junho de 1994. *Aprova o Código de Ética Profissional do Servidor Público Civil do Poder Executivo Federal*. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto/d1171.htm. Acesso em 08 jun 2021.
- Buchanan, B. G. A. (2005). (Very) Brief History of Artificial Intelligence. *AI Magazine*, 26(4), 53-60. DOI: 10.1609/aimag.v26i4.1848. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1848>.
- Charalabidis, Y., Medaglia, R. & Noordt, C. (2024). Introduction to the Research handbook on public management and artificial intelligence. (p. 1-6) In Y. Charalabidis, R. Medaglia & C. Noordt. *Research handbook on public management and artificial intelligence*. Edward Elgar Publishing: Cheltenham. <http://dx.doi.org/10.4337/9781802207347>
- Cortina, A. & Martínez, E. (2005). *Ética*. São Paulo: Edições Loyola.
- Deleuze, G. & Guattari, F. (1992). *O que é Filosofia?* Rio de Janeiro: Editora 34.
- Downe, J., Cowell, R. & Morgan, K. (2016). What determines ethical behavior in public organizations: Is it rules or leadership? *Public Administration Review*, 76(6), 898-909. DOI: 10.1111/puar.12562.
- Ferraz, C. A. (2014). *Elementos de ética*. Pelotas: NEPFIL OnLine. acesso: <http://nepfil.ufpel.edu.br>

- Franzke, A. S., Muis, I. & Schäfer, M. T. (2021). Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology*, 23, 551–567. <https://doi.org/10.1007/s10676-020-09577-5>
- Gal, U., Hansen, S. & Lee, A. S. (2022). Research Perspectives: Toward Theoretical Rigor in Ethical Analysis: The Case of Algorithmic Decision-Making Systems. *Journal of the Association for Information Systems*, 23(1), 1634-1661. DOI: 10.17705/1jais.00784
- Garcia-Murillo, M. & Macinnes, I. (2024). Introduction to the handbook of artificial intelligence at work: interconnections and policy implications. (p. 1-14). In M. Garcia-Murillo & I. A. R. Macinnes. *Handbook of artificial intelligence at work: interconnections and policy implications*. Edward Elgar Publishing: Cheltenham. <http://dx.doi.org/10.4337/9781800889972>
- Garcia, A. C. B. (2020). Ética e inteligência artificial. *Computação Brasil*, 14-22. DOI: 10.5753/CompBR.2020.43.1791
- García Rodríguez, M. J., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E. D. & Signor, R. Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*, 133. <https://doi.org/10.1016/j.autcon.2021.104047>
- Giarmoleo, F. V., Ferrero, I., Rocchi, M. & Pellegrini, M. M. (2024). What ethics can say on artificial intelligence: Insights from a systematic literature review. *Business and Society Review*, 1–35. <https://doi.org/10.1111/basr.12336>
- Guerreiro Ramos, A. (1984). Modelos de homem e teoria administrativa. *Revista de Administração Pública*, 18(2), 3-12.
- Hagendorff, T. (2022a). Blind spots in AI ethics. *AI and Ethics*, 2, 851–867. doi.org/10.1007/s43681-021-00122-8.
- Hagendorff, T. (2022b). A Virtue-Based Framework to Support Putting AI Ethics into Practice. *Philosophy & Technology*. 35(55). <https://doi.org/10.1007/s13347-022-00553-z>
- Hagendorff, T. (2020). The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines*, 30, 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hallamaa, J. & Kalliokoski, T. (2022). AI Ethics as applied ethics. *Frontiers in Computer Science*, 4. DOI:10.3389/fcomp.2022.776837
- Hanna, R. & Kazim, E. (2021). Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach. *AI and Ethics*, 1, 405–423. <https://doi.org/10.1007/s43681-021-00040-9>
- Hobuss, J. (2009). *Virtude e mediedade em Aristóteles*. Pelotas: Editora da UFPel.

- Hupe, P. (2013). Dimensions of discretion: specifying the object of street-level bureaucracy research. *dms – der moderne staat – Zeitschrift für Public Policy, Recht und Management*, 6(2), 425-440.
- Japiassú, H. (2006). *O sonho transdisciplinar e as razões da filosofia*. Rio de Janeiro: Imago.
- Japiassú, H & Marcondes, D. (2006). *Dicionário básico de filosofia*. Rio de Janeiro: Zahar.
- Kant, I. (2006). *Antropologia de um ponto de vista pragmático*. São Paulo: Iluminuras.
- Kant, I. (2012). *Immanuel Kant: Textos seletos*. Petrópolis: Vozes.
- Kant, I. (2019). *Fundamentação da metafísica dos costumes*. Lisboa: Edições 70.
- Kaufman, D. (2016). Inteligência artificial: questões éticas a serem enfrentadas. *IX Simpósio Nacional ABCiber: cibercultura, democracia e liberdade no Brasil*. PUC/SP 8 a 10 de dezembro.
- Lotta, G. & Santiago, A. Autonomia e discricionariedade: matizando conceitos-chave para o estudo de Burocracia. *Revista Brasileira de Informação Bibliográfica em Ciências Sociais*, 83, 21-42.
- Maas, J. (2022). A neo-republican critique of AI ethics. *Journal of Responsible Technology*, 9. <https://doi.org/10.1016/j.jrt.2021.100022>
- Meyer-Sahling, J.-H. & Mikkelsen, K. S. (2022). Codes of Ethics, Disciplinary Codes, and the Effectiveness of Anti-Corruption Frameworks: Evidence from a Survey of Civil Servants in Poland. *Review of Public Personnel Administration*, 42(1), 142–164. <https://doi.org/10.1177/0734371X20949420>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Munn, L. (2023). The uselessness of AI ethics. *AI Ethics*, 3, 869–877. <https://doi.org/10.1007/s43681-022-00209-w>
- OCDE. (2024). OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. *OECD Legal Instruments*. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#mainText>
- Oldenhof, L., Postma, J. & Putters, K. (2014). On justification work: How compromising enables public managers to deal with conflicting values. *Public Administration Review*, 74(1), 52-63. DOI:10.1111/puar.12153
- European Parliament. (2023). *DRAFT Compromise Amendments on the Draft Report*. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf
- Piteira, M., Aparicio, M. & Costa, C. J. (2019). A ética na inteligência artificial: Desafios. In A. Rocha, I. Pedrosa, M. P. Cota & R. Goncalves.

- (Ed.). 14th *Iberian Conference on Information Systems and Technologies* (CISTI). Coimbra: IEEE.
- Rochel, J. & Evéquoz, F. (2021). Getting into the engine room: a blueprint to investigate the shadowy steps of AI ethics. *AI & Society*, 36, 609–622. <https://doi.org/10.1007/s00146-020-01069-w>
- Santos, L. S., Leal, F. G., Serafim, M. C. & Moraes, M. C. B. (2018). Values and public administration: A discussion on rationality and parenthetical attitude. *Revista de Administração Mackenzie*, 19(3). DOI:10.1590/1678-6971/eRAMG170136
- Siau, K. & Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal of Database Management*, 31(2). DOI: 10.4018/JDM.2020040105
- Sousa, W. G., Melo, E. R., Bermejo, P. H., Farias, R. A. & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, 36(4). <https://doi.org/10.1016/j.giq.2019.07.004>
- Steden, R. (2020). Blind spots in public ethics and integrity research: what public administration scholars can learn from Aristotle. *Public Integrity*, 33(3), 236-244. <https://doi.org/10.1080/10999922.2020.1714412>
- Straub, V. J., Morgan, D., Brigh, J. & Margetts, H. (2023). Artificial intelligence in government: concepts, standards, and a unified framework. *Government Information Quarterly*, 40(4). <https://doi.org/10.1016/j.giq.2023.101881>
- Thomann, E., Engen, N. & Tummers, L. (2018). The Necessity of Discretion: A Behavioral Evaluation of Bottom-Up Implementation Theory. *Journal of Public Administration Research and Theory*, 28(4), 583–601. <https://doi.org/10.1093/jopart/muy024>
- Valle-Cruz, D., Gil-Garcia, J. R. & Sandoval-Almazan, R. (2024). Artificial intelligence algorithms and applications in the public sector: a systematic literature review based on the PRISMA approach. (p. 8-26) In Y. Charalabidis, R. Medaglia & C. Noordt. *Research handbook on public management and artificial intelligence*. Edward Elgar Publishing: Cheltenham. <http://dx.doi.org/10.4337/9781802207347>
- Waelen, R. (2022). Why AI Ethics is a critical theory. *Philosophy & Technology*, 35(9). <https://doi.org/10.1007/s13347-022-00507-5>
- Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2), 1-37. DOI:10.2478/jagi-2019-0002
- Wernaart, B. F. W. (2022). An introduction to moral design and technology. (p. 13-23) In B. F. W. WERNAART (ed). *Moral design and technology*. Wageningen Academic

Publishers. DOI: 10.3920/978-90-8686-922-0_1

Sandro Trescastro Bergue (sandro.bergue@gmail.com)* trabalhou na introdução, referencial teórico, discussão dos argumentos e nas considerações finais.

Data de Submissão: 11/04/2024 Data de Aprovação: 20/06/2024.

Editor-Chefe: Diogo Henrique Helal.

Editor Adjunto: Bruno Melo Moura.

Esta obra está licenciada sob uma Licença Creative Commons Atribuição 4.0 Internacional (CC BY NC 4.0). Esta licença permite que outros distribuam, remixem, adaptem e criem a partir do trabalho, para fins não comerciais, desde que lhe atribuam o devido crédito pela criação original. Texto da licença: https://creativecommons.org/licenses/by-nc/4.0/deed.pt_BR.