

QUESTÕES ÉTICAS NA ORGANIZAÇÃO DO CONHECIMENTO: DOS SISTEMAS DE RECOMENDAÇÃO À INTELIGÊNCIA ARTIFICIAL GENERATIVA

SARA DE CARVALHO*

MARIA MANUEL BORGES**

Resumo: *Este trabalho apresenta uma sistematização das questões éticas na Organização do Conhecimento tendo como eixos paradigmáticos o estudo dos Sistemas de Recomendação e a Inteligência Artificial Generativa. Partindo de uma revisão crítica da literatura visa refletir sobre as implicações e as contribuições que a Organização do Conhecimento, enquanto campo de estudo, pode trazer ao debate ético transdisciplinar contemporâneo. Os resultados apontam para o facto de a Organização do Conhecimento assumir um papel de fundamento/estrutura crítica e os seus pressupostos serem a condição para a operacionalização dos princípios éticos, no paradigma da Inteligência Artificial centrada no humano.*

Palavras-chave: *Ética; Sistemas de Recomendação; Inteligência Artificial Generativa; Organização do Conhecimento.*

Abstract: *This paper presents a systematisation of ethical issues in Knowledge Organisation, using the study of Recommendation Systems and Generative Artificial Intelligence as paradigmatic axes. Based on a critical review of the literature, it aims to reflect on the implications and contributions that Knowledge Organisation, as a field of study, can bring to the contemporary transdisciplinary ethical debate. The results point to the fact that Knowledge Organisation takes on the role of a critical foundation/structure and its presuppositions are the condition for the operationalisation of ethical principles in the paradigm of human-centred Artificial Intelligence.*

Keywords: *Ethics; Recommender Systems; Generative Artificial Intelligence; Knowledge Organisation.*

1. MAPEAMENTO DOS PRINCÍPIOS ÉTICOS NA INTELIGÊNCIA ARTIFICIAL

A poderosa cosmovisão que a Inteligência Artificial (IA) representa reflete-se também na abundância de propostas e divulgação de princípios e orientações para uma «Inteligência Artificial Ética», oriundos dos mais diversos quadrantes, de organizações governamentais e intergovernamentais a instituições de ensino e investigação, empresas privadas, associações profissionais, entre outras. Neste sentido, a investigação demonstra o consenso sobre a necessidade de colocar a Ética no centro das

* Universidade de Coimbra, Faculdade de Letras, Centro de Estudos Interdisciplinares – CEIS20. Email: uc1997000129@student.uc.pt. ORCID: <https://orcid.org/0000-0001-6868-6801>.

** Universidade de Coimbra, Faculdade de Letras, Centro de Estudos Interdisciplinares – CEIS20. Email: mmb@fl.uc.pt. ORCID: <https://orcid.org/0000-0002-7755-6168>.

preocupações, evidenciando-se, contudo, muitos debates sobre o que constitui uma IA Ética, quais os seus fundamentos, requisitos, normas e melhores práticas.

Jobin, Ienca e Vayena (2019) mapearam, numa revisão sistemática da literatura, o *corpus* de princípios e orientações sobre IA Ética e identificaram, nos oitenta e quatro documentos criteriosamente selecionados, onze grupos temáticos, demonstrando a convergência global em cinco princípios: transparência, justiça e equidade, não-maleficência, responsabilidade e privacidade. Não obstante, admitem divergências substantivas no que se refere a quatro dimensões: «(1) como são interpretados os princípios éticos; (2) porque são considerados importantes; (3) a que questão, domínio ou atores se referem; e (4) como devem ser postos em prática» (Jobin, Ienca e Vayena 2019, p. 396).

De igual relevância é o estudo que se propõe «contribuir para colmatar a lacuna entre princípios e práticas, construindo uma tipologia que possa ajudar os programadores com preocupações práticas a aplicar a Ética em cada fase da cadeia de desenvolvimento» (Morley et al. 2020, p. 2141). Esta investigação procura, primeiramente, construir tipologias que façam corresponder as ferramentas e métodos identificados aos princípios éticos e, numa segunda fase, exemplificar os instrumentos e métodos, bem como as empresas e pessoas, enquadrando-as na respetiva tipologia.

Merecem também destaque as conclusões da AI4People, uma iniciativa do Atomium – European Institute for Science, Media and Democracy (EISMD), liderada por Luciano Floridi e concebida para pensar as bases de uma «Boa Sociedade da IA». Nestas conclusões, no âmbito do quadro ético, ao equacionarem-se as oportunidades, os riscos e as recomendações, estabelecem-se os princípios que devem estar na génese do seu desenvolvimento. A análise recaiu sobre seis documentos paradigmáticos, dos quais resultaram quatro princípios, a saber: «beneficência: promover o bem-estar, preservar a dignidade e sustentar o planeta»; «não-maleficência: privacidade, segurança e “preservação de capacidades”»; «autonomia: o poder de decidir (se se deve decidir)»; «justiça: promover a prosperidade e preservar a solidariedade» (Floridi et al. 2018, pp. 697-698). Tais princípios foram transpostos da Bioética e adaptam-se aos desafios colocados pela Ética Digital. A estes, o grupo liderado por Floridi acrescenta um quinto princípio, o da «explicabilidade: possibilitar os outros princípios através da inteligibilidade e responsabilidade» (Floridi et al. 2018, p. 698; Floridi e Cowsls 2019).

Nesta senda, o Grupo de Peritos de Alto Nível da Inteligência Artificial da Comissão Europeia lança as *Orientações Éticas para uma IA de Confiança*, identificando os seguintes princípios éticos como as bases de uma IA de confiança: «respeito pela autonomia humana, prevenção de danos, equidade e explicabilidade» (GPAN IA 2019, p. 10).

No entanto, este debate centrado nos princípios e declaração de valores não deixa de ser alvo de críticas e objeções, sobretudo, no que concerne aos seus atores.

Considerando que grande parte destas iniciativas é promovida pela própria indústria, os seus objetivos e agendas são questionáveis porque não são isentos. Em última instância, centrar a discussão em questões abstratas, que não se traduzem em recomendações concretas, nem refletem as tensões normativas e políticas dos próprios princípios que enumeram, pode ser encarado como um mero subterfúgio, com o intuito de protelar a regulamentação do setor. Por outro lado, outros autores destacam que «qualquer compromisso alcançado até agora em torno de princípios fundamentais para a Ética da IA não reflete um consenso significativo sobre uma direção prática comum para o “bom” desenvolvimento e governança da IA» (Mittelstadt 2019, p. 6).

Conscientes destas fragilidades, e do ponto de vista do mapeamento, destaca-se a perspectiva de Fjeld et al. (2020). Através de um estudo comparativo exaustivo, avaliam as diversas propostas de Ética na IA e apresentam um cruzamento dos temas e princípios com os direitos humanos consagrados internacionalmente. O estudo estrutura-se em oito temas: privacidade, responsabilidade, segurança e proteção, transparência e explicabilidade, equidade e não discriminação, controlo humano da tecnologia, responsabilidade profissional e promoção de valores. O relatório contempla, além disso, o desdobramento de cada tema nos princípios que lhe estão subjacentes, com visualização dos dados numa proposta de «retrato de alto nível do estado atual do pensamento na governação da IA».

Por seu turno, Ryan e Stahl (2021), partindo dos grupos temáticos elencados no estudo de Jobin, Ienca e Vayena (2019), procuram fazer uma descrição e compilação pormenorizada do conteúdo normativo que é abrangido pelos princípios, considerando a iniciativa fundamental não apenas do ponto de vista do aprofundamento teórico, mas também como um contributo para a criação de orientações práticas implementáveis para os criadores e utilizadores de IA.

A *Recomendação sobre a Ética na Inteligência Artificial* da UNESCO identifica os seguintes princípios: proporcionalidade e não causar dano; segurança e proteção; justiça e não discriminação; sustentabilidade; direito à privacidade e proteção de dados; transparência e explicabilidade; supervisão humana e determinação; responsabilidade e prestação de contas; consciencialização e alfabetização; governança e colaboração adaptáveis e com múltiplas partes interessadas (UNESCO 2021, pp. 20-23).

Por fim, é de salientar o estudo de Laine, Minkkinen e Mäntymäki (2024), pela sua virtuosa conceção pragmática, uma vez que, através de uma revisão sistemática da literatura, expõe os princípios éticos aplicados às auditorias dos sistemas de IA, destacando-se os seguintes temas: justiça, equidade, não-maleficência, responsabilidade, privacidade, confiança, beneficência e liberdade/autonomia.

Tendo por base este *corpus* de investigação sobre os princípios éticos na IA, e sabendo «que existe um grande e espinhoso fosso entre a articulação destes conceitos de alto nível e a sua concretização efetiva no mundo real» (Fjeld et al. 2020,

p. 66), este trabalho procura indagar qual a sua relação com a Organização do Conhecimento (OC), concretamente, no universo dos Sistemas de Recomendação (SR) e no domínio da Inteligência Artificial Generativa (IAGen). O percurso investigativo proposto congrega, neste sentido, a organização intelectual ou cognitiva e a própria organização social do conhecimento (Hjørland 2008; Hjørland 2016; Hjørland 2023).

Depois de percorrido o panorama dos princípios éticos na Inteligência Artificial, aprofundam-se as noções de SR e IAGen no contexto das suas implicações éticas na OC. Assim, este percurso conduz ao exercício dialógico, num duplo movimento de abordagem epistemológica. Por um lado, empreende-se o levantamento das problemáticas que estes universos convocam nos sistemas e processos de OC; por outro, reflete-se sobre a forma como os postulados teórico-metodológicos da própria OC podem ser decisivos para uma visão crítica de «como o conhecimento é socialmente organizado» (Hjørland 2008, p. 87).

Face a este contexto, urge questionar: qual o papel da OC no pensamento ético-reflexivo da contemporaneidade? Que desafios enfrenta a comunidade de OC e que oportunidades se vislumbram?

2. ABORDAGEM METODOLÓGICA

Para responder à pergunta de investigação que norteia este trabalho — que questões éticas se apresentam à OC ao estudar os SR e a IAGen — adotou-se uma abordagem metodológica qualitativa, de natureza exploratória, descritiva/interpretativa. Como objetivos específicos pretende-se identificar os principais desafios éticos na OC, avaliar as implicações éticas dos Sistemas de Recomendação e explorar os desafios éticos da Inteligência Artificial Generativa.

O estado da arte é suportado pelos resultados da pesquisa bibliográfica na Web of Science, Library & Information Science Source, b-on e Scopus. Num primeiro momento, os artigos recuperados foram analisados — título, palavras-chave e resumo. Posteriormente, os dados foram sistematizados, eliminando-se os duplicados e os artigos não pertinentes. À medida que se aprofundaram as leituras, procedeu-se à inclusão de pistas bibliográficas que se revelaram frutíferas para a investigação. Por fim, procedeu-se à análise crítica e sistematização, no sentido de perceber as interrogações éticas e as metamorfoses nas práticas e resultados do universo do conhecimento.

3. SISTEMAS DE RECOMENDAÇÃO: PROBLEMÁTICAS VERSUS DESCONSTRUÇÃO

Ao longo da investigação, torna-se evidente que os autores apresentam diferentes conceptualizações quando descrevem e analisam os SR, primando, contudo, pelas abordagens técnicas. Por conseguinte, são escassos os estudos que refletem sobre os princípios e desafios éticos que se desenham no ecossistema informacional dos SR e

menos ainda os que investigam as suas implicações na OC. A abundância de recursos confere aos SR um papel de grande utilidade, já que «fornecem uma visão personalizada desses espaços, priorizando itens que provavelmente serão do interesse do utilizador» (Burke, Felfernig e Göker 2011, p. 13). Por outro lado, no que diz respeito às técnicas de recomendação, estas desdobram-se em quatro grandes categorias: sistemas de filtragem colaborativa, baseada em conteúdo, em conhecimento e híbridos.

Começando pelos *SR de filtragem colaborativa*, estes «agregam dados sobre os hábitos ou preferências de compra dos clientes e fazem recomendações a outros utilizadores com base na semelhança dos padrões gerais de compra» (Burke 2000, p. 1). Por sua vez, os *SR baseados em conteúdo* induzem «um classificador capaz de distinguir entre os itens que provavelmente interessam ao utilizador e os que provavelmente não interessam» (Burke 2000, p. 1). Quanto aos *sistemas baseados em conhecimento*, como a própria designação indica, partem do conhecimento sobre os utilizadores e produtos «para gerar a recomendação, raciocinando sobre os produtos que satisfazem os requisitos do utilizador» (Burke 2000, p. 1). Finalmente, os *SR híbridos* resultam da combinação das diferentes técnicas.

Independentemente da evolução das tipologias descritas, oferecem vantagens e inconvenientes que devem ser equacionados de acordo com o contexto e objetivos específicos, sendo inquestionável a sua relevância crescente nos processos de apoio à decisão humana. Além disso, otimizam a experiência do utilizador no universo *Web* e são considerados como muito «pertinentes no domínio da interação humano-computador (IHC), pois melhoram a eficiência da relação através de mecanismos de *feedback*» (Xia et al. 2024, p. 196914).

Ao mesmo tempo, assistimos ao disseminar da sua atuação e, se inicialmente o seu domínio de excelência foram as plataformas de comércio eletrónico, hoje em dia estão plasmados nas redes sociais, motores de pesquisa, plataformas das mais diversas áreas, nomeadamente na saúde e educação. No cenário de sugestões produzidas por «raciocínio» algorítmico urge repensar o processo de assimilação deste seu «papel constitutivo ou performativo na ordenação do mundo em nosso nome» (Beer 2017, p. 6).

Na literatura técnica de avaliação dos SR têm existido muitos esforços para repensar as suas métricas. Se a maximização da acurácia¹ nas recomendações feitas marcou as primeiras investigações enquanto fator decisivo, a evolução dos estudos, e das suas implicações, trouxe outros fatores à colação, como a diversidade, a serendipidade, a novidade e a equidade.

Abdollahpouri, Burke e Mobasher (2017) salientam também a forma como, em muitos ensaios, a análise se centrou nos utilizadores, desvalorizando o papel fundamental

¹ Utilizamos o termo acurácia no sentido de proporção de previsões corretas em relação ao total de previsões feitas.

dos outros intervenientes/agentes, designadamente: utilizadores, fornecedores, sistemas/ plataformas. Outras investigações advogam o papel essencial de considerar a própria sociedade como um dos agentes neste universo (Jannach e Zanker 2022). Jannach e Adomavicius (2016) procedem ao exercício de enumerar os objetivos das recomendações, exemplificando a diferença entre os pontos de vista do utilizador e do fornecedor, sendo pertinente verificar que os objetivos e os «interesses» em causa são distintos e, por vezes, até conflitantes, quer do ponto de vista pragmático quer ético.

Outras investigações merecem destaque, como a *Estrutura geral de recomendações éticas centradas no utilizador* de Paraschakis, em que são construídas propostas de ferramentas que o próprio utilizador pode gerir e ajustar, centradas em quatro blocos conceptuais: «recolha e filtragem de dados»; «publicação e anonimização de dados»; «opacidade algorítmica, enviesamento e manipulação de comportamentos»; e «experiências online» (Paraschakis 2017).

Noutro contexto disciplinar, e embora as questões éticas possam ser consubstanciadas através de múltiplos paradigmas, alguns investigadores focam-se na análise das ações e consequências. Milano, Taddeo e Floridi (2020; 2021) analisam os impactos éticos dos SR numa dupla dimensão: de que modo podem afetar negativamente ou violar os direitos das partes interessadas e em que medida isso constitui um impacto imediato ou futuro (Milano, Taddeo e Floridi 2020, pp. 959-960). Colhendo os contributos transdisciplinares, os autores identificam seis áreas de preocupação ética nos SR, respetivamente: conteúdo inadequado, privacidade, autonomia, identidade pessoal, opacidade e equidade (Milano, Taddeo e Floridi 2020, pp. 960-964).

Através deste prisma, refletir sobre os SR levanta questões de grande complexidade. Em primeiro lugar, importa acentuar o movimento a que assistimos, ou seja, à medida que os universos algorítmicos e de *big data* avançam, os SR tornam-se «omnipresentes» no quotidiano dos utilizadores, uma vez que «moldam as [suas] preferências [...] e orientam as suas escolhas, tanto a nível individual como social. Este impacto é significativo e merece um escrutínio ético» (Milano, Taddeo e Floridi 2020, p. 957).

Questiona-se: que parâmetros estão presentes nestes sistemas, se o que os rege são espaços de «semelhança», de «escolhas passadas», de movimentos que identificam «padrões» e se nutrem de previsões de comportamentos e preferências dos utilizadores, segmentando e/ou ampliando perfis? Nestes processos discursivos, de predições pré-programadas e estigmatizantes — que consequentemente anulam a diferença, as vozes minoritárias, a importância do contexto e evolução da agência humana —, que tipo de conhecimento da realidade se está a construir?

Que fluxo informacional se materializa nestes percursos? E como se gerem os seus fenómenos e metamorfoses, especialmente as «câmaras de eco» e as «bolhas informacionais», que arrastam consigo fenómenos de polarização e desinformação e

que tanto limitam a exposição a novos conhecimentos, pontos de vista, ideias? Nesta mundividência, surge o discurso sobre as desigualdades (Zajko 2022) e os vieses discriminatórios (Broughton 2019) que confluem em todo este universo de rede, onde se interrogam até os fundamentos da «democracia participativa» (Vicente 2023).

Aprofundando a análise, estas temáticas revelam pistas de investigação pertinentes e operacionalizáveis do ponto de vista da OC. Que implicações tem este ambiente informacional? Desde logo, há que questionar a própria preparação dos dados, o que constitui as bases de dados, da amálgama do «rasto» deixado pelo utilizador até à recolha sistemática de informação; o caminho e as operações que se executam para extrair conclusões; as consequências das previsões feitas; os critérios utilizados para definir o que é relevante; a tecnicidade do algoritmo, que induz uma falsa «garantia» de imparcialidade; o seu impregnar na vida quotidiana, modelando o próprio utilizador e o sujeito do conhecimento; a forma como a interação é já uma retroalimentação dos sistemas.

Tendo por base o conceito de algoritmo, Gillespie (2014) chama a atenção para o fenómeno que designa de *algoritmos de relevância pública*, realçando a necessidade de se questionar não só a sua representação como uma abordagem ao conhecimento particular, mas também as suas implicações. A sua reflexão desdobra-se em seis temas específicos, nomeadamente: «padrões de inclusão», «ciclos de antecipação», «avaliação de relevância», «promessa de objetividade algorítmica», «emaranhamento com a prática» e «produção de públicos calculados» (Gillespie 2014, p. 168).

É nesta confluência que os caminhos da OC e da Ética se encontram e permitem exemplificar problemáticas e desafios ao estudar os SR. A sua atuação atinge o âmago da relação cognoscitiva, e as questões da produção, organização, validação e disseminação do conhecimento, numa sociedade mediada por estas instâncias algorítmicas, passam a ter uma dimensão que é, em si, profundamente ética.

4. INTELIGÊNCIA ARTIFICIAL GENERATIVA E INTEGRIDADE NA INVESTIGAÇÃO CIENTÍFICA

A IAGen tem vindo a povoar a investigação científica e, na opinião predominante, de forma disruptiva (Vasconcelos e Marušić 2025), ampliando, ao mesmo tempo, as questões da integridade em todo o processo científico de investigação. No seu cerne, a IAGen pode ser definida como uma tecnologia de IA que:

gera conteúdo de forma automática como resposta a comandos escritos em interfaces de conversação em linguagem natural. [...] na verdade, produz novo conteúdo. Esses conteúdos podem aparecer em formatos que compreendem todas as representações simbólicas do pensamento humano: textos escritos em linguagem natural, imagens [...], vídeos, música e código de software (UNESCO 2023, p. 8).

De facto, a sua expansão atingiu níveis impressionantes na rapidez de adesão e também na diversidade de públicos que conquistou, apresentando índices de utilização nunca atingidos por qualquer outra tecnologia (Bobula 2024). Modelos de linguagem de grande escala (LLM), como o ChatGPT da OpenAI, tornaram-se acessíveis ao público em geral, pois a sua utilização/ativação acontece em linguagem natural, eliminando as barreiras técnicas e, nesta fase, económicas, que outras tecnologias apresentam. O modelo de base é o *Generative Pre-trained Transformer* (GPT), alicerçado na arquitetura *Transformer* (Vaswani et al. 2017), permitindo a criação de conteúdo fluído e plausível.

Uma análise cuidada revela que os LLM identificam «a probabilidade de ocorrência de uma palavra (ou conjunto de palavras) dada a presença de outros padrões de linguagem numa passagem de texto» (Bail 2024, p. 2). Contudo, o salto do reconhecimento de padrões da aprendizagem automática para a geração de novos conteúdos, com a consequente mimetização de conversação humana, tem suscitado as mais diversas questões. Neste sentido, os autores questionam se é de facto um discurso novo e criativo ou se os LLM não passam de «papagaios estocásticos» (Bender et al. 2021). Apesar de esta «recombinação» poder aparentar uma incursão no domínio da criatividade, a maioria dos autores considera que a «motivação», o «dar sentido» e a «compreensão do problema» ainda pertencem à esfera eminentemente humana (Haase e Hanel 2023).

Independentemente da perspectiva adotada, a IAGen está a disseminar-se nas diversas fases do processo científico, nomeadamente, na análise de dados, formulação de hipóteses, modelação e escrita científica, ensino e aprendizagem. Existem vários inquéritos a investigadores que nos permitem desenhar o panorama das oportunidades, mas também dos desafios e riscos que a IA, e a IAGen em particular, representam para a investigação, apresentando-se também os cenários prospetivos.

O inquérito do Conselho Europeu de Investigação (ERC) procura mostrar de que forma os investigadores utilizam a IA, e a IAGen, mapeando a sua evolução. Quando lhes foi solicitado que perspetivassem o papel da IA na investigação científica até 2030, a «maioria afirmou que a IA servirá de ferramenta de apoio, desempenhará um papel fulcral ou essencial e, em alguns casos, acelerará, revolucionará ou transformará determinados elementos do processo científico» (European Research Council 2023, p. 6). A maioria das respostas dos inquiridos vai no sentido de considerarem que a IA desempenhará o papel de «assistente» ou «suporte» aos cientistas, ao longo do processo científico (European Research Council 2023, p. 8).

No que diz respeito, especificamente, à IAGen, nos desafios e riscos identificados destacam-se, por ordem decrescente de importância: em primeiro lugar, «difundir informações falsas ou conhecimentos inexatos»; «afetar a integridade da investigação»; «conduzir a uma dependência excessiva da IA»; «suscitar questões de direitos de propriedade intelectual» (European Research Council 2023, p. 9).

Simultaneamente, a *Nature* inquiriu também os cientistas e as respostas têm similaridades, se compararmos os possíveis impactos da IAGen com as problemáticas identificadas no Relatório do ERC. Desde logo, 68% revelam-se preocupados com a «proliferação de desinformação»; igual número considerou que «facilitaria o plágio» e 66% estavam preocupados com a «introdução de erros ou imprecisões nos trabalhos de investigação» (Van Noorden e Perkel 2023, p. 673). Não obstante, uma minoria de 28% afirma «utilizar LLM todos os dias ou mais do que uma vez por semana» (Van Noorden e Perkel 2023, p. 673).

Numa perspetiva diferente, mas igualmente pertinente, a editora Wiley apresenta o seu estudo *ExplanAltions*, onde identifica as cinco principais formas de utilização da IAGen por parte dos investigadores/inquiridos, respetivamente: «auxílio na tradução» 40%; «revisão e correção de textos académicos e trabalhos para publicação» 38%; «brainstorming/ideias» 38%; «revisão de grandes quantidades de informação» 26%; «descobrir a investigação mais recente que é relevante» 24% (Wiley 2025, p. 11).

Este estudo indica os casos de utilização sistematizando-os nas diversas etapas do processo de investigação. Nas fases de «determinar o que investigar» e da própria «realização da investigação», os inquiridos consideram estar prontos para utilizar a AI, sobretudo, na análise de grandes dados estruturados e não estruturados. A fase da preparação para publicação é a área onde a utilização da IA é já mais efetiva, concretamente nas questões de autoverificação e no auxílio à redação dos seus manuscritos. Na fase da revisão por pares os autores valorizam a avaliação humana e, por fim, na divulgação, partilha e pós-publicação sobressai a relevância da IA na geração de resumos e na vertente de tornar a informação mais acessível.

Os três estudos referidos permitem contextualizar o ponto de situação na utilização e implementação da IA junto das comunidades científicas e, ao mesmo tempo, evidenciam a forma como os investigadores perspetivam o futuro próximo.

Importa perceber o que está no centro do debate, isto é: «a forma como os sistemas de IA podem ser alinhados com os objetivos de manter a integridade e a confiança na ciência» (Vasconcelos e Marušić 2025, p. 1924). O *Código Europeu de Conduta para a Integridade da Investigação* elege como princípios fundamentais a fiabilidade, a honestidade, o respeito e a responsabilidade (All European Academies 2023, p. 5). E considera que «a integridade da investigação é crucial para a preservação da confiança no sistema científico e nos seus resultados» (All European Academies 2023, p. 3). Dada a vastidão de questões que a existência da IAGen coloca à integridade da investigação, os autores desdobram-se em propostas, na tentativa de encontrar caminhos que possam ser operacionalizáveis e que garantam, de facto, essa integridade.

Vasconcelos e Marušić consideram que «os quadros de investigação devem agora incluir o reforço da agência humana na apresentação, condução, comunicação e revisão da ciência» (2025, p. 1927).

Eacersall estrutura os desafios da AIGen de forma muito consistente e sistemática, ressaltando-se, para o nosso estudo, os seguintes: «qualidade dos resultados: alucinações, precisão, referências, atualidade» (2024, p. 2); proteção de dados, com as consequentes questões da privacidade e autoria; direitos autorais e propriedade intelectual; transparência; acesso e viés; «impactos sociais: [...] desafios em torno da equidade social nas áreas de acesso, enviesamento e emprego» (Eacersall 2024, p. 3). Nesta senda, as *Living Guidelines on the Responsible Use of Generative AI in Research* da Comissão Europeia (2025) apresentam um conjunto de recomendações não só para os investigadores, mas também para as Organizações e Instituições de Financiamento que pode ser da maior utilidade.

Ora, estes desafios relacionam-se, acima de tudo, com os princípios éticos que apresentámos na primeira parte, sobretudo, os elencados pela iniciativa da AI4People, e cruzam-se com o paradigma da OC, no seu sentido amplo. A disseminação da IA, e da IAGen em particular, nas mundividências contemporâneas torna fundamental o questionamento dos pressupostos teóricos, epistemológicos e ontológicos que subjazem à OC, obrigando a um movimento de reflexão que interroge os critérios e fundamentos a partir dos quais os Processos de Organização do Conhecimento (POC) e os Sistemas de Organização do Conhecimento (SOC) são pensados (Ibekwe-SanJuan e Bowker 2017).

Circunscrevendo o pensamento, a complexidade das questões-chave é de natureza eminentemente epistemológica. Que relação epistémica se estabelece na IAGen? Que autoridade epistémica emerge nesta dimensão (Heersmink et al. 2024)? E como valorizar a geração de hipóteses geradas por IA? E quando esta se torna peça-chave no próprio processo de descoberta científica (Wang et al. 2023)? Que problemáticas se desenhem quando existe a possibilidade de gerar amostras simuladas por IA ou de utilizar dados sintéticos na investigação (Bail 2024; Argyle et al. 2023)? E como equacionar as questões da verdade, veracidade, rastreabilidade e reprodutibilidade nestes sistemas?

CONCLUSÕES

Através de uma abordagem exploratória, procurámos proceder ao levantamento dos grandes desafios e problemáticas éticos que se colocam no domínio da IA, evidenciando como as comunidades de OC são os interlocutores, por excelência, deste debate.

As temáticas dos «Sistemas de Recomendação» e «Inteligência Artificial Generativa» permitiram ilustrar a relação efetiva entre a Ética e a OC, concluindo-se que o fundamento da possibilidade de operacionalização dos princípios éticos, para um uso responsável e benéfico da IA, está nos pressupostos epistemológicos que tecem a OC.

No entanto, reconhecemos que o universo ético convocado no terreno da IA implica uma metodologia hermenêutica de reconfiguração das próprias teorias do conhecimento, da relação do eu com o mundo, da liberdade, dignidade, autonomia humana, responsabilidade e, em última instância, da Democracia participativa. Quer o domínio técnico quer o social têm no seu âmago o desafio seminal que é esta dupla face Ética e de OC. Deste modo, a possibilidade de «florescimento» da agência humana, na encruzilhada da IA, dependerá exclusivamente do projeto humano que todos quisermos construir.

REFERÊNCIAS

- ABDOLLAHPOURI, H., R. BURKE, e B. MOBASHER, 2017. Recommender Systems as Multistakeholder Environments. Em: *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 9-12 July, 2017, Bratislava, Slovakia. Nova Iorque: ACM, pp. 347-348. DOI: <https://doi.org/10.1145/3079628.3079657>.
- ALL EUROPEAN ACADEMIES, 2023. *The European Code of Conduct for Research Integrity* [Em linha]. Rev. ed. [consult. 2025-07-08]. Disponível em: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/european-code-of-conduct-for-research-integrity_horizon_en.pdf.
- ARGYLE, L. P., et al., 2023. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*. **31**(3), 337-351. DOI: <https://doi.org/10.1017/pan.2023.2>.
- BAIL, C. A., 2024. Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*. **121**(21). DOI: <https://doi.org/10.1073/pnas.2314021121>.
- BEER, D., 2017. The social power of algorithms. *Information, Communication & Society*. **20**(1), 1-13. DOI: <https://doi.org/10.1080/1369118X.2016.1216147>.
- BENDER, E. M., et al., 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. Em: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, 3-10 March 2021, Canada*. Nova Iorque: ACM, pp. 610-623. DOI: <https://doi.org/10.1145/3442188.3445922>.
- BOBULA, M., 2024. Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education*. (30). DOI: <https://doi.org/10.47408/jldhe.vi30.1137>.
- BROUGHTON, V., 2019. The Respective Roles of Intellectual Creativity and Automation in Representing Diversity: Human and Machine Generated Bias. *Knowledge Organization*. **46**(8), 596-606. DOI: <https://doi.org/10.5771/0943-7444-2019-8-596>.
- BURKE, R., A. FELFERNIG, e M. H. GÖKER, 2011. Recommender Systems: An Overview. *AI Magazine*. **32**(3), 13-18. DOI: <https://doi.org/10.1609/aimag.v32i3.2361>.
- BURKE, R., 2000. Knowledge-based recommender systems. Em: *Encyclopedia of library and information science* [Em linha]. **69** (Supplement 32), 180-201 [consult. 2025-07-08]. Disponível em: https://www.researchgate.net/publication/2378325_Knowledge-Based_Recommender_Systems.
- COMISSÃO EUROPEIA, 2025. *Living Guidelines on the Responsible Use of Generative AI in Research* [Em linha] [consult. 2025-07-08]. Disponível em: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en.
- EACERSALL, D., 2024. *AI literacy: Principles for ethical generative artificial intelligence* [Em linha]. Human-AI Collaborative Knowledgebase for Education and Research (HACKER) [consult. 2025-07-08]. Disponível em: <https://research.usq.edu.au/item/z7461/ai-literacy-principles-for-ethical-generative-artificial-intelligence>.

- EUROPEAN RESEARCH COUNCIL, 2023. *Foresight: Use and Impact of Artificial Intelligence in the Scientific Process* [Em linha]. European Research Council [consult. 2025-07-08]. Disponível em: https://erc.europa.eu/sites/default/files/2023-12/AI_in_science.pdf.
- FJELD, J., et al., 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. DOI: <https://doi.org/10.2139/ssrn.3518482>.
- FLORIDI, L., e J. COWLS, 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*. **1**(1). DOI: <https://doi.org/10.1162/99608f92.8cd550d1>.
- FLORIDI, L., et al., 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*. **28**, 689-707. DOI: <https://doi.org/10.1007/s11023-018-9482-5>.
- GILLESPIE, T., 2014. The Relevance of Algorithms. Em: T. GILLESPIE, P. J. BOCZKOWSKI, e K. A. FOOT, eds. *Media Technologies*. Cambridge, MS: The MIT Press, pp. 167-194. ISBN 978-0-262-52537-4.
- GPAN IA – Grupo de Peritos de Alto Nível da Inteligência Artificial, 2019. *Orientações éticas para uma IA de Confiança* [Em linha]. Comissão Europeia [consult. 2025-07-08]. Disponível em: <https://digital-strategy.ec.europa.eu/pt/library/ethics-guidelines-trustworthy-ai>.
- HAASE, J., e P. H. P. HANEL, 2023. Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity*. **33**(3), 100066. DOI: <https://doi.org/10.1016/j.yjoc.2023.100066>.
- HEERSMINK, R., et al., 2024. A phenomenology and epistemology of large language models: transparency, trust, and trustworthiness. *Ethics and Information Technology*. **26**(3). DOI: <https://doi.org/10.1007/s10676-024-09777-3>.
- HJØRLAND, B., 2023. Education in Knowledge Organization (KO). *Knowledge Organization*. **50**(3), 160-181. DOI: <https://doi.org/10.5771/0943-7444-2023-3-160>.
- HJØRLAND, B., 2016. Knowledge Organization (KO). *Knowledge Organization*. **43**(6), 475-484. DOI: <https://doi.org/10.5771/0943-7444-2016-6-475>.
- HJØRLAND, B., 2008. What is Knowledge Organization (KO)? *Knowledge Organization*. **35**(2-3), 86-101. DOI: <https://doi.org/10.5771/0943-7444-2008-2-3-86>.
- IBEKWE-SANJUAN, F., e G. C. BOWKER, 2017. Implications of Big Data for Knowledge Organization. *Knowledge Organization*. **44**(3), 187-198. DOI: <https://doi.org/10.5771/0943-7444-2017-3-187>.
- JANNACH, D., e G. ADOMAVICIUS, 2016. Recommendations with a Purpose. Em: *Proceedings of the 10th ACM Conference on Recommender Systems, 15-19 September, Boston, Massachusetts, USA*. Nova Iorque: ACM, pp. 7-10. ISBN 978-1-4503-4035-9. DOI: <https://doi.org/10.1145/2959100.2959186>.
- JANNACH, D., e M. ZANKER, 2022. Value and Impact of Recommender Systems. Em: F. RICCI, L. ROKACH, e B. SHAPIRA, eds. *Recommender Systems Handbook*. Nova Iorque: Springer, pp. 519-546. ISBN 978-1-07-162196-7.
- JOBIN, A., M. IENCA, e E. VAYENA, 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*. **1**(9), 389-399. DOI: <https://doi.org/10.1038/s42256-019-0088-2>.
- LAINE, J., M. MINKKINEN, e M. MÄNTYMÄKI, 2024. Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information & Management*. **61**(5), 103969. DOI: <https://doi.org/10.1016/j.im.2024.103969>.
- MILANO, S., M. TADDEO, e L. FLORIDI, 2021. Ethical aspects of multi-stakeholder recommendation systems. *The Information Society*. **37**(1), 35-45. DOI: <https://doi.org/10.1080/01972243.2020.1832636>.
- MILANO, S., M. TADDEO, e L. FLORIDI, 2020. Recommender systems and their ethical challenges. *AI & Society*. **35**(4), 957-967. DOI: <https://doi.org/10.1007/s00146-020-00950-y>.
- MITTELSTADT, B., 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*. **1**(11). DOI: <https://doi.org/10.1038/s42256-019-0114-4>.

- MITTELSTADT, B., et al., 2016. The Ethics of algorithms: mapping the debate. *Big Data & Society*. 3(2), 1-21.
- MORLEY, J., et al., 2020. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*. 26(4), 2141-2168. DOI: <https://doi.org/10.1007/s11948-019-00165-5>.
- PARASCHAKIS, D., 2017. Towards an ethical recommendation framework. Em: *2017 11th International Conference on Research Challenges in Information Science (RCIS), 10-12 May 2017, Brighton, United Kingdom*. [S.l.]: IEEE, pp. 211-220. ISBN 978-1-5090-5476-3. DOI: <https://doi.org/10.1109/RCIS.2017.7956539>.
- RYAN, M., e B. C. STAHL, 2021. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*. 19(1), 61-86. DOI: <https://doi.org/10.1108/JICES-12-2019-0138>.
- UNESCO, 2023. *Guia para IA Generativa na Educação e na Pesquisa* [Em linha] [consult. 2025-07-08]. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000390241>.
- UNESCO, 2021. *Recomendações sobre a Ética na Inteligência Artificial* [Em linha] [consult. 2025-07-08]. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por.
- VAN NOORDEN, R., e J. M. PERKEL, 2023. AI and science: what 1,600 researchers think. *Nature*. 621(7980), 672-675. DOI: <https://doi.org/10.1038/d41586-023-02980-0>.
- VASCONCELOS, S., e A. MARUŠIĆ, 2025. Gen AI and research integrity: Where to now?: The integration of Generative AI in the research process challenges well-established definitions of research integrity. *EMBO Reports*. 26(8), 1923-1928. DOI: <https://doi.org/10.1038/s44319-025-00424-6>.
- VASWANI, A., et al., 2017. Attention Is All You Need. Em: *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*. Nova Iorque: Curran Associates Inc. DOI: <https://doi.org/10.48550/ARXIV.1706.03762>.
- VICENTE, P. N., 2023. *Os algoritmos e nós*. Lisboa: Fundação Francisco Manuel dos Santos. ISBN 978-989-9118-82-9.
- WANG, H., et al., 2023. Scientific discovery in the age of artificial intelligence. *Nature*. 620(7972), 47-60. DOI: <https://doi.org/10.1038/s41586-023-06221-2>.
- XIA, Z., et al., 2024. Contemporary Recommendation Systems on Big Data and Their Applications: A Survey. *IEEE Access*. 12, 196914-196928. DOI: <https://doi.org/10.1109/ACCESS.2024.3517492>.
- WILEY, 2025. *ExplanAItions: An AI Study by Wiley* [Em linha] [consult. 2025-07-08]. Disponível em: https://www.wiley.com/content/dam/wiley-dotcom/en/b2c/content-fragments/explanaitions-ai-report/pdfs/Wiley_ExplanAItions_AI_Study_February_2025.pdf.
- ZAJKO, M., 2022. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*. 16(3), e12962. DOI: <https://doi.org/10.1111/soc4.12962>.