26º ENANCIB
Encontro Nacional de Pesquisa e Pós-Graduação em Ciência da Informação
A Amazônia na sociedade global em rede: assimetrias e potencialidades informacionais, tecnológicas e ambientais
Belém, Pará, 09 a 13 de novembro de 2026
Universidade Federal do Pará

XXVI ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO - XXVI ENANCIB

ISSN 2177-3688

GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação

FLORIDI, CAPURRO E ARENDT EM DIÁLOGO: CONTRIBUIÇÕES ÉTICAS E CONCEITUAIS PARA O DESENVOLVIMENTO DA INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL

FLORIDI, CAPURRO AND ARENDT IN DIALOGUE: *ETHICAL AND CONCEPTUAL CONTRIBUTIONS TO EXPLAINABLE ARTIFICIAL INTELLIGENCE*

Aline Laureano Suave – Universidade Federal de São Carlos (UFSCar), Instituto de Informação para Inovação (i2i), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)
Rogério Sá Ramalho – Universidade Federal de São Carlos (UFSCar), Instituto de Informação para Inovação (i2i), Universidade Estadual de Londrina (UEL)
Marilde Terezinha Prado Santos – Universidade Federal de São Carlos (UFSCar), Instituto de Informação para Inovação (i2i)

Modalidade: Trabalho Completo

Resumo: O avanço da denominada Inteligência Artificial Explicável tem sido apresentado como uma possível resposta às crescentes demandas por transparência algorítmica. Contudo, persiste uma lacuna conceitual e teórica, uma vez que os pressupostos que definem o que constitui uma explicação adequada não são culturalmente neutros nem universalmente compartilhados, configurando um campo ainda pouco explorado pela Ciência da Informação. Nesse contexto, este trabalho tem como objetivo promover um diálogo interdisciplinar entre os estudos de Luciano Floridi, Rafael Capurro e Hannah Arendt, visando à construção de fundamentos éticos e conceituais para o desenvolvimento de sistemas de Inteligência Artificial Explicável em contextos culturalmente heterogêneos. Articula-se a macroética ontocêntrica de Floridi, fundamentada nos conceitos de infosfera, entropia informacional e agência moral, com a ética intercultural da informação de Capurro, ancorada na pluralidade dos mundos informacionais. Complementarmente, mobilizam-se as categorias de pluralidade, ação e julgamento de Arendt para incorporar a dimensão política da explicabilidade algorítmica. A pesquisa adota uma abordagem teórico-conceitual qualitativa, combinando análise conceitual filosófica, genealogia intelectual e hermenêutica. Como resultado, propõe-se um framework composto por quatro princípios orientadores para o desenvolvimento de sistemas XAI: integridade informacional, legibilidade culturalmente situada, promoção da capacidade de julgamento e redução de assimetrias de poder. Conclui-se que a articulação entre Floridi, Capurro e Arendt oferece uma contribuição

original da Ciência da Informação para o desenvolvimento de sistemas de IA mais transparentes, eticamente fundamentados e sensíveis à pluralidade sociocultural.
Palavras-chave: Ética da informação; Inteligência artificial explicável; Filosofia da informação; interculturalidade; justiça informacional.

Abstract: The advancement of so-called Explainable Artificial Intelligence has been presented as a possible response to the growing demands for algorithmic transparency. However, a conceptual and theoretical gap persists, as the assumptions that define what constitutes an adequate explanation are neither culturally neutral nor universally shared, making this a field that remains underexplored within Information Science. In this context, this study aims to promote an interdisciplinary dialogue among the works of Luciano Floridi, Rafael Capurro, and Hannah Arendt in order to establish ethical and conceptual foundations for the development of Explainable Artificial Intelligence systems in culturally heterogeneous contexts. Floridi's ontocentric macroethics, grounded in the concepts of infosphere, informational entropy, and moral agency, is articulated with Capurro's intercultural information ethics, rooted in the plurality of informational worlds. Additionally, Arendt's concepts of plurality, action, and judgment are mobilized to incorporate the political dimension of algorithmic explainability. The research adopts a qualitative theoretical-conceptual approach, combining philosophical conceptual analysis, intellectual genealogy, and hermeneutics. As a result, a framework composed of four guiding principles for the development of XAI systems is proposed: informational integrity, culturally situated legibility, promotion of the capacity for judgment, and reduction of power asymmetries. It is concluded that the articulation between Floridi, Capurro, and Arendt offers an original contribution from Information Science to the development of AI systems that are more transparent, ethically grounded, and sensitive to sociocultural plurality.
Keywords: Information Ethics; Explainable Artificial Intelligence; Philosophy of Information; Interculturality; Informational Justice.

1 INTRODUÇÃO

A Inteligência Artificial (IA) deixou de ser uma promessa futurista para tornar-se uma infraestrutura presente em processos cotidianos de seleção, classificação, recomendação e tomada de decisão. Sistemas automatizados influenciam a concessão de crédito, a distribuição de conteúdo em plataformas digitais, o reconhecimento facial e a seleção de candidatos em processos seletivos. Diante dessa crescente presença, ampliam-se as demandas sociais, jurídicas e éticas por transparência algorítmica. Nesse contexto, a Inteligência Artificial Explicável (Explainable Artificial Intelligence – XAI) emerge como um campo voltado ao desenvolvimento de sistemas capazes de fornecer explicações compreensíveis para suas decisões, buscando responder às críticas dirigidas aos algoritmos caracterizados como "caixas-pretas".

A XAI responde, em termos técnicos, ao problema dos modelos ditos "caixa-preta", sobretudo redes neurais profundas, cuja lógica interna não é diretamente interpretável por humanos. A literatura da área distingue, em linhas gerais, duas famílias de abordagem: modelos intrinsecamente interpretáveis, como árvores de decisão e modelos lineares esparsos, cuja própria estrutura já é legível; e métodos de explicação post-hoc, aplicados a modelos opacos após o treinamento, como a atribuição de importância a variáveis, explicações contrafactuais e mapas de saliência (Doshi-Velez; Kim, 2017). Essa distinção situa a XAI como um objeto de pesquisa relevante por si mesmo, e não apenas como pretexto para a discussão filosófica que se segue, na medida em que a escolha entre essas famílias de método já compromete, de saída, os tipos de explicação que podem ser oferecidos aos sujeitos afetados.

Entretanto, a própria noção de explicabilidade pressupõe critérios de inteligibilidade que raramente são problematizados. Como observa Lessig (2006), ao afirmar que “o código é a lei”, as arquiteturas técnicas incorporam escolhas normativas que influenciam comportamentos e formas de compreensão do mundo. Aplicada à XAI, essa perspectiva sugere que os critérios pelos quais uma explicação é considerada adequada não são neutros, mas dependem de formas culturalmente situadas de justificação, interpretação e confiança epistêmica. Uma explicação inteligível em determinado contexto cultural pode não produzir o mesmo efeito em outros contextos.

É nesse cenário que a Ciência da Informação (CI) encontra um problema filosófico relevante: como desenvolver sistemas explicáveis em contextos marcados pela pluralidade de formas de vida informacional? Em que medida as convergências e tensões entre as perspectivas de Luciano Floridi, Rafael Capurro e Hannah Arendt podem fundamentar princípios éticos orientadores da Inteligência Artificial Explicável? Partindo dessa questão, o presente trabalho investiga em que medida a macroética ontocêntrica de Luciano Floridi pode ser articulada à ética intercultural da informação de Rafael Capurro, em diálogo com as categorias de pluralidade, ação e julgamento de Hannah Arendt, para fundamentar princípios orientadores para o desenvolvimento ético de sistemas de Inteligência Artificial Explicável.

Do ponto de vista teórico, sustenta-se que a tensão entre o universalismo de Floridi e o particularismo intercultural de Capurro não constitui um obstáculo, mas uma condição produtiva para a construção de fundamentos éticos da XAI. Enquanto Floridi oferece princípios voltados à preservação da integridade da infosfera, Capurro evidencia a necessidade de

contextualizá-los nas diferentes formas de vida informacional. Arendt, por sua vez, acrescenta a dimensão política da explicabilidade ao enfatizar a pluralidade e a capacidade de julgamento dos sujeitos. A partir desse diálogo, propõe-se um framework composto por princípios orientadores para o desenvolvimento de sistemas XAI mais transparentes, responsáveis e sensíveis à diversidade sociocultural.

2 REFERENCIAL TEÓRICO

O trabalho exige mais do que a aproximação de autores: requer o mapeamento das tensões internas ao campo da Ética da informação e a identificação dos pontos de complementaridade produtiva entre tradições filosóficas distintas. As três subseções a seguir reconstroem, em ordem crescente de complexidade argumentativa, as contribuições de Floridi, Capurro e Arendt, preparando o terreno para a discussão comparativa da seção 4.

2.1 A MACROÉTICA ONTOCÊNTRICA DE FLORIDI: INFOSFERA, ENTROPIA E OBRIGAÇÃO DE TRANSPARÊNCIA

A Ética da Informação proposta por Luciano Floridi emerge de um diagnóstico preciso: as macroéticas clássicas, consequencialismo, deontologismo e contratualismo, são insuficientes para tratar os problemas morais gerados pelas tecnologias de informação e comunicação porque operam, todas elas, a partir de um pressuposto *action-oriented* e *anthropocentric* (Floridi, 1999, p. 39). Enquanto essas tradições concentram sua atenção no valor moral das ações humanas ou no agente que as pratica, os problemas informacionais contemporâneos frequentemente envolvem entidades não-humanas, ações virtuais de consequências prolixas e um desgaste progressivo da responsabilidade individual, como ocorre quando uma decisão de crédito, uma triagem de currículo ou uma seleção de conteúdo é inteiramente delegada a um algoritmo sem que nenhum sujeito humano específico possa ser responsabilizado pelo resultado.

Diante dessa carência, Floridi propõe a Ética da Informação (EI) como macroética genuinamente nova: *object-oriented* e *ontocentric*. Seu argumento central é que a EI "é primariamente uma ética do ser, e não da conduta ou do devir" (Floridi, 1999, p. 43, tradução nossa), o que significa que seu objeto de valor não é a ação humana, mas o estado da infosfera, definida como "o ambiente constituído pela totalidade de entidades informacionais, incluindo todos os agentes, seus processos, propriedades e relações mútuas" (Floridi, 1999, p. 44, tradução nossa). Exemplificando: por analogia com a termodinâmica, entropia designa o

grau de desordem ou perda de conteúdo em um sistema; Floridi trata informação e entropia como grandezas inversamente relacionadas, quanto mais uma região da infosfera se degrada, perdendo precisão, integridade, autenticidade ou completude, maior sua entropia (Floridi, 1999, p. 44-45, nota 7). Nessa perspectiva, toda entidade informacional, seja um documento, um perfil de dados, uma obra cultural ou um ser humano, possui uma dignidade ontológica mínima que merece respeito moral, independentemente de sua utilidade para agentes humanos específicos.

O núcleo normativo da proposta de Floridi organiza-se em quatro leis morais graduais, listadas em ordem crescente de valor:

> 0. entropia não deve ser causada na infosfera (lei nula); 1. entropia deve ser prevenida na infosfera; 2. entropia deve ser removida da infosfera; 3. o bem-estar informacional deve ser promovido pela extensão (quantidade), melhoria (qualidade) e enriquecimento (variedade) da infosfera (Floridi, 1999, p. 47, tradução nossa).

Aplicadas ao problema da explicabilidade algorítmica, essas leis têm implicações diretas: um sistema XAI que produz explicações falsas, incompletas ou intencionalmente enganosas viola a lei nula ao aumentar a entropia da infosfera; um sistema que não previne a amplificação de vieses discriminatórios viola a lei preventiva; um sistema que não oferece mecanismos de contestação e correção viola a lei reparadora. Portanto, o cumprimento das quatro leis não é uma exigência meramente técnica, mas uma obrigação ética de primeira ordem.

Outro conceito central para a aplicação da EI à XAI é o de *me-hood*: a identidade informacional do sujeito. Para Floridi (1999, p. 53, tradução nossa), "somos nossa informação, e quando uma entidade informacional é um ser humano no extremo receptor de uma ação, podemos falar de um me-hood". Desta forma decorre que violações algorítmicas de privacidade, como o uso não transparente de dados pessoais para treinar modelos preditivos, não são apenas infrações jurídicas: são desfigurações ontológicas do sujeito como entidade informacional, atingindo sua integridade, autenticidade e autonomia. Esse entendimento torna a violação da privacidade mais do que uma infração técnica, configurando-se como um problema ético fundamental para a CI.

A força da proposta de Floridi consiste, em oferecer um princípio de avaliação ética aplicável a qualquer sistema informacional: o dever de preservar e enriquecer a infosfera,

prevenindo o aumento de entropia informacional. Contudo, sua limitação é estrutural: ao construir esse princípio sobre uma ontologia formal da informação que se pretende universalmente válida, Floridi pressupõe tacitamente que os critérios de "integridade", "autenticidade" e "bem-estar informacional" são culturalmente homogêneos, pressuposto este que a abordagem de Capurro contestará de forma sistemática.

2.2 A ÉTICA DA INFORMAÇÃO INTERCULTURAL DE CAPURRO: FACTUALIDADE, PLURALIDADE E OS LIMITES DO UNIVERSALISMO

Rafael Capurro parte de uma pergunta incômoda: será que os princípios que orientam a Ética da informação global reproduzem, sob aparência de universalidade, os valores e interesses de uma parcela culturalmente específica da humanidade? Essa questão não é retórica. Conforme apresenta o autor, a conferência realizada em Oxford em 2005, dedicada à Ética da informação em perspectiva intercultural, discutiu explicitamente "se a ética da informação, neste sentido ontológico ou global, pode ser tendenciosa em favor de valores e interesses ocidentais" (Capurro, 2008). Essa suspeita constitui o ponto de partida da Ética da Informação Intercultural (EII), campo que Capurro ajudou a sistematizar e que propõe uma abordagem filosófica radicalmente distinta da proposta por Floridi.

A divergência não é meramente de destaque: ela é estrutural. Enquanto Floridi fundamenta a EI em uma ontologia formal, o ser compreendido como informação, Capurro fundamenta a EII na factualidade da existência humana, na tradição fenomenológica heideggeriana. Para o autor, "o enigmático 'fato' do nosso ser-no-mundo, nossa facticidade, é o 'primeiro chamado' ou *primum movens* da nossa vontade" (Capurro, 2008). Essa facticidade não é um ponto de chegada universal, mas um ponto de partida compartilhado que se desdobra em interpretações culturais radicalmente distintas: os estados de ânimo pelos quais os seres humanos se relacionam com o mundo, o espanto grego, a compaixão budista, a solidariedade ubuntu, não são meras variações superficiais sobre um mesmo fundo, mas modos constituidos diferentes de habitar a existência e, de estabelecer obrigações morais.

A consequência mais relevante dessa posição para o problema da XAI aparece com clareza no caso analisado por Capurro a partir de Nishigaki: enquanto a epistemologia ocidental valoriza a justificação por razões explícitas e universalizáveis, tradição que Nishigaki identifica (apud Capurro, 2008) como tendo "forte afinidade com a busca judaico-cristã por uma interpretação universal dos textos sagrados", a tradição zen budista orienta-se pela

eliminação de interpretações convencionais e pela "encenação" do conhecimento como história de ações no mundo (Capurro, 2008). Essa diferença tem implicações diretas para o design e a avaliação de sistemas XAI. Um sistema que explica suas decisões através de pesos atribuídos a variáveis, o padrão dominante na literatura técnica sobre explicabilidade (Doshi-Velez; Kim, 2017), opera segundo um modelo de justificação inteligível dentro da epistemologia analítica ocidental, mas que pode ser cognitivamente estranho a usuários formados em tradições distintas.

Capurro não decorre dessa constatação um relativismo ético que inviabilizaria qualquer avaliação transcultural. Sua proposta metodológica distingue dois tipos de postura Ética intercultural: a postura *pros hen*, que busca pontos de convergência normativa a partir de um horizonte comum, e a *postura hothen*, que retorna às fontes existenciais da moralidade antes de buscar qualquer convergência. O autor defende a segunda: "em ressonância com o conceito aristotélico de Charles Ess de uma ética *pros hen* que busca a interpretação e aplicação pluralistas de normas éticas compartilhadas, defendo uma abordagem *hothen* que volte a atenção para a questão das fontes das normas éticas, incluindo a experiência cognitivo-emocional múltipla de tais fontes" (Capurro, 2008). Para o objetivo deste trabalho, a postura *hothen* implica que princípios éticos de avaliação da XAI devem ser construídos a partir de formas de vida informacional concretas e culturalmente situadas, e não impostos sobre elas a partir de padrões técnico-jurídicos externos.

Conforme argumenta Suave e Ramalho (2025) em diálogo com essa perspectiva, "a ética intercultural da informação deve considerar as diferentes formas de vida informacionais e seus valores plurais" (Suave; Ramalho, 2025). González de Gómez (2003, citada em Suave; Ramalho, 2025) igualmente sustenta que a informação deve ser compreendida como prática social mediada por valores, disputas e assimetrias, o que reforça a posição de Capurro: uma ética informacional que ignore as condições concretas de produção e circulação do conhecimento permanece normativa e abstrata, incapaz de orientar práticas reais.

2.3 HANNAH ARENDT: PLURALIDADE, JULGAMENTO E A DIMENSÃO POLÍTICA DA EXPLICABILIDADE

A incorporação de Hannah Arendt ao referencial teórico deste trabalho não é aleatória: ela responde a uma necessidade identificada dedicada às contribuições de Arendt para a CI. Suave (2024) demonstrou que os conceitos arendtianos de trabalho, obra e ação, as três

atividades fundamentais da vida ativa, oferecem instrumentos para compreender a informação em sua dimensão política e não apenas técnica. A ação, em particular, corresponde ao domínio em que os sujeitos aparecem publicamente, revelam quem são e iniciam processos que transformam o mundo compartilhado. Barreto (2002), conforme mapeado nessa tese, aproxima essa categoria do que denomina Tempo do Conhecimento Interativo, o momento em que a informação deixa de ser gerenciada ou processada para tornar-se constitutiva da identidade e da participação social dos sujeitos (Suave, 2024, p. 17–18).

A inclusão de Arendt também responde a uma lógica de justificação análoga à empregada para Floridi e Capurro: há tanto convergência quanto tensão a explicitar. Com Capurro, Arendt converge na recusa de uma ética informacional puramente formal e técnico-normativa e na valorização da pluralidade. Em Capurro, pluralidade de mundos informacionais culturalmente situados; em Arendt, pluralidade como "condição de toda vida política" (Arendt, 2010, p. 10), dois sentidos distintos, porém mutuamente potentes. Com Floridi, há uma tensão produtiva: o ontocentrismo de Floridi desloca o centro da preocupação moral da ação humana para o estado da infosfera, incluindo entidades não humanas. Arendt, ao contrário, é radicalmente antropocêntrica, pois para ela é a obediência automática e não refletida que esvazia a ação de sentido político. Essa tensão não é uma falha da articulação, mas sua justificativa mais forte: a presença de Arendt reintroduz deliberadamente um contrapeso antropocêntrico ao ontocentrismo de Floridi, evitando que a preservação da infosfera se torne um fim em si mesma, desconectado do juízo dos sujeitos humanos concretos. Há, uma razão mais concreta e específica para essa inclusão: decisões algorítmicas nocivas, quando naturalizadas como "resultado do algoritmo", reproduzem estruturalmente o mecanismo que Arendt identificou no julgamento de Adolf Eichmann, a obediência sem reflexão crítica sobre as consequências dos próprios atos, isto é, a banalidade do mal. Neste ponto, o problema que a XAI deve resolver não é apenas técnico, tornar a decisão legível, porém é, em sentido arendtiano, o risco de dissolução da responsabilidade moral individual atrás da suposta neutralidade do algoritmo.

Para Arendt (2010, p. 221–222), a ação "em seu sentido mais geral, significa tomar iniciativa, iniciar", e essa capacidade de início é inseparável da pluralidade humana, o fato de que o mundo é habitado por sujeitos radicalmente distintos entre si, cada um portador de uma perspectiva singular e insubstituível. A pluralidade não é, para Arendt, um problema a

ser gerenciado pela técnica ou pelo consenso: é "a condição de toda vida política" (Arendt, 2010, p. 10), o fundamento sem o qual a esfera pública se dissolve em uniformidade ou silêncio. Traduzida para o problema da IA, essa categoria produz uma exigência normativa precisa: sistemas algorítmicos que organizam a visibilidade, o discurso e a circulação de informações são também sistemas que moldam as condições de aparecimento público dos sujeitos e, deste modo, têm implicações diretamente políticas sobre quem pode agir e como.

A contribuição mais específica de Arendt para a avaliação da XAI reside no conceito de julgamento. Inspirado na terceira Crítica de Kant, o julgamento arendtiano designa a faculdade de pensar que "percorre" as perspectivas dos outros, não para igualar a elas, mas para ampliar a própria compreensão do caso particular em deliberação. Para Arendt (2010, p. 180), "a essência da ação humana é estar com os outros, no modo de revelação mútua", condição que os sistemas algorítmicos contemporâneos frequentemente comprometem ao substituir o encontro plural por filtragens automatizadas. Conforme argumentam Suave e Ramalho (2025), o risco central das redes algorítmicas contemporâneas é precisamente o "esvaziamento ético da ação política", a substituição da ação autêntica por "comportamentos automatizados e reativos". A avaliação Ética da XAI, à luz de Arendt, não limitar-se à verificação técnica das explicações produzidas: ela examina se essas explicações efetivamente restituem ao sujeito sua condição de agente político capaz de julgamento autônomo.

3 PROCEDIMENTOS METODOLÓGICOS

Este trabalho adota uma abordagem teórico-conceitual de natureza qualitativa, coerente com a tradição filosófico-informacional que orienta a pesquisa no campo da Ciência da Informação (CI). Pesquisas de natureza teórico-conceitual não produzem dados empíricos primários, mas constroem conhecimento original por meio do trabalho rigoroso sobre conceitos, tradições filosóficas e suas inter-relações, o que, conforme argumentam Suave e Ramalho (2025), exige um método que "não apenas descreva a literatura, mas compreenda os sentidos subjacentes aos discursos científicos, explorando contradições, ambiguidades e disputas simbólicas".

O percurso metodológico combina três procedimentos analíticos complementares. O primeiro é a análise conceitual filosófica, que consiste na identificação, definição rigorosa e comparação sistemática dos conceitos centrais nos textos primários: infosfera, entropia,

agência moral e *me-hood* em Floridi (1999); factualidade, estados de ânimo, pluralidade dos mundos informacionais e distinção *pros hen/hothen* em Capurro (2008); pluralidade, ação e julgamento em Arendt (2010). Sem esse procedimento, qualquer articulação entre os autores correria o risco de ser superficial, fundada em semelhanças terminológicas antes do que em compatibilidade argumentativa real.

O segundo procedimento é a genealogia intelectual, que mapeia as adoções filosóficas de cada autor (a), a tradição analítica em Floridi, a fenomenologia heideggeriana e a hermenêutica em Capurro, o republicanismo e a leitura de Kant em Arendt, com o objetivo de identificar onde os sistemas compartilham pressupostos e onde divergem estruturalmente. Conforme observa Suave e Ramalho (2025), a hermenêutica "parte do princípio de que todo texto carrega valores, pressupostos e intenções, e que sua interpretação depende da interação entre o leitor e o material analisado", o que torna a genealogia intelectual condição de possibilidade de uma leitura crítica não anacrônica.

O terceiro procedimento é a análise hermenêutica fundamentada em Gadamer (1999) e Ricoeur (1983), que orienta a leitura dos textos primários em movimento circular entre partes e totalidade, o chamado círculo hermenêutico, atenta aos pressupostos implícitos, aos silêncios argumentativos e às tensões internas que não se manifestam na superfície dos textos. Como aporte documental, o artigo analisa documentos normativos internacionais sobre Ética da IA, especialmente a Recomendação da UNESCO sobre Ética da Inteligência Artificial (UNESCO, 2021) e as Diretrizes da OCDE para IA (OCDE, 2019), verificando em que medida incorporam, ignoram ou contradizem os princípios filosóficos discutidos. Esses documentos funcionam como "casos de teste" documentais da pergunta de pesquisa, materializando em linguagem normativa a tensão entre universalismo ético e sensibilidade intercultural. Esses documentos funcionam, portanto, como parâmetros normativos externos que permitem confrontar e contextualizar, e não como dados empíricos primários, a discussão teórico-conceitual desenvolvida a partir de Floridi, Capurro e Arendt.

4 RESULTADOS E DISCUSSÃO

A análise dos referenciais de Floridi e Capurro permite identificar importantes pontos de convergência para a construção de fundamentos éticos da Inteligência Artificial Explicável. A primeira convergência reside na rejeição da neutralidade técnica. Para Floridi (1999), sistemas informacionais integram a infosfera e possuem implicações morais

intrínsecas; para Capurro (2008), tecnologias da informação jamais são neutras, pois sua apropriação produz diferenças culturais e morais. Em ambos os casos, a explicabilidade algorítmica deixa de ser uma questão meramente técnica para constituir um problema ético.

A segunda convergência refere-se ao reconhecimento do valor moral das entidades informacionais. Enquanto Floridi fundamenta esse valor no princípio da igualdade ontológica, Capurro enfatiza a dignidade das formas de vida informacional culturalmente situadas. Embora partam de perspectivas distintas, ambos defendem que dados, conhecimentos, identidades e demais entidades informacionais não podem ser reduzidos à sua utilidade instrumental.

A terceira convergência encontra-se na responsabilidade moral. Floridi relaciona a responsabilidade à preservação da integridade da infosfera e à redução da entropia informacional, ao passo que Capurro destaca a responsabilidade ética como resposta situada às demandas concretas dos diferentes contextos culturais. Ambos rejeitam a naturalização das decisões tecnológicas e a diluição da responsabilidade em mecanismos técnicos aparentemente autônomos.

Essas convergências indicam que Floridi e Capurro não são referenciais incompatíveis, mas complementares. Floridi fornece um fundamento normativo para a preservação da integridade informacional, enquanto Capurro oferece uma perspectiva intercultural capaz de contextualizar a aplicação desses princípios em diferentes formas de vida informacional.

As tensões entre os autores emergem precisamente do ponto de partida ontológico. Floridi sustenta uma ontologia informacional de pretensão universal, segundo a qual toda entidade possui valor intrínseco. Capurro, por sua vez, questiona a universalidade dessas categorias, argumentando que toda concepção ética está historicamente situada e atravessada por contextos culturais específicos. Essa divergência torna-se particularmente relevante para a XAI, uma vez que os modelos dominantes de explicabilidade pressupõem formas de justificação e compreensão derivadas da tradição epistemológica ocidental.

Essa tensão não é apenas de escopo, universal X local, mas de método filosófico. O próprio Capurro nomeia o problema ao afirmar que o debate sobre uma fundamentação ontológica da Ética da informação, suas questões, terminologia e finalidade, está, até o momento, profundamente enraizado na Filosofia ocidental (Capurro, 2006a, p. 184 apud Capurro, 2008, tradução nossa), uma crítica que atinge em cheio qualquer projeto de

macroética ontocêntrica de validade universal, como o de Floridi. Isso não torna os dois autores incompatíveis, mas exige explicar em que nível cada um opera: a Ética da Informação de Floridi fornece o conteúdo normativo de primeira ordem, o que deve ser preservado, a integridade da infosfera, sem o qual a Ética da XAI recairia em relativismo; a Ética da Informação Intercultural de Capurro fornece a instância crítica e reflexiva de segunda ordem, que impede esse conteúdo de ser tomado como culturalmente neutro, exigindo sua tradução para formas de vida informacionais concretas. Assim o framework aqui proposto resulta, de uma articulação genuína, e não de mera justaposição de contribuições distintas, pois cada autor (a) desempenha uma função diferente e necessária na arquitetura dos princípios propostos.

Nesse sentido, a explicabilidade não pode ser compreendida apenas como transparência técnica. Diferentes culturas atribuem significados distintos à privacidade, à responsabilidade, à interpretação e à própria noção de explicação. Assim, um sistema capaz de explicar adequadamente suas decisões em determinado contexto cultural pode não produzir o mesmo efeito em outros contextos.

Longe de representar um obstáculo, a tensão entre o universalismo de Floridi e o particularismo de Capurro constitui o principal potencial analítico desta pesquisa. O primeiro impede que a ética da IA se dissolva em relativismo; o segundo evita que princípios éticos sejam convertidos em formas de imperialismo epistêmico. A articulação entre ambos oferece bases mais robustas para o desenvolvimento de sistemas de Inteligência Artificial Explicável sensíveis à pluralidade dos contextos socioculturais.

Com base na discussão precedente, propomos um framework composto por quatro princípios orientadores para o desenvolvimento ético de sistemas de Inteligência Artificial Explicável (XAI), articulando as contribuições de Floridi, Capurro e Arendt. Esses princípios destinam-se a orientar a concepção, implementação e governança de sistemas explicáveis em contextos culturalmente heterogêneos.

Princípio 1: Integridade informacional - Fundamentado nas leis morais de Floridi e nas propriedades da infosfera, este princípio estabelece que sistemas XAI devem ser concebidos de modo a preservar a integridade da infosfera e evitar a produção de entropia informacional por meio das explicações que oferecem. Conforme a tabela de propriedades informacionais apresentada por Floridi (1999, p. 46), as dimensões mais relevantes para a avaliação são: *accuracy* (precisão), *integrity* (integridade), *authenticity* (autenticidade) e

completeness (completude). Uma explicação algorítmica que seleciona apenas as variáveis favoráveis à decisão tomada, omite fatores determinantes ou apresenta contrafactuais enganosos viola a lei nula de Floridi ao introduzir ativamente desinformação no ecossistema informacional. A incorporação deste princípio ao desenvolvimento exige: (a) verificação da fidelidade da explicação ao processo decisório interno do modelo; (b) auditoria da completude dos fatores apresentados; (c) verificação da ausência de vieses sistemáticos na seleção das informações explicadas.

Princípio 2: Legibilidade culturalmente situada - Fundamentado na abordagem *hothen* de Capurro, este princípio estabelece que sistemas XAI devem ser desenvolvidos de modo que suas explicações sejam legíveis e significativas para os sujeitos inseridos em diferentes formas de vida informacional. Conforme González de Gómez (2003), a informação deve ser compreendida como "prática social mediada por valores, disputas e assimetrias", o que implica que a explicabilidade não é apenas um problema de interface, mas um problema de reconhecimento epistêmico. A operacionalização deste princípio durante o desenvolvimento exige: (a) mapeamento das formas de justificação epistêmica valorizadas nas comunidades afetadas; (b) participação dessas comunidades na definição do formato das explicações; (c) validação intercultural das explicações produzidas.

Princípio 3: Promoção da capacidade de julgamento - Fundamentado nas categorias arendtianas de ação e julgamento, este critério estabelece que um sistema XAI eticamente desenvolvido não é aquela que simplesmente registra os fatores determinantes da decisão automatizada: é aquela que efetivamente habilita o sujeito a exercer sua capacidade crítica sobre ela. Para Arendt (2010, p. 221), a ação em sentido pleno significa "tomar iniciativa, iniciar", capacidade que pressupõe que o sujeito não seja reduzido a receptor passivo de decisões externas, mas preserve sua condição de agente político em relação aos sistemas que o governam. Conforme argumentam Suave e Ramalho (2025), o risco central das redes algorítmicas contemporâneas é precisamente o "esvaziamento ético da ação política". A incorporação deste princípio ao desenvolvimento requer: (a) existência de mecanismos efetivos de contestação e revisão acessíveis aos sujeitos afetados; (b) clareza sobre os responsáveis pelas decisões contestadas; (c) garantia de que a explicação não funciona apenas como legitimação post hoc de decisões irreversíveis.

É preciso distinguir julgamento de contestação. O julgamento, em Arendt, designa uma faculdade mental, não um procedimento formal, que consiste em apreciar um caso particular

à luz de uma pluralidade de perspectivas, sem subsumi-lo mecanicamente sob uma regra geral. É o que Kant chamava de juízo reflexionante, retomado por Arendt como "mentalidade alargada" (Arendt, Lições sobre a filosofia política de Kant). A contestação, por sua vez, é um dispositivo institucional, um canal formal de recurso ou revisão: um sistema pode oferecer contestação sem habilitar julgamento algum, assim como é possível exercer julgamento crítico mesmo sem qualquer canal formal de contestação. O Princípio 3, não se satisfaz apenas com a existência de mecanismos de contestação, exige que a explicação forneça informação suficiente para que o sujeito afetado e terceiros na posição de espectadores, no sentido arendtiano, possam formar um juízo próprio sobre o caso particular, considerando-o a partir de múltiplos pontos de vista. Nesse sentido, Aldo Barreto, já antecipava essa mesma tríade arendtiana ao perguntar: "Como ficará o pensar, o querer e o julgar nessas novas configurações de tempo e de espaço e informação?" (Barreto, 2002, p. 74).

Critério 4: Redução de assimetrias **de poder -** Fundamentado na articulação entre Capurro, Arendt e a produção crítica da CI, este critério estabelece que sistemas XAI implantados em contextos de vulnerabilidade social diferencial devem ser concebidos de forma a minimizar assimetrias informacionais e evitar a reprodução de desigualdades estruturais. Conforme a análise de Zuboff (2019) sobre o capitalismo de vigilância, "dados comportamentais são transformados em produtos preditivos", configurando "uma nova forma de dominação invisível" que afeta desproporcionalmente os grupos mais vulneráveis. Butler (2015) acrescenta que "certos corpos são mais expostos à violência simbólica e à exclusão discursiva nas redes". A implementação deste princípio requer: (a) análise do impacto diferencial das explicações sobre grupos historicamente marginalizados; (b) verificação de que a XAI não funciona como legitimação técnica de discriminações estruturais; (c) incorporação de perspectivas das comunidades mais vulneráveis na definição dos parâmetros avaliativos. O quadro 1 a seguir sintetiza o framework proposto:

Quadro 1 - Princípios orientadores para o desenvolvimento ético de sistemas de Inteligência Artificial Explicável

Princípio	Fundamentação	Dimensão	Diretriz de desenvolvimento
Integridade informacional	Floridi (1999): leis morais e propriedades da infosfera	Ontológica	Desenvolver explicações fiéis, completas e autênticas

Princípio	Fundamentação	Dimensão	Diretriz de desenvolvimento
Legibilidade culturalmente situada	Capurro (2008): abordagem hothen	Cultural-epistêmica	Adaptar explicações aos contextos culturais dos usuários
Promoção da capacidade de julgamento	Arendt (2010): ação e julgamento	Política	Favorecer contestação, compreensão e agência dos sujeitos
Redução de assimetrias de poder	Suave; Ramalho (2025)	Sociopolítica	Minimizar desigualdades informacionais e ampliar participação

Fonte: elaborado pelos autores (2026)

A originalidade do framework reside precisamente em sua recusa de reduzir o desenvolvimento ético da XAI a qualquer dimensão isolada: a integridade ontológica sem legibilidade cultural é imperialismo epistêmico; a legibilidade cultural sem integridade ontológica é populismo técnico; a capacidade de julgamento sem redução de assimetrias é formalismo político; e a redução de assimetrias sem os três princípios anteriores é intervencionismo sem fundamento filosófico.

Neste framework frente a propostas já consolidadas na literatura de Ética de IA. Os cinco princípios da AI4People: beneficência, não maleficência, autonomia, justiça e explicabilidade (Floridi *et al*., 2018), e as Ethics Guidelines for Trustworthy AI do grupo de peritos de alto nível da União Europeia (2019) organizam-se, em ambos os casos, em torno de um registro majoritariamente universalista e procedimental. O framework aqui proposto se diferencia ao incorporar explicitamente uma exigência de legibilidade culturalmente situada (Princípio 2), e de redução de assimetrias de poder (Princípio 4), dimensões ausentes ou apenas tangenciadas nesses outros marcos, o que constitui sua originalidade frente a propostas equivalentes.

Convém registrar que o framework proposto encontra ressonância e ausência simultânea, nos documentos regulatórios internacionais mais relevantes sobre Ética da IA. A Recomendação da UNESCO sobre Ética da Inteligência Artificial (UNESCO, 2021) reconhece explicitamente que "os valores culturais e as tradições influenciam a percepção e a aplicação" dos princípios éticos da IA, e inclui entre seus valores fundamentais a diversidade cultural e linguística. Contudo, o documento não oferece mecanismos concretos para operacionalizar essa sensibilidade intercultural nos processos de design e avaliação de sistemas específicos,

lacuna que o framework aqui proposto pretende, ao menos parcialmente, preencher. As Diretrizes da OCDE - Organização para a Cooperação e Desenvolvimento Econômico para IA (OCDE, 2019) estabelecem a explicabilidade como princípio, mas a definem em termos que pressupõem tacitamente a epistemologia técnica ocidental como padrão universal. Essa ausência confirma a relevância da pergunta de pesquisa que orienta este trabalho e reforça a contribuição que a Ciência da Informação, em diálogo com a Filosofia, pode oferecer ao debate sobre governança de IA.

5 CONSIDERAÇÕES FINAIS

O percurso teórico desenvolvido neste trabalho evidencia que a Ciência da Informação não constitui apenas um campo de aplicação da ética da informação, mas também um espaço de produção de problemas filosóficos originais diante dos desafios colocados pela Inteligência Artificial. Ao confrontar sistemas de IA com contextos marcados pela diversidade cultural, pela mediação informacional e por diferentes formas de produção de conhecimento, a área revela limites das abordagens éticas universalistas e demanda perspectivas capazes de articular dimensões informacionais, culturais e políticas.

Partindo da questão sobre a possibilidade de articulação entre a macroética ontocêntrica de Floridi, a ética intercultural da informação de Capurro e as reflexões de Arendt sobre pluralidade, ação e julgamento, o estudo demonstrou que essa convergência é não apenas possível, mas teoricamente produtiva. Floridi oferece fundamentos normativos voltados à preservação da integridade da infosfera e à responsabilidade moral; Capurro evidencia a necessidade de reconhecer a pluralidade dos contextos culturais e informacionais; e Arendt contribui para compreender a explicabilidade como condição para o exercício do julgamento e da agência dos sujeitos.

Como resultado, foi proposto um framework composto por quatro princípios orientadores para o desenvolvimento ético de sistemas de Inteligência Artificial Explicável: integridade informacional, legibilidade culturalmente situada, promoção da capacidade de julgamento e redução de assimetrias de poder. Em conjunto, esses princípios oferecem subsídios para a concepção de sistemas mais transparentes, responsáveis e sensíveis à diversidade sociocultural, contribuindo para aproximar o debate sobre explicabilidade das preocupações da Ciência da Informação com justiça informacional, mediação e pluralidade.

Por fim, destaca-se que o framework possui natureza teórico-conceitual e demanda validação empírica em diferentes contextos sociotécnicos. Investigações futuras poderão explorar sua aplicação em estudos de caso, análises comparativas de sistemas de IA e processos de governança algorítmica, ampliando sua capacidade explicativa e seu potencial de contribuição para o desenvolvimento de tecnologias mais inclusivas e socialmente responsáveis.

Para concluir, retoma-se o aforismo medieval mobilizado por Floridi (1999, p. 56) como divisa da Ética da Informação: "*Agere sequitur esse*", a ação segue o ser. No contexto da Inteligência Artificial Explicável, essa máxima adquire um significado renovado. Antes de definir como um sistema deve explicar suas decisões, é necessário perguntar para quem essa explicação se dirige, quais formas de vida ela pretende tornar inteligíveis e quais horizontes culturais e informacionais reconhece ou silencia. A explicabilidade não é apenas uma questão técnica, mas uma questão ética, política e informacional. Por isso, o desafio não consiste apenas em construir sistemas capazes de explicar, mas em desenvolver sistemas capazes de dialogar com a pluralidade dos mundos que pretendem afetar. É nesse diálogo interdisciplinar, entre diferentes tradições filosóficas, culturas informacionais e experiências humanas, que reside a possibilidade de uma Inteligência Artificial verdadeiramente explicável.

REFERÊNCIAS

ARENDT, Hannah. **A condição humana**. 11. ed. Rio de Janeiro: Forense Universitária, 2010.

BARRETO, Aldo de Albuquerque. A condição da informação. **São Paulo em Perspectiva**, São Paulo, v. 16, n. 3, p. 67-74, jul./set. 2002. Disponível em: https://www.scielo.br/j/spp/a/5Q85NCzRFvJ8BLjjd54jLMv/?lang=pt. Acesso em: 10 maio 2025.

BUTLER, Judith. **Quadros de guerra**: quando a vida é passível de luto? Rio de Janeiro: Civilização Brasileira, 2015.

CAPURRO, Rafael. Ética da informação intercultural. In: HIMMA, Kenneth E.; TAVANI, Herman T. (orgs.). **The handbook of information and computer ethics**. Nova Jersey: Wiley, 2008. p. 639–665.

DOSHI-VELEZ, Finale; KIM, Been. Towards a rigorous science of interpretable machine learning. **arXiv**, [*s. l.*], v. 1, 2017. Disponível em: https://arxiv.org/abs/1702.08608 . Acesso em: 10 maio 2025.

FLORIDI, Luciano. Information ethics: on the philosophical foundation of computer ethics. **Ethics and Information Technology**, [*s. l.*], v. 1, n. 1, p. 37-56, 1999. Disponível em: https://link.springer.com/article/10.1023/A:1010018611096. Acesso em: 10 maio 2025.

FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica; CHATILA, Raja; CHAZERAND, Patrice; DIGNUM, Virginia; LUETGE, Christoph; MADELIN, Robert; PAGALLO, Ugo; ROSSI, Francesca; SCHAFER, Burkhard; VALCKE, Peggy; VAYENA, Effy. AI4People - An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. **Minds and Machines**, [*s. l.*], v. 28, n. 4, p. 689-707, 2018. Disponível em: https://doi.org/10.1007/s11023-018-9482-5 . Acesso em: 7 set. 2026.

GADAMER, Hans-Georg. **Verdade e método.** Petrópolis: Vozes, 1999.

GONZALEZ DE GÓMEZ, Maria Nélida. Reflexões sobre ética da informação. *In*: CAPURRO, Rafael; GONZALEZ DE GÓMEZ, Maria Nélida (org.). **Ética da informação**. Brasília: UnB, 2003. p. 19-44.

LESSIG, Lawrence. **El código 2.0**. Madrid: Traficantes de Sueños, 2006.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **Recommendation on the ethics of artificial intelligence**. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em: 10 maio 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. **Recommendation of the council on artificial intelligence**. Paris: OCDE, 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 10 maio 2025.

RICOEUR, Paul RICOEUR, Paul. **Tempo e narrativa** (Tomo 1). Tradução de Constança Marcondes César. Campinas: Papirus Editora, 1994. (Publicado originalmente em 1983).

SUAVE, Aline Laureano. **Hannah Arendt e suas contribuições para a Ciência da Informação**. 2024. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista "Júlio de Mesquita Filho", Marília, 2024. Disponível em: https://hdl.handle.net/11449/255712. Acesso em: 10 maio 2025.

SUAVE, Aline Laureano; RAMALHO, Rogério Aparecido Sá. Filosofia e ética informacional na Ciência da Informação: uma leitura hermenêutica dos dilemas das redes digitais. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 25., 2025, Rio de Janeiro. **Anais** [...]. Rio de Janeiro: ANCIB, 2025. Disponível em: https://enancib.ancib.org/enancib/en/article/view/377. Acesso em: 10 maio 2025.

UNIÃO EUROPEIA. High-Level Expert Group on Artificial Intelligence (AI HLEG). **Ethics guidelines for trustworthy AI**. Bruxelas: Comissão Europeia, 2019. Disponível em: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai . Acesso em: 7 set. 2026.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância:** a luta por um futuro humano na nova fronteira do poder. Tradução de George Schlesinger. Rio de Janeiro: Intrínseca, 2019.