



SEÇÃO: CIBERCULTURA

## Inteligência Artificial Responsável: notas conceituais para os estudos da comunicação

*Responsible Artificial Intelligence: conceptual notes for communication studies*

*Inteligencia Artificial Responsable: notas conceptuales para los estudios de la comunicación*

**Francisco Leite<sup>1</sup>**

[orcid.org/0000-0003-0710-3339](http://orcid.org/0000-0003-0710-3339)  
[leitefco@gmail.com](mailto:leitefco@gmail.com)

**Recebido em:** 30 jan. 2024.

**Aprovado em:** 05 nov. 2024.

**Publicado em:** 17 jan 2025.

**Resumo:** Este artigo tem como objetivo delinear e explicar o conceito de Inteligência Artificial Responsável (IAR), implicando nesse exercício sobretudo a comunicação estratégica no campo das organizações. É finalidade também pontuar alguns dos princípios-chave da IAR, visando fomentar e oferecer como resultado algumas reflexões e direcionamentos críticos que contribuam para informar e apoiar os estudos brasileiros da comunicação que tenham e ou terão como foco a IA como objeto de pesquisa nos próximos anos. A metodologia atende uma pesquisa de cunho conceitual, que fomenta e mobiliza uma analítica exploratória e propositiva sobre a temática agendada. Como complemento, alguns casos são observados como objetos de exemplificação.

**Palavras-chave:** inteligência artificial responsável; comunicação; organizações; transformações sociais.

**Abstract:** This paper aims to delineate and explain the concept of Responsible Artificial Intelligence (RAI), with a particular focus on strategic communication within organizational contexts. Additionally, it seeks to outline key principles of RAI, offering critical reflections and guidelines intended to inform and support Brazilian communication studies that focus on, or will focus on, AI as a research subject in the coming years. The methodology employed is a conceptual research approach, facilitating an exploratory and propositional analysis of the topic. As a complement, selected cases are included as illustrative examples.

**Keywords:** responsible artificial intelligence; communication; organizations; social change.

**Resumen:** Este artículo tiene como objetivo delinear y explicar el concepto de Inteligencia Artificial Responsable (IAR), abordando en este ejercicio, sobre todo, la comunicación estratégica en el ámbito organizacional. También se propone señalar algunos de los principios clave de la IAR, con el fin de fomentar y ofrecer, como resultado, reflexiones y orientaciones críticas que contribuyan a informar y apoyar los estudios brasileños de comunicación que tengan o lleguen a tener la IA como objeto de investigación en los próximos años. La metodología adoptada es de carácter conceptual, promoviendo y movilizando un análisis exploratorio y propositivo sobre la temática planteada. Como complemento, son observados algunos casos como objetos de exemplificación.

**Palabras clave:** inteligencia artificial responsable; comunicación; organizaciones; cambio social.

### Introdução



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

Envolvido pelo presente de uma sociedade em profunda transformação, especialmente, desafiada pelos avanços e inovações das tecnologias de inteligência artificial (IA), este artigo se volta para o

<sup>1</sup> Universidade de São Paulo (USP), São Paulo, SP, Brasil.

futuro ao refletir sobre o que, provavelmente, terá importância crescente para as pesquisas e práticas da comunicação nos próximos anos: a necessária abordagem responsável à IA, pautada por atributos éticos e valores humanos (Dignum, 2020, 2017; Baeza-Yates, 2023).

Nessa perspectiva, este trabalho tem como objetivo delinear e explicar as noções conceituais de inteligência artificial responsável (IAR), implicando, nesse exercício, sobretudo a comunicação estratégica no campo das organizações (Zerfass et al., 2018; Kunsch, 2018). Esta pesquisa também busca pontuar alguns dos princípios-chave da IAR, visando estimular reflexões e direcionamentos que contribuam para informar e apoiar os estudos brasileiros da comunicação que têm ou terão como foco a IAR como objeto de pesquisa (Romanini, 2023). Com efeito, é expectativa ainda deste texto, a partir das lentes da IAR, instigar o desenvolvimento de olhares críticos que questionem quando necessário a intervenção e os impactos da IA nos processos socioculturais que passam, progressivamente, a ser mediados por algoritmos (Cordeiro, 2021; Santaella, 2022).

A metodologia deste trabalho atende uma pesquisa de cunho conceitual (MacInnis, 2011), que mobiliza e fomenta uma análise exploratória propositiva (Gil, 2008) sobre a temática agendada neste escrito. Como complemento, alguns casos são indicados como objetos de exemplificação.

Como fontes, considera-se especialmente (não exclusivamente) meta-análises, revisões e estudos empíricos recentes, que foram identificados e originados no campo de pesquisa da comunicação, da ciência da computação, sociologia, antropologia, marketing, entre outros. Este artigo, ao reunir e articular intencionalmente esta produção científica, constrói a sua fundamentação integrando descobertas e perspectivas de autoras e autores de diferentes partes do mundo que, de realidades e contextos sociais diversos de pesquisa – do Sul ao Norte Global (Santaella, 2022; Arun, 2020) –, podem falar de uma gama

variada de pontos de partida socioculturais, teóricos, conceituais e metodológicos acerca dos diferentes estágios de desenvolvimento e aprendizagem da IA. A literatura explorada neste artigo é publicada em inglês e português.

Desse modo, este texto está organizado em três partes. Na primeira, as noções conceituais de IA e outros termos adjacentes são explicados introdutoriamente. Posteriormente, na segunda parte, são abordadas as ideias acerca da IAR e alguns dos seus princípios-chave. Por fim, considerando as questões comunicacionais nesse contexto, na terceira parte, se observa como a IAR está ou não sendo assimilada pelas organizações (públicas e privadas) e sociedade em geral.

## 1 Inteligência Artificial

Segundo Arbix (2021, p. 71), “a busca por uma IA começou a ganhar corpo durante a segunda guerra, com os trabalhos de [...] Turing, e elevou seu estatuto na década de 1950”. Esse autor ainda explica, que a história da IA, “porém, esteve longe de uma ascensão linear. Viveu oscilações fortes, com surtos positivos e retração de investimento, tanto financeiro quanto humano, sempre ligados aos resultados que prometia e que nem sempre se realizavam [...]” (Arbix, 2021, p. 271).<sup>2</sup> No contemporâneo, Cozman e Neri (2021) pontuam que a definição de IA ainda é vaga, porém o seu escopo demonstra-se relativamente claro, tendo em vista que “hoje toda a ação em IA está centrada em computadores digitais [...] cujos programas representam e raciocinam sobre conhecimento e crenças, tomam decisões e aprendem, e interagem com seu ambiente, realizando todas essas atividades ou pelo menos algumas com nível alto de sofisticação (Cozman; Neri, 2021, p. 23).

Desse modo, minimamente, a ideia de IA denota os sistemas ou artefatos construídos por pessoas para cumprir alguns objetivos e ou tarefas determinadas (Dignum, 2020). No entanto, “o termo em si é muito amplo e abrange múltiplas técnicas, incluindo a mais popular, *machine learning*”. Como explica Domingos, a ideia de *machine*

<sup>2</sup> Para uma leitura mais atenciosa sobre a história da IA recomenda-se a obra de Nilsson (2010), para um panorama global, e, especificamente considerando o Brasil, o trabalho de Costa et al. (2021).

*learning* (aprendizagem de máquina)

assume muitas formas diferentes e tem muitos nomes diferentes: reconhecimento de padrões, modelagem estatística, mineração de dados (*data mining*), descoberta de conhecimento, análise preditiva, ciência de dados, sistemas adaptativos, sistemas auto-organizados e muito mais. Cada um deles é usado por comunidades diferentes e tem associações diferentes. [...] uso o termo *machine learning* para me referir amplamente a todos eles (Domingos, 2015, p. 8, tradução própria).

Este artigo segue essa perspectiva de Domingos. *Machine learning* também inclui as redes neurais artificiais (RNA) que, essencialmente, são "operações matemáticas com matrizes que usam nós interconectados inspirados em neurônios do cérebro humano. Elas que quando possuem muitas camadas são chamadas de *deep learning*" (Baeza-Yates, 2023, p. 407, tradução própria). Devidos aos limites deste espaço, os conceitos de *machine learning* e *deep learning* não serão aprofundados. Para leituras sobre tais abordagens recomenda-se Domingos (2015) e Kaufman e Santaella (2020).

No entanto, para uma noção geral dessa dinâmica, Cordeiro (2020) explica que é basilar observar a ideia-chave de algoritmo. Segundo essa autora, ao decompor a IA a nível unitário, chega-se ao algoritmo. Conforme Cormen et al. (2009), um algoritmo é qualquer procedimento computacional bem definido que recebe algum valor, ou conjunto de valores, como entrada (*input*) e produz algum valor, ou conjunto de valores, na saída (*output*). Desse modo, um algoritmo pode ser também compreendido como "uma ferramenta para resolver um problema computacional bem especificado" (Cormen et al., 2009, p. 5, tradução própria).

Gillespie (2016) agrega a essa discussão pontuando que para os seus desenvolvedores, "o algoritmo só surge após a geração de um 'modelo'.

O modelo é a formalização de um problema e seu objetivo, articulado em termos computacionais" (Gillespie, 2016, p. 19, tradução própria). Ele conforma os sistemas algorítmicos, ou seja, os algoritmos processados "em interação com estruturas de dados implementados em softwares [...] – que são capilarizados na nossa vida social" (Cordeiro, 2021, p. 211), ou seja, na cultura (Santaella, 2022).

Os sistemas algorítmicos de IA, para Baeza-Yates (2023), são a espinha dorsal da economia digital e dos processos automatizados. Eles são a face pública da IA amplamente utilizada, por exemplo: comércio eletrônico, serviços de streaming, chatbots, plataformas de redes sociais, tomada de decisão automatizada etc. Nesse sentido, Burrell (2016) reforça que os algoritmos, portanto, operam se configurando e se aperfeiçoando com base em dados extraídos das experiências socioculturais.

Baeza-Yates (2023) explana que a AI se tornou realmente popular há pouco mais de dez anos. Essa popularização se acentuou na consciência pública após o lançamento, em 30 de novembro de 2022, do ChatGPT<sup>3</sup>, um chatbot alimentado por IA generativa, que utiliza RNA, para criar conteúdos inovadores. Esse sistema de IA interage com humanos e permite obter respostas em linguagem natural<sup>4</sup> (LN) a partir de comandos (*prompts*) escritos. O ChatGPT foi desenvolvido pela empresa estadunidense OpenAI. Mas o que se entende por IA generativa?

Como brevemente pontuado, é com base nos processos técnicos de *machine learning* e nos seus algoritmos mais avançados de *deep learning* que a IA se configura e se dinamiza em diversos modelos de RNA como, por exemplo, os discriminativos ou gerativos. Quando se fala em IA generativa ou IA discriminativa deve-se compreender que tais termos denotam sistemas

<sup>3</sup> Em comunicação pessoal, Almeida (2023) explicou que o "ChatGPT é um grande modelo de linguagem (Large Language Model - LLM), com cerca de 175 bilhões de parâmetros, treinado em uma imensa base de dados, capaz de aprender de forma autônoma e produzir textos sofisticados, aparentemente inteligentes. Os parâmetros são chave para os algoritmos de aprendizagem de máquina, pois representam aquilo que é aprendido pelo modelo com os dados de treinamento". Os LLM são um tipo de modelo de aprendizagem de máquina que processa e gera texto em linguagem natural. Além do ChatGPT outras ferramentas de IA fazem uso desse modelo como: assistentes virtuais, chatbots ou outros geradores de texto, similares ao ChatGPT, como o Bard, da Google, lançado em março de 2023.

<sup>4</sup> Processamento de Linguagem Natural (PNL) é vertente da IA, que usa *machine learning*, para ajudar as máquinas computacionais a entender, interpretar e manipular a linguagem humana.

que se baseiam em modelos de *deep learning* treinados com grandes conjuntos de dados (texto e imagem) nem sempre moderados.

Por exemplo, a IA discriminativa opera com algoritmos discriminativos, aplicados para classificar ou predizer informações com base nas diferenças dos dados de entrada (p. ex., algoritmos de sugestão e recomendação dos mecanismos de busca, os algoritmos que apoiam tomadas de decisão etc.). Por sua vez, a IA generativa opera com algoritmos generativos, utilizados para criar dados mais complexos (imagens, textos, vídeos, áudios etc.). Desse modo, IA discriminativa é utilizada para classificar ou predizer dados/informações com base nas diferenças e a AI generativa, vai além dessas classificações operando para produzir dados ou conteúdos novos, ao responder comandos humanos (Botelho; Nöth, 2021).

Nesse contexto, Sodré (2021) chama atenção para o fato problemático que são organizações estadunidenses como Google, Amazon, Meta, Microsoft e IBM, também conhecidas como *big techs*, as administradoras do *big data*; isto é, das grandes massas de dados ou macrodados que manipulam a maioria dos algoritmos complexos de IA. São essas organizações privadas globais que igualmente dominam e operam, mediante as suas plataformas digitais, a maioria dos sistemas algorítmicos que tem "invadido de forma crescente todas as atividades e aspectos de nossas vidas, inclusive nossas práticas culturais, por meio de sugestões de livros pela Amazon ou de filmes e séries nos serviços de *streaming*" (Santaella, 2022, p. 73).

Em suma, hoje, a IA é propriedade das *big techs* e se as regulações governamentais, ao redor do globo, não forem aceleradamente estabelecidas inscrevendo regimes efetivos de responsabilidades que coloquem os interesses do público acima das promessas empresariais, essas poucas e poderosas organizações privadas podem aproveitar das "suas posições para definir a agenda política da IA [...]" (Não se engane [...] , 2023). Essa brecha, na perspectiva da vulnerabilidade existente ou potencial, pode gerar insegurança, riscos e danos aos indivíduos, grupos,

sociedade, economia, ecossistemas, valores e normas sociais (ONU, 2023).

Outro aspecto problemático que integra esse contexto refere-se à opacidade dos sistemas de IA baseados em aprendizagem de máquina. Eles, geralmente, não oferecem direcionamentos transparentes para se compreender a lógica definida e operada pelos seus algoritmos. Burrell (2016) considera que essa opacidade dos sistemas de IA vem tomando o centro das preocupações sobre os algoritmos entre reguladores e cientistas sociais. Nesse contexto, Maranhão, Abrusio e Almada (2021) chamam atenção que a preocupação atual com essa opacidade se refere, centralmente, à interdição e neutralização (ou não) das contestações sociais direcionadas aos algoritmos de IA.

Nessa perspectiva, esses autores também ressaltam o risco que essa opacidade pode gerar na incorporação, voluntária ou não, de vieses nos sistemas de IA, "que resultem em construção de perfis ou tomadas de decisão discriminatórias, ou ainda da possibilidade de tomadas de decisão que ignorem valores humanos ou desrespeitem direitos fundamentais e a dignidade humana" (Maranhão, Abrusio; Almada, 2021, p. 291-292).

Dado esse breve panorama, é possível compreender explicitamente, de acordo com Santaella (2022), que não é por acaso que os mais diversos aspectos da IA, atualmente, constam com destaque nas agendas e debates de governos e de entidades internacionais de relevância. Logo, isso denota que paralelamente "à corrida do mundo empresarial rumo à transformação digital, emerge a necessidade de enfrentamento de dilemas relativos ao acesso a dados, viés algorítmico, ética e transparência e responsabilidade legal por decisões resultantes de IA" (Santaella, 2022, p. 73). Esses dilemas vêm enredando a todos, especificamente, tendo em vista a capacidade, direta ou indireta, que os algoritmos de IA têm de interferir e influenciar os processos e as relações sociais (Cordeiro, 2021).

Portanto, é na emergência desses debates e embates acerca tanto dos potenciais benefícios e oportunidades quanto dos riscos e danos

que as tecnologias de IA podem oferecer para indivíduos, grupos, sociedades e para o planeta, que a ideia de inteligência artificial responsável vem se difundindo e se estabelecendo como possibilidade de intervenção crítica.

## 2 Inteligência Artificial Responsável

O conceito de inteligência artificial responsável (IAR), segundo Dignum (2020), reflete uma abordagem que trata da responsabilidade humana no desenvolvimento de sistemas inteligentes, em conformidade com princípios e valores humanos fundamentais, de modo a garantir o progresso e o bem-estar social em um mundo sustentável. Nessa direção, a premissa central da IAR postula que, para que os sistemas de IA sejam seguros, reconhecidos e confiáveis, eles devem ser concebidos para considerar ponderações éticas e consequências morais das suas ações e decisões, de maneira responsável, transparente e *accountable*<sup>5</sup> (Dignum, 2017).

Por sua vez, Baeza-Yates (2023) elucida que a IAR objetiva, portanto, criar sistemas que beneficiem os indivíduos, as sociedades e o meio ambiente. Ela abrange todos os aspectos éticos, legais e técnicos ao longo do desenvolvimento (análise, projeto, construção, implantação e avaliação) e usos de tecnologias benéficas de IA, bem como busca garantir que sistemas de IA não interfiram na agência humana, não causem riscos e danos, não discriminem/oprimam ou desperdicem recursos ambientais.

Com efeito, De Laat (2021) pontua que o termo IAR vem sendo utilizado para denotar uma IA que seja justa e imparcial, transparente e explicável, segura e protegida, que garanta privacidade, *accountable* e benefícios à humanidade. Ainda segundo esse autor, desde 2016, muitas organizações ao redor do mundo, voluntariamente, vêm prometendo fidelidade a esses princípios.

De Laat (2021) ainda ressalta, com base em Steurer (2013), que a IA, para assegurar a sua abordagem responsável, observando os seus benefícios e riscos, deve ser objeto-alvo para a cobertura e desenvolvimento de três tipos básicos de regimes regulatórios: a) regulamentação pelo governo (legislação mestra); b) autorregulamentação por parte das empresas (ou suas associações comerciais), por exemplo, recomendando a adoção voluntária de boas práticas, códigos de conduta etc.; e c) regulação pela sociedade civil e suas organizações sociais via, por exemplo, contestações públicas geralmente suportadas por investidores críticos, funcionários críticos, consumidores críticos etc., bem como iniciativas de organizações como a Associação Brasileira de Ciências (ABC), que recentemente elaborou e difundiu de modo amplo as suas *Recomendações para o avanço da Inteligência Artificial no Brasil* (2023).

Esse documento da ABC, divulgado em novembro de 2023, compreende a IA responsável como um conjunto de iniciativas que vem emergindo em resposta às preocupações sociais com a IA como, por exemplo, as violações de privacidade, criação de ambientes anticompetitivos, opressão, racismo, manipulação de comportamentos, ocorrência de desastres ambientais, direitos autorais, entre outras.

A publicação também chama atenção para as áreas estratégicas que potencialmente poderão ser afetadas pela IA, a saber: saúde, energia elétrica, financeira, meio ambiente, educação, agricultura, pesquisa científica, competitividade das empresas e eficiência e governança do setor público. As áreas da comunicação e mídia não são implicadas nesse documento. No entanto, nesse conjunto de recomendações da ABC, de modo substancial, consta demarcada a urgência para o desenvolvimento e estabelecimento de

<sup>5</sup> Não há em português uma tradução literal que abarque adequadamente os sentidos desses vocábulos em inglês: *accountable* ou *accountability*. Grosso modo, os sentidos desses termos refletem a ideia de compromisso com resultados, ou seja, envolvem "responsabilidade (objetiva e subjetiva), controle, transparência, obrigação de prestação de contas, justificativas para as ações que foram ou deixaram de ser empreendidas, premiação e/ou castigo" (Pinho; Sacramento, 2009, p. 1364). Dignum chama atenção para a corrente confusão entre as ideias de *accountability* e responsabilidade elucidando que a primeira se refere à capacidade de explicar ou relatar o papel de algo ou alguém "em eventos ou ações, enquanto responsabilidade é o dever de responder pelas próprias ações. Responsabilidade implica dever e existe antes que a tarefa ou ação seja realizada. A *accountability* só é evidente depois que a ação é realizada ou não" (Dignum, 2020, p. 219, tradução própria).

princípios, regras e legislações (regulação), no Brasil, que tratem, por exemplo, de minimizar os riscos e danos das tecnologias de IA; da formação de recursos humanos; da IA na educação; de investimento em pesquisa e desenvolvimento de IA; e de políticas de proteção de dados mais adequadas às características da IA.

Nesse horizonte, é preciso informar que ainda não há no mundo<sup>6</sup> uma legislação que regulamente de modo geral a IA. No Brasil, por exemplo, no Senado Federal tramita o Projeto de Lei n.º 2338, de 2023, que dispõe sobre o uso da Inteligência Artificial no país, focando em princípios como a igualdade de direitos, a não discriminação e a centralidade no elemento humano.

Já o Ministério da Ciência, Tecnologia, Inovações e Comunicações<sup>7</sup> (MCTIC) instituiu a IA como uma prioridade em projetos de pesquisa e no desenvolvimento de tecnologias e inovações para o período 2020 a 2023. No desdobramento dessa iniciativa governamental foi anunciada, em 2021, a *Estratégia Brasileira de Inteligência Artificial* (EBIA) para direcionar as prioridades do Brasil em torno do tema, considerando especialmente contribuir para a elaboração de princípios éticos para o desenvolvimento e uso da IAR.

Abrindo e avançando com o racional dessa iniciativa, em 2024, a pedido do presidente Luiz Inácio Lula da Silva (2023-), o MCTI apresentou o Plano Brasileiro de Inteligência Artificial (PBIA) 2024-2028. Entre os seus objetivos o documento prevê cerca de R\$ 23 bilhões em recursos para desenvolver a infraestrutura nacional, incluindo um supercomputador, além de incentivos à indústria local de IA. O plano abrange cinco eixos centrais de investimento: infraestrutura e desenvolvimento de IA; difusão, formação e capacitação em IA; IA para melhorias dos serviços públicos; IA para inovação empresarial; apoio ao processo

regulatório e de governança de IA. Em síntese, o PBIA foca o desenvolvimento de "soluções em IA que melhorem significativamente a qualidade de vida da população, otimizando a entrega de serviços públicos e promovendo a inclusão social" (Plano [...], 2024).

É importante pontuar que essas iniciativas do governo brasileiro se orientam pelos princípios de IAR estabelecidos pela *Organization for Economic Co-operation and Development* (OECD)<sup>8</sup>, que abarcam: a) crescimento inclusivo, desenvolvimento sustentável e bem-estar; b) valores centrados no ser humano e na justiça; c) transparência e explicabilidade; d) robustez<sup>9</sup>, segurança e proteção; e e) *accountability*.

Retomando o conceito de IAR, como visto *a priori*, termos de expressões humanas como ética, confiável e responsável estão sendo associados a sistemas e máquinas nos debates atuais. Logo, "existe o perigo de que a pessoa comum presuma que são intercambiáveis" (Baeza-Yates, 2023, p. 408, tradução própria). Desse modo, conforme Baeza-Yates (2023), é preciso observar as sensibilidades distintas que tais termos carregam para se compreender por que o qualificador "responsável" vem sendo considerado como o mais apropriado para envolver o contexto das discussões sobre a IA. Baeza-Yates explica que a ideia de:

**IA ética** implica um grau de agência moral, uma condição humana que requer julgamentos de valor e comportamento pretendidos. No entanto, na sua encarnação atual, a IA não é um agente moral. A ética é, portanto, algo que apenas um ser humano pode demonstrar, assim transferir isso para uma máquina não é uma boa ideia. A **IA confiável** é ainda mais problemática. A IA não é humana, por isso devemos evitar anexar capacidades humanísticas não apenas a um algoritmo, mas também a uma máquina projetada para operar no vácuo. [...]. Simplificando, a IA não deve ser colocada em nenhum desses grupos. Tanto "éticas" como "confiáveis" são capacidades que podem ser

<sup>6</sup> No dia 12 de julho de 2024, o Conselho da União Europeia aprovou a primeira lei abrangente de IA do mundo, a *Artificial Intelligence Act* (AI Act), com o objetivo de harmonizar as regras sobre IA na UE. Essa lei será aplicada progressivamente até 2026. O Reino Unido, os EUA e a China, ainda não publicaram as suas próprias regulações de IA (Como inteligência [...], 2024). No entanto, esses dois últimos países, seguidos pela União Europeia, estão na liderança da corrida global pela regulação da IA (Röhe, 2024).

<sup>7</sup> É relevante informar que a Medida Provisória 980/2020 recriou o Ministério das Comunicações a partir de desmembramento do MCTIC, que passou a ser denominado Ministério da Ciência, Tecnologia e Inovações (MCTI). A MP entrou em vigor em 10 de junho de 2020 (Júnior, 2020).

<sup>8</sup> Disponível em: <https://oecd.ai/en/dashboards/countries/Brazil>. Acesso em: 14 jan. 2024.

<sup>9</sup> Grosso modo, nos estudos computacionais, o termo robustez refere-se à capacidade de um sistema ou algoritmo de lidar com condições desafiadoras, adversas, erros ou variações imprevistas mantendo o seu desempenho.

aplicadas ao elemento humano no processo [...] (Baeza-Yates, 2023, p. 408, grifo nosso, tradução própria).

Obviamente, ainda segundo esse autor, a responsabilidade é também outra característica humana, mas esse conceito guarda-chuva tem a vantagem de abranger em seus sentidos tanto ponderações éticas quanto de *accountability*, o que, portanto, o viabilizaria como integrador dos sentidos que contornam a ideia de IAR demarcando a agência humana no realizar do seu processo.

Nesse cenário, ainda conforme Baeza-Yates (2023), há um número crescente de organizações públicas e privadas se esforçando para compreender o que realmente é a IAR, bem como os caminhos para aderir e implementar os seus princípios na concepção, implementação e ou uso dos seus sistemas de IA próprios ou não.

Neste ponto, antes de avançar, é importante alinhar e reforçar a seguinte compreensão: algumas problemáticas da IA estão fora do escopo das organizações. Logo, requerem ações regulatórias governamentais como, por exemplo, o impacto no campo do trabalho, usos maliciosos e criminosos da IA – adoção indevida de *deep fakes*<sup>10</sup>, violação de privacidade etc. Já outras podem ser geridas nos espaços das organizações que desenvolvem e ou usam sistemas de IA no cotidiano das suas atividades e relacionamentos. Posto isso, a seguir abordam-se, com mais ênfase, alguns princípios-chave da IAR, considerando as suas implicações no social, para, posteriormente, finalizar este texto refletindo sobre como os profissionais da comunicação, a partir dos espaços das organizações, poderiam colaborar com os processos de construção de uma IAR no Brasil.

Para Baeza-Yates (2023), entre os princípios que, de modo geral, sinalizam o que a IA responsável pode ser constam: legitimidade, equidade (*fairness*), transparência, *accountability* (prestação de contas), robustez, privacidade, segurança e sustentabilidade. Esse autor também agrupa a esse quadro os princípios cunhados pela As-

sociation for Computing Machinery, em 2022: a) legitimidade e competência; b) minimizar danos; c) segurança e privacidade; d) transparência; e) interpretabilidade e explicabilidade; f) capacidade de manutenção; g) contestabilidade e auditabilidade; h) *accountability* e responsabilidade; e i) limitar/conter os impactos ambientais.

Dignum (2020) sugere os seguintes princípios para abranger a ideia e as práticas de IAR: *accountability*, responsabilidade e transparência (ART). Essa autora ainda ressalta que tais princípios também podem enquadrar todo o sistema sociotécnico da IA. Isto é, a condução deles pode orientar um tratamento sociotécnico para

o projeto, a implantação e o uso de sistemas, entrelaçando soluções de software com governança e regulamentação. Além disso, embora cada um dos princípios do ART possa ser aplicado a todos os aspectos dos sistemas de IA, cada um é imperativo para uma característica específica (Dignum, 2020, p. 408, tradução própria).

Logo, de acordo com Dignum (2020), não se pode ter autonomia dos sistemas de IA sem alguma forma de responsabilidade, interação sem *accountability* ou adaptabilidade sem transparência.

A ONU (2023), no seu recente relatório *Governing AI for Humanity*, estabelece cinco princípios orientadores: a) a IA deve ser governada de forma inclusiva, por e para o benefício de todos; b) a IA deve ser governada para o interesse público; c) a governança da IA deve ser construída em sintonia com a governança de dados e a promoção de dados comuns; d) a governança da IA deve ser universal, interligada e enraizada na colaboração adaptativa entre as diversas partes interessadas; e) a governança da IA deve estar ancorada na Carta das Nações Unidas, no Direito Internacional dos Direitos Humanos e em outros compromissos internacionais acordados, como os Objetivos de Desenvolvimento Sustentável. Em ação complementar, em 2024, a ONU aprovou uma resolução sobre a governança da IA, com o

<sup>10</sup> Slee explica que as "técnicas de aprendizagem profunda chamadas 'Redes Adversariais Generativas (GANs)' tornaram-se excelentes na geração de imagens, vídeos ou textos que parecem ter sido criados por humanos ou retratam artefatos do 'mundo real'. Esses usos foram agrupados sob o nome de '*deep fakes*' (Slee, 2020, p. 112, tradução própria).

objetivo de assegurar que a tecnologia crie um mundo mais seguro e equitativo (ONU [...], 2024).

Alguns conjuntos de princípios de IAR também vêm sendo delineados, discutidos e abordados em diretrizes governamentais de diversos países, a exemplo do Brasil (EBIA e PBIA), com vistas a subsidiar futuras regulações sobre a IA. Por exemplo, em 2023, tendo o Brasil como signatário, a Reunião de Altas Autoridades sobre Direitos Humanos do Mercosul aprovou a *Declaração de Princípios de Direitos Humanos no âmbito da IA no Mercosul*. Alguns dos principais pontos da declaração abarcam o desenvolvimento responsável da IA, a garantia de igualdade e não discriminação, a promoção da educação e formação digital, a eliminação de segmentos étnico-raciais na IA e a utilização de tecnologias de reconhecimento facial com transparência, privacidade e sem segmentos raciais (Em plenária [...], 2023).

O governo dos EUA, em 2022, apresentou o *Blueprint for an AI Bill of Rights* (U.S. Office of Science and Technology Policy, 2023) pontuando cinco princípios-chave, a saber: a) os sistemas de IA devem ser seguros e eficazes; b) devem apresentar proteções contra discriminação algorítmica; c) devem assegurar a privacidade de dados; d) devem informar e explicar quando são utilizados e como contribuem para os resultados; e e) aos usuários devem ser oferecidas alternativas humanas aos sistemas de IA, consideração e reserva (poder de recusar os sistemas IA).

Em 2023, vinte e oito países – entre eles China, EUA, União Europeia (EU), Brasil e Arábia Saudita – assinaram a *Declaração de Bletchley* para um desenvolvimento global seguro da IA. O documento foi acordado durante a Cimeira de Segurança da IA, no Reino Unido. Esse documento se configura como mais um esforço global para direcionar coletivamente os desafios e riscos potenciais da IA, bem como inscreve direcionamentos para apoiar a construção de regulamentações convergentes entre os países signatários. Nessa direção, o documento afirma como princípios que, para o bem de todos, a IA deve ser concebida, desenvolvida, implantada e utilizada, de uma forma que seja segura, centrada

no ser humano, confiável e responsável (The Bletchley [...], 2023). Nesse sentido, essa declaração ainda orienta aos países signatários que a IA deve ser operada para promover o crescimento econômico inclusivo, o desenvolvimento sustentável e a inovação, proteger os direitos humanos e as liberdades fundamentais, bem como promover a confiança do público em tais sistemas.

Por fim, para além do marco da UE que, como informado, em 2024, aprovou a primeira regulação da IA no mundo – a *AI Act*, que estabelece que os sistemas de IA usados no bloco devem atender princípios que garantam: segurança, transparência, rastreabilidade, não discriminação (gênero, sexo, idade etc.) e sustentabilidade –, se registra também a participação da UE, bem como Reino Unido e mais onze países (Argentina, Austrália, Canadá, Costa Rica, Santa Sé, Israel, Japão, México, Peru, Estados Unidos da América e Uruguai) na negociação do histórico primeiro tratado internacional de IA (Conselho [...], 2024). Esse documento acordado “aborda os riscos que a IA pode representar, ao mesmo tempo em que pretende promover a inovação responsável”, considerando os direitos humanos, a democracia e o estado de direito. Contudo, segundo Di Lorenzo (2024), especialistas apontam lacunas e fragilidades nesse tratado, por exemplo, os seus signatários têm a obrigação de cumprir as regras estabelecidas, porém não há nenhum tipo de punição caso elas sejam descumpridas.

Dessa forma, apesar de ser mais um passo significativo para a governança global da IA, essa iniciativa não garante o cumprimento rigoroso das normas que determina. Ela, a exemplo de outras iniciativas (como as observadas neste artigo), de acordo com Verdegem (2021, p. 12, tradução própria), não consegue garantir uma “democratização radical da IA” que, ao impedir desigualdades de poder, transforme a sociedade para melhor e permita a emancipação humana. Nesse sentido, a “democratização radical é necessária para evitar uma situação em que apenas algumas organizações, sejam governamentais ou corporativas, tenham o poder econômico e político para decidir que tipo de IA será desenvolvida

e quais propósitos ela servirá" (Verdegem, 2021, p. 12, tradução própria).

Assim, o caminho imprescindível a ser fortalecido por tais iniciativas, pautadas nos princípios da IAR, é fomentar a elaboração de regulações da IA (legislação mestra) que, conforme sugere a ABC (2023), consiga efetivamente enfrentar o desafio duplo de assegurar proteção à sociedade e não atrasar o desenvolvimento tecnológico, compreendendo desse modo as implicações globais e territoriais demandadas pela IA.

### 3 Pensando a comunicação das organizações nesse contexto

Sensível a tais princípios acerca da IAR que, em síntese, abrangem preocupações societais, legais e técnicas, Dignum (2020) orienta que para assimilá-los, as organizações (públicas ou privadas) precisam, inicialmente, ponderar tanto as questões macro quanto microambientais dos espaços onde atuam e, em conformidade, com tais contextos buscar implementar tais diretrizes nas suas práticas. Por exemplo, do ponto de vista do desenvolvimento e ou implementação de sistemas de IA, uma abordagem responsável pode requerer das organizações novas formas e novos métodos que impliquem internamente a incorporação desses princípios desde o processo de concepção e engenharia dos algoritmos.

Envolvendo também outras áreas das organizações, de modo geral, para a assimilação e integração de tais princípios na cultura organizacional são necessários esforços para a formação e sensibilização de todos os seus públicos internos e externos intervenientes (incluindo pesquisadores, *designers*, programadores, gestores, responsáveis pelas aquisições, usuários e consumidores), que permita a cada um deles compreender e assumir o seu papel no processo amplo e contínuo, que a IAR enquanto compromisso reivindica (Dignum, 2020).

Acredita-se que a construção desses processos de formação e sensibilização, a partir das organizações, que demandam por processos comunicativos estratégicos e relacionamentos assertivos com tais públicos, possa ser liderada

e ou facilitada por profissionais da comunicação como, de modo especial, os(as) relações públicas. Esses(as) profissionais que, tradicionalmente, vêm zelando por assegurar dentro das organizações a construção ética de políticas de responsabilidade e sustentabilidade. Como, por exemplo, a implementação e gestão das diretrizes e práticas da atual tendência de *Environmental, Social and Governance* (ESG), ou Governança Ambiental, Social e Corporativa, que requer das organizações compromissos efetivos tanto com o meio ambiente quanto com a sociedade. As práticas de ESG, inclusive, vêm sendo potencializadas dentro das organizações em linha com os usos de sistemas de IA (Pinsky, 2023).

Mas de que forma essas tarefas de capacitação e sensibilização para incorporar a IAR nas organizações poderia ser iniciada pelos profissionais de comunicação? Como desenvolver uma cultura organizacional que assimile e pratique a IAR sendo, portanto, sensível às questões éticas e aos valores humanos? Para estimular e apoiar a elaboração de reflexões que direcionem exercícios de respostas a tais questões, Dignum (2020) enfatiza que para assegurar a implementação da IAR nos espaços organizacionais é necessário, essencialmente, ação. É preciso desenvolver e orquestrar práticas que permitam efetivamente concretizar os discursos que as listas de princípios contemplam.

Nesse sentido, iniciativas que visem fomentar o desenvolvimento de regulamentações (internas ou externas), a adoção de dispositivos de certificação, auditorias, bem como de códigos de conduta, que abarquem e orientem todos os estágios de desenvolvimento e usos da IA, podem ser observadas como potenciais mecanismos e ferramentas para ajudar a direcionar e materializar práticas efetivas para uma abordagem responsável à IA nas organizações. Em adição a esse conjunto de ações, De Laat (2021) sugere também a criação de comitê de ética, de conselho consultivo, a promoção constante de orientações atualizadas sobre a IA, *checklists*, cursos, capacitação, novas ferramentas de software etc. É oportuno também compartilhar que a

OECD disponibiliza em seu site um rico catálogo de cases, ferramentas e métricas para apoiar o desenvolvimento e a gestão da IA responsável<sup>11</sup> nas organizações ao redor do mundo.

Dignum (2020) orienta que iniciativas como as sugeridas *a priori* devem ser amparadas e organizadas sobretudo considerando três pilares sensibilizadores igualmente relevantes. O primeiro denota que a sociedade, em geral, deve estar preparada (e ou se preparar) para assumir a responsabilidade pelos impactos da IA. Esse eixo, ao posicionar IAR como uma questão prioritária de governança, envolve processos educativos e comunicativos, ao considerar especialmente, não exclusivamente, que os pesquisadores e criadores da IA devem ser capacitados para estarem conscientes da responsabilidade que possuem no desenvolvimento desses sistemas e dos impactos diretos deles na sociedade. Tais processos podem ser fomentados com a criação e fornecimento de materiais pedagógicos e informativos, bem como a elaboração de códigos de conduta para orientar a criação e o manejo da IA. Esse contexto também requer métodos e ferramentas para auxiliar a compreensão e a integração de valores morais, sociais e legais ao desenvolvimento tecnológico da IA etc. Por fim, esse pilar ainda enquadra e reforça o entendimento geral que cabe aos governos e a sociedade civil, coletivamente, construir e determinar as regulações acerca da IA.

O segundo pilar articula que a IAR exige a concepção de processos e mecanismos que permitam continuamente aos sistemas de IA agir e ou responder de acordo com a ética e os valores humanos, tendo em vista que "quer os projetemos dessa forma ou não, os sistemas de IA tomarão e já estão tomando decisões que considerariamos ter um sabor ético se fossem tomadas por pessoas. Estar consciente disso é o que a IA responsável demanda" (Dignum, 2020, p. 220, tradução própria). Já o último pilar, salienta o papel da participação, da colaboração e responsabilidade compartilhada. Ele demarca

a necessidade de se compreender como diferentes pessoas trabalham e convivem com as tecnologias de IA em todas as culturas, a fim de desenvolver estruturas assertivas para uma IAR. Nesse sentido, reforça Dignum (2020), é importante lembrar que a IA não é independente, mas deve ser entendida como parte dos contextos e relações sociotécnicas em todas as suas diversidades e especificidades.

À guisa de conclusão, em linha com as reflexões postas, algumas pesquisas identificadas na literatura internacional amparam e fortalecem as reflexões acionadas neste artigo sobre os impactos da IA responsável nas organizações e junto aos seus públicos internos e externos. Por exemplo, o trabalho de Benjamins, Barbado e Sierra (2019), no qual esses autores desenvolvem um estudo de caso crítico com o objetivo de descrever como uma grande organização, como a Telefónica (24 países, 127.000 funcionários, 343 milhões de acessos, 52 bilhões em receitas) está abordando e gerindo a IA. Nessa direção, os autores apresentam e discutem os princípios para IAR desenvolvidos e implementados pela empresa, bem como a metodologia "Responsible AI by Design" também elaborada pela Telefónica para incorporar tais princípios na sua cultura organizacional.

Outra pesquisa recente *Companies Committed to Responsible AI: From Principles towards Implementation and Regulation?*, desenvolvida por De Laat (2020), inscreve contribuições importantes às discussões em tela ao refletir e atender assertivamente três objetivos: a pesquisa identifica empresas globais que praticam e desenvolvem IA e que recentemente se comprometeram publicamente com os princípios ou diretrizes de IA responsável; pontua se essas empresas introduziram ferramentas de governança interna para a IAR; e sinaliza se elas produziram/publicaram propostas e capacitações internas sobre a governança adequada da IA. Nas discussões críticas desse trabalho, timidamente, o autor aborda o papel das ações e ofensivas de relações públi-

<sup>11</sup> Para consultas: <https://oecd.ai/en/catalogue/overview>. Acesso em: 16 jan. 2024.

cas nesse contexto interpelando, diretamente, as organizações a agirem de maneira aberta e transparente na operação e gestão da IA com apoio desses profissionais de comunicação.

No Brasil, de modo geral, ainda são poucos os trabalhos que abordam de modo similar, especificamente, as expressões da AIR nas organizações. Na comunicação inclusive. Por exemplo, como exercício ilustrativo, uma rápida busca<sup>12</sup> em alguns dos importantes periódicos do campo utilizando os termos "Inteligência Artificial" e "Inteligência Artificial Responsável" obteve-se os seguintes resultados: na revista *Organicom*, da USP, foram identificados 21 artigos sobre IA; na revista *E-Compós*, vinculada à Associação Nacional dos Programas de Pós-Graduação em Comunicação; e na revista *Intercom*, da Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, apareceram apenas três trabalhos que tangenciam o tema. Por fim, na revista *Matrizes*, da USP, há o registro de 24 artigos e dois editoriais que abordam o termo IA, bem como na revista *Famecos*, da PUCRS, tem-se o registro de cinco trabalhos que abordam a IA. Nesses periódicos não há registro de artigo que explore o termo "Inteligência Artificial Responsável".

Na área empresarial brasileira, de modo incipiente, algumas empresas como, por exemplo, a companhia aérea GOL e a empresa de telefonia móvel VIVO, marca comercial da Telefónica no Brasil, começam a compartilhar na mídia os aprendizados e desafios, especialmente, com a IA generativa, bem como o impacto transformativo que essa tecnologia vem gerando em seus negócios. Por exemplo, em evento recente sobre a tendência da IA generativa, ocorrido em São Paulo, no final de 2023, os executivos de ambas as organizações apresentaram publicamente algumas experiências com a IA. Thales Chieco, Head de Tecnologia na GOL, pontou que a adoção da IA generativa está sendo rápida e iniciou na empresa com a implementação de estudos de caso internos, como, por exemplo, a iniciativa de reformular e simplificar para consultas

o código de ética da GOL de 200 páginas. Para essa tarefa, um time multidisciplinar foi reunido e, em 10 minutos, usando o OpenAI, o trabalho foi concluído qualificadamente.

A companhia também está aprimorando a assistente virtual, a Gal, para responder perguntas frequentes, como regras de viagem, mas a proposta é que ela ajude os clientes com buscas e dicas de turismo hipersonalizadas. "Queremos transformar a nossa IA em uma verdadeira agente de viagens", para isso a atenção da Gol está na experimentação e na garantia da segurança desse processo. "A cibersegurança é de suma importância para nós e reflete nosso lema: 'zero erro'", finaliza. (Confira os insights [...], 2023).

Já William Bueno, Head de Produtos IA da Vivo, contou que frente às inovações da IA, a meta em 2023 foi focar em diversos testes com essa tecnologia para desenvolver uma plataforma segura e com governança centralizada. Como próximos passos, a Vivo está trabalhando para materializar a IA no atendimento ao cliente e ganhar eficiência e agilidade. A empresa também lançou o VivoGPT, dando oportunidade aos colaboradores de testar os benefícios do ChatGPT em ambiente controlado, otimizando consultas às informações e processos internos. Igualmente foram realizados testes de análise das interações na central de atendimento telefônico e em pesquisas de satisfação. Como desafios urgentes a serem enfrentados, para implementar a IA dentro da companhia, se têm a qualificação de profissionais, regulamentação e equilíbrio ambiental. Esse último ponto, citado por Bueno, implica entre outras questões, "a energia consumida pelo uso intensivo de IA pelas empresas" (Confira [...], 2023). Essa já é uma questão na agenda dos debates atuais sobre a AI (Saenko, 2023).

Muitas outras empresas vêm, semelhantemente, compartilhando midiaticamente as suas iniciativas com os sistemas de IA, por exemplo, em matérias jornalísticas que convocam: "Confira 10 grandes empresas que usam a inteligência artificial e lucram com ela" (Nascimento, 2023);

<sup>12</sup> Levantamento exploratório realizado nos respectivos sites dos periódicos citados em 4 de nov. 2024.

"12 empresas que utilizam inteligência artificial em suas operações" (12 empresas [...], 2023); "10 empresas que inovam no Brasil usando Inteligência Artificial" (10 empresas [...], 2023); entre outras.

Nesses materiais midiáticos, bem como em seus sites institucionais, rara e vagamente as empresas informam as suas iniciativas e compromissos com foco na gestão da IAR, bem como os seus processos comunicativos nessa direção. Por exemplo, temas e efeitos críticos inerentes ao uso da IA pelas organizações, – como a redução de custos, automação e substituição de mão de obra humana, acúmulo de trabalho, novas exigências inexequíveis aos trabalhadores e consumidores etc. –, são silenciados e não problematizados nesses posicionamentos midiáticos que, via de regra, comunicam posturas acríticas e ideias romantizadas de trabalho e processos otimizados pela parceria entre humanos e máquinas (Brandão, 2023).

Enfim, essas e outras lacunas, além de demarcarem um desafio para a compreensão empírica de tais práticas no contexto brasileiro, se colocam como uma oportunidade significativa e central a ser considerada na agenda de pesquisas e intervenções do campo comunicacional que, a partir de perspectivas críticas, objetivem desenvolver e contribuir com a construção de conhecimentos e o avanço das práticas acerca da IAR no Brasil nos próximos anos.

### Considerações finais

Este trabalho buscou delinear e explicar o conceito de IAR, implicando nesse exercício a comunicação a partir dos espaços das organizações. Nessa tarefa, também foram apresentados alguns dos princípios-chave da IAR, visando inscrever, de modo frutífero, algumas elucubrações e direcionamentos que contribuam para apoiar os estudos brasileiros da comunicação que tenham e ou terão como foco a IAR como objeto de pesquisa e intervenção.

Nesse sentido, espera-se que a articulação teórica desenvolvida para edificar as reflexões deste artigo, que abrange trabalhos de autoras e autores do Sul ao Norte Global, seja um quadro

introdutório que estimule o interesse de pesquisadoras e pesquisadores da comunicação a explorarem em suas investigações os impactos da IA no cotidiano sociocultural de um país, como o Brasil, ainda marcado profundamente pelas desigualdades sociais.

Com efeito, sumariamente, o saber organizado neste texto alinha-se e complementa o pensamento de Dignum (2020). Para essa autora, diante dos contextos de transformações profundas que a IA vem produzindo socioculturalmente, a educação novamente deve desempenhar um papel importante para garantir que o conhecimento do potencial da IA seja generalizado, para conscientizar as pessoas de que elas podem e devem participar na definição do desenvolvimento social e, fundamentalmente, para garantir a diversidade e a inclusão. Nessa perspectiva, se observou neste texto ser imperativo envolver a comunicação e os seus processos midiáticos, que indiscutivelmente também operam, a partir dos seus limites, sentidos pedagógicos significativos na sociedade contemporânea midiatizada (Sodré, 2021). Portanto, implicar a comunicação nesse cenário é ação-chave para potencializar a sensibilização e a conscientização social sobre a implementação e o uso de tais tecnologias, bem como acerca da importância da participação ampla da sociedade civil nos debates e processos de regulamentação, desenvolvimento e impactos dos sistemas de IA. É essa participação social que, conforme Dignum (2020), permitirá direcionar sistemas de IA que representem e sejam acessíveis a todos, bem como sirvam para o bem da sociedade.

Entre as brechas que este artigo deixa, que podem ser exploradas em outras oportunidades e estudos futuros, se ressalta a necessidade de pensar com mais atenção, a partir das perspectivas teóricas e empíricas do Sul Global, as especificidades do Brasil nas discussões sobre a IAR e o lugar da comunicação nesses contextos e debates. Por exemplo, na literatura e documentos analisados e utilizados neste trabalho que versam sobre recomendações e princípios-chave para o avanço da IAR, as questões comunicacionais e midiáticas são raramente abordadas. Santaella

(2022) faz observação semelhante acerca das questões culturais pouco consideradas nessas agendas.

Por fim, acredita-se ser relevante que pesquisas futuras também abordem as dimensões de sentidos da IAR, que vão além dos esforços necessários de regulação, visando explorar as particularidades e os impactos que o avanço dos sistemas de IA vêm produzindo, ou não, nos espaços e dinâmicas sociais que conformam as áreas logotécnicas da comunicação (publicidade, jornalismo, relações públicas, audiovisual etc.). Nessa direção, espera-se que este artigo tenha contribuído, de alguma forma, para facilitar as primeiras reflexões dessas tarefas que convocam a todos e todas diante dos desafios e transformações aceleradas que estão desenhandando os sentidos da comunicação do século XXI.

## Referências

- 10 EMPRESAS que inovam no Brasil usando Inteligência Artificial. In: Terra. [S. l.], 26 set. 2023. Disponível em: <https://bit.ly/3SukgQg>. Acesso em: 4 nov. 2024.
- 12 EMPRESAS que utilizam inteligência artificial em suas operações. Exame, [S. l.], 13 out. 2023. Disponível em: <https://bit.ly/48Lw2Hb>. Acesso em: 4 nov. 2024.
- ABC – ACADEMIA BRASILEIRA DE CIÊNCIAS. Recomendações para o avanço da inteligência artificial no Brasil. GT-IA da ABC. RJ: ABC, 2023.
- ARBIX, G. "Algoritmos não são inteligentes nem têm ética, nós temos": a transparência no centro da construção de uma IA ética. In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. Inteligência artificial: avanços e tendências. São Paulo: IEA-USP, 2021. p. 262-286.
- ALMEIDA, V. IA Generativa e a Web no Brasil. In: IA Generativa: oportunidades, riscos e governança. IS. l.: UFMG: Institute of Advanced Studies of the University of São Paulo, 18 maio de 2023. Disponível em: <https://bit.ly/3OdPfJk>. Acesso em: 4 nov. 2024.
- ARUN, C. AI and the Global South: Designing for Other Worlds. In: DUBBER, M. D.; PASQUALE, F.; DAS, S. (ed.). The Oxford Handbook of Ethics of AI. Oxford: Oxford Academic, 2020.
- ASCHER, W. Forecasting: An Appraisal for Policy-makers and Planners. Baltimore, MD: Johns Hopkins University Press, 1979.
- BAEZA-YATES, R. LECTURE HELD AT THE ACADEMIA EUROPAEA BUILDING BRIDGES CONFERENCE 2022: An Introduction to Responsible AI. European Review, [S. l.], v. 31, n. 4, p. 406-421, 2023.
- BENJAMINS, R.; BARBADO, A.; SIERRA, D. Responsible AI by design in practice. In: HUMAN-CENTERED AI: TRUSTWORTHINESS OF AI MODELS & DATA (HAI) TRACK AT AAAI FALL SYMPOSIUM (ArXiv), 2019, Arlington. Proceedings [...]. Arlington: AAAI, 2019. <https://doi.org/10.48550/arXiv.1909.12838>.
- BOTELHO, T. H. F.; NÖTH, W. Deepfake: Inteligência Artificial para discriminação e geração de conteúdos. TECCOGS, [S. l.], n. 23, p. 69-78, jan./jun. 2021.
- BRANDÃO, M. Brasil lidera automação de processos, mas enfrenta desafios. In: Consumidor Moderno. [S. l.], 5 jun. 2023. Disponível em: <https://bit.ly/3NR7xzJ>. Acesso em: 4 nov. 2024.
- BURRELL, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, [S. l.], v. 3, n. 1, p. 1-12, 2016.
- BRASIL. Senado Federal. Projeto de Lei n.º 2338, de 2023. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023. Disponível em: <https://bit.ly/3YTzOvP>. Acesso em: 4 nov. 2024.
- COMO inteligência artificial impacta empregos no mundo e no Brasil, segundo FMI. In: BBC News Brasil. [S. l.], 15 jan. 2024. Disponível em: <https://bit.ly/3UcwWlh>. Acesso em: 4 nov. 2024.
- CONFIRA os insights de grandes lideranças sobre Transformação Digital impulsionada por IA Generativa. Exame, [S. l.], 21 dez. 2023. Disponível em: <https://bit.ly/3UhJUop>. Acesso em: 4 nov. 2024.
- CONSELHO da Europa aprova primeiro tratado internacional sobre IA. In: Compete 2030. [S. l.], 18 jun. 2024. Disponível em: <https://bit.ly/3CbgfFs>. Acesso em: 2 nov. 2024.
- CORDEIRO, V. D. Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais. In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. Inteligência artificial: avanços e tendências. São Paulo: IEA-USP, 2021. p. 2024-226.
- CORMEN, T. et al. Introduction to Algorithms. Cambridge, MA: MIT Press: McGraw-Hill, 2009.
- COZMAN, F. G.; NERI, H. O que, afinal, é Inteligência Artificial? In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. (org.). Inteligência artificial: avanços e tendências. São Paulo: IEA-USP, 2021. p. 19-29.
- COSTA, A. H. et al. Trajetória acadêmica da Inteligência Artificial no Brasil. In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. (org.). Inteligência artificial: avanços e tendências. São Paulo: IEA-USP, 2021. p. 30-66.
- DE LAAT, P. Companies Committed to Responsible AI: From Principles towards Implementation and Regulation? Philos. Technol., [S. l.], v. 34, p. 1135-1193, 2021.
- Dignum, V. Responsible Autonomy. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 26., 2017, Palo Alto. Proceedings [...]. Palo Alto, CA: AAAI Press, 2017. p. 4698-4704.

DIGNUM, V. Responsibility and Artificial Intelligence. In: DUBBER, M. D.; PASQUALE, F.; DAS, S. (ed.). *The Oxford Handbook of Ethics of AI*. Oxford: Oxford Academic, 2020. p. 30-66.

DI LORENZO, A. Países assinam primeiro tratado internacional de IA; entenda. In: Olhar Digital. [S. l.], 6 set. 2024. Disponível em: <https://bit.ly/4f8oeCw>. Acesso em: 2 nov. 2024.

DOMINGOS, P. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. [S. l.]: Basic Books, 2015.

EM PLENÁRIA final, RAADH aprova Declaração de Princípios [...] In: Ministério dos Direitos Humanos e da Cidadania. [S. l.], 24 nov. 2023. Disponível em: <https://bit.ly/48yStjl>. Acesso em: 4 nov. 2024.

EUROPEAN UNION. *The EU Artificial Intelligence Act*. [S. l.], 2021. Disponível em: <https://artificialintelligenceact.eu>. Acesso em: 4 nov. 2024.

GIL, A. C. *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas, 2008.

GILLESPIE, T. Algorithm. In: PETERS, B. *Digital Keywords*. Princeton: Princeton University Press, 2016. p. 18-30.

JÚNIOR, J. Governo edita medida provisória para recriar Ministério das Comunicações. In: Agência Câmara de Notícias. [S. l.], 12 jun. 2020. Disponível em: <https://bit.ly/3Aiq2dc>. Acesso em: 1 nov. 2024.

KAUFMAN, D.; SANTAELLA, L. O papel dos algoritmos de inteligência artificial nas redes sociais. FAMECOS, Porto Alegre, v. 27, p. 1-10, jan./dez. 2020.

KUNSCH, M. M. K. A comunicação estratégica nas organizações contemporâneas. *Media & Jornalismo*, [S. l.], v. 33, p. 13-24, 2028.

MACINNIS, D. J. A framework for conceptual contributions in marketing. *Journal of Marketing*, [S. l.], v. 75, n. 4, p.136-154, 2011.

MARANHÃO, J.; ABRUSIO, J.; ALMADA, M. Inteligência Artificial e o Direito: duas perspectivas. In: COZMAN, F. G.; PLONSKI, G. A.; NERI, H. *Inteligência artificial: avanços e tendências*. São Paulo: Instituto de Estudos Avançados, 2021. p. 287-310.

MCTI anuncia revisão da Estratégia Brasileira de Inteligência Artificial. In: Ministério da Ciência, Tecnologia e Inovação. [S. l.], 11 dez. 2023. Disponível em: <https://bit.ly/42bLOZx>. Acesso em: 4 nov. 2024.

NASCIMENTO, A. Confira 10 grandes empresas que usam a inteligência artificial e lucram com ela. *O Tempo*, [S. l.], 10 jul. 2023. Disponível em: <https://bit.ly/48HoO7g>. Acesso em: 4 nov. 2024.

NILSSON, N. J. *The quest for artificial intelligence: a history of ideas and achievements*. New York, NY: Cambridge University Press, 2010.

NÃO SE engane – a IA é propriedade das grandes empresas de tecnologia. In: MIT Technology Review. [S. l.], 19 jan. 2024. Disponível em: <https://bit.ly/428PrPl>. Acesso em: 4 nov. 2024.

OECD. *Artificial Intelligence in Society*. Paris: OECD Publishing, 2019. Disponível em: <https://bit.ly/47Tskdg>. Acesso em: 4 nov. 2024.

ONU – Organização das Nações Unidas. Interim Report: Governing AI for Humanity. [S. l.]: ONU, dez. 2023. Disponível em: [www.un.org/en/ai-advisory-body](http://www.un.org/en/ai-advisory-body). Acesso em: 4 nov. 2024.

ONU adota por consenso resolução para regular a inteligência artificial. In: ONU News. [S. l.], 21 mar. 2024. Disponível em: <https://news.un.org/pt/story/2024/03/1829446>. Acesso em: 2 nov. 2024.

PLANO Brasileiro de Inteligência Artificial é apresentado na Finep em palestra do professor Virgílio Almeida. In: Finep. [S. l.], 26 set. 2024. Disponível em: <https://bit.ly/3UBCwn6>. Acesso em: 4 nov. 2024.

PINHO, J. A. G.; SACRAMENTO, A. R. S. Accountability: já podemos traduzi-la para o português? *Revista de Administração Pública*, [S. l.], v. 43, n. 6, p. 1343-1368, nov. 2009.

PINSKY, V. Como a inteligência artificial pode potencializar a agenda ESG nas empresas. *Época Negócios*, [S. l.], 27 nov. 2023. Disponível em: <http://glo.bo/3UcAYRc>. Acesso em: 14 jan. 2023.

PLANO Brasileiro de Inteligência Artificial (PBIA) 2024-2028. – LNCC. In: Gov.br; MCTI Laboratório Nacional de Computação Científica. [S. l.], 7 ago. 2024. Disponível em: <https://bit.ly/4e1CY5P>. Acesso em: 1 nov. 2024.

RÖHE, A. A corrida global pela regulação das Inteligências Artificiais. In: ABES. [S. l.], 4 jan. 2024. Disponível em: <https://bit.ly/3UgvmHa>. Acesso em: 4 nov. 2024.

ROMANINI, V. A inteligência artificial como um objeto da pesquisa comunicacional: a contribuição da semiótica peirceana. In: PEREZ, C. et al. *Comunicação na agenda do século XXI*. São Paulo: Estação das Letras e Cores, 2023. p. 69-88.

SAENKO, K. IA generativa tem seus problemas – e um dos maiores é o impacto ambiental. In: Fast Company. [S. l.], 1 jun. 2023. Disponível em: <https://bit.ly/3Cokan8>. Acesso em: 4 nov. 2024.

SANTAELLA, L. Inteligência Artificial e cultura: oportunidades e desafios para o Sul Global. In: NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR (ed.). *Inteligência artificial e cultura: perspectivas para a diversidade cultural na era digital*. São Paulo: Comitê Gestor da Internet no Brasil, 2022. p. 69-95.

SLEE, T. *The Incompatible Incentives of Private-Sector AI*. In: DUBBER, M. D.; PASQUALE, F.; DAS, S. (ed.). *The Oxford Handbook of Ethics of AI*. Oxford: Oxford Academic, 2020. p. 107-123.

SODRÉ, M. *A sociedade incivil: mídia, iliberalismo e finanças*. Petrópolis, RJ: Ed. Vozes, 2021.

STEURER, R. Disentangling governance: A synoptic view of regulation by government, business and civil society. *Policy Science*, [S. l.], 46, p. 387-410, 2013.

THE BLETCHLEY Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023. In: GOV.UK. [S. l.], 1 nov. 2023. Disponível em: <https://bit.ly/3U7Eo7E>. Acesso em: 4 nov. 2024.

U.S. Office of Science and Technology Policy. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. In: White House. [S. l.], Oct. 2022. Disponível em: <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>. Acesso em: 4 nov. 2024.

VERDEGEM, P. Introduction: Why We Need Critical Perspectives on AI. In: VERDEGEM, P. (ed.). AI for Everyone? Critical Perspectives. London: University of Westminster Press, 2021. p. 1-18.

ZERFASS, A.; VERČIČ, D.; NOTTHAFT, H.; WERDER, K. P. Strategic communication: Defining the field and its contribution to research and practice. International Journal of Strategic Communication, [s. l.], v. 12, n. 4, p. 487-505, 2018.

---

### Francisco Leite

Doutor e mestre em Ciências da Comunicação pela Escola de Comunicações e Artes, da Universidade de São Paulo (ECA-USP), em São Paulo, SP, Brasil, com estágio de doutoramento PDSE/CAPES na Università di Bologna e na Università degli Studi di Trento, Itália; e pós-doutorado em Comunicação, Consumo e DEI pela USP/FAPESP. Professor no Instituto de Educação Continuada da PUC Minas. Vice-líder do grupo de pesquisa ArC2 – Estudos Antirracistas em Comunicação e Consumos, CNPq/USP. Pesquisador do grupo de estudos “Inteligência Artificial Responsável” (divisão IA, Comunicação e Artes) na Cátedra Oscar Sala, do Instituto de Estudos Avançados da USP.

---

### Endereço para correspondência

**Francisco Leite**

IEA-USP

Cátedra Oscar Sala

Rua da Praça do Relógio, 109, térreo

Cidade Universitária, 05508-050

São Paulo, SP, Brasil

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