Universidade de Brasília
Faculdade de Economia, Administração e Contabilidade
Programa de Pós-Graduação em Administração
Curso de Doutorado Acadêmico

GOVERNANÇA DE INTELIGÊNCIA ARTIFICIAL NAS ORGANIZAÇÕES PÚBLICAS: APLICAÇÃO DE FUZZY E CRISP-SET QCA A PROCESSOS E PRÁTICAS QUE CONSIDEREM PRINCÍPIOS ÉTICOS

PATRICIA GOMES RÊGO DE ALMEIDA

Brasília - DF

2023

Universidade de Brasília (UnB)
Faculdade de Economia, Administração e Contabilidade (FACE)
Programa de Pós-Graduação em Administração
Curso de Doutorado Acadêmico

GOVERNANÇA DE INTELIGÊNCIA ARTIFICIAL NAS ORGANIZAÇÕES PÚBLICAS: APLICAÇÃO DE FUZZY E CRISP-SET QCA A PROCESSOS E PRÁTICAS QUE CONSIDEREM PRINCÍPIOS ÉTICOS

PATRICIA GOMES RÊGO DE ALMEIDA

Tese de Doutorado submetida ao Programa de Pós-Graduação em Administração da Universidade de Brasilia como requisito à obtenção do grau de Doutora em Administração.

Data de aprovação: _____de _____ de 2023.

_____

Doutor Carlos Denner dos Santos Júnior – Professor Orientador
Programa de Pós-Graduação em Administração – Universidade de Brasília (UnB)

_____

Doutora Raquel Janissek Muniz – Membro Externo (PPGAdm/UFRGS)

_____

Doutor Pedro Jácome de Moura Júnior – Membro Externo (PPGA/UFPB)

_____

Doutor Herbert Kimura – Membro Interno (PPGA/UnB)

_____

Doutora Eliane Mosconi – Suplente (Business School/Université de Sherbrooke)

Dedico esta tese ao meu pai, Raimundo.

Professor que sempre teve clara sua missão na
luta incessante pela busca do conhecimento.
Pelas lições diárias sobre ética na vida pessoal e profissional, e
imensa generosidade para que outras pessoas tivessem uma vida digna.

AGRADECIMENTOS

À Deus, pela contínua força e luz que me fizeram chegar até aqui,

Aos meus pais, por dedicarem suas vidas à educação das filhas,

Ao meu marido, Nelson, pelo irrestrito apoio em todas as minhas aspirações e desafios, e muito especialmente no decorrer desta tese;

Às minhas filhas Letícia e Luiza, pela compreensão e cuidados comigo nessa jornada;

Ao prof. Carlos Denner, pelas orientações durante o curso de maneira a tornar possível a exploração das diferentes abordagens que a pesquisa sobre a regulação da IA pode proporcionar, e especialmente pela confiança em mim desde o primeiro dia de orientação;

À profª Josivânia Farias, pelas preciosas observações ao longo de todo o meu curso, e pelo apoio nos momentos mais difíceis que passei nessa jornada;

Ao prof. Cleidson Dias, pela generosidade e disponibilidade em esclarecer questões sobre o método;

À profª Solange Alfinito e à Edvânia, pela incomparável presteza nas inúmeras vezes que precisei de apoio do PPGA;

Aos professores Pedro Jácome, Raquel Muniz e Herbert Kimura, pelas contribuições durante minha banca de qualificação;

Aos demais professores do PPGA, que me permitiram construir uma base fundamental para esta pesquisa;

Aos amigos Raquel Dias, Anderson Brandão, Eloísa Torlig, Roberto Barbosa, e Renato Calhau, pela generosidade de compartilharem informações e experiências em benefício desta tese.

A um conjunto de pessoas muito especiais, alguns amigos de longa data, outros conhecidos durante a jornada desta tese, que me auxiliaram na busca por organizações públicas que atendessem aos critérios da pesquisa, no fornecimento de documentos, na participação da avaliação e pré-testes dos instrumentos de coleta de dados, e nas ricas sugestões durante toda a jornada desta tese: prof. André Ponce de Leon , prof. Glauco Arbix, prof. Gerhard Hammerschmid, Ludovic Delépine, Bruno Bazzana, Ahto Saks, Tracey Jessup, Carlo Marchetti, Katrin Sutter, Christian Busch, Frode Rein, Sari Wellinus, Niko Ruostetsaari, Neemias Muachendo, Soufiane BenMoussa, Luis Kimaid, Tiberio Loureiro, Marcus Chevitarese, Márcio Fonseca, Edmundo Andrade, Aymara Neves, Matheus Nascimento, Rodrigo Brandão, Fabricio Santana, Hallisson Rêgo, Avinash Bikha, Andy Williamson, Mariane Piccin, Felipe Lauritzen, Lucia Russo, Julia Maragno, Colin van Noordt, Prateek Sibal, Alex Moltzau, e Deru Schelhaas.

A todos os profissionais das organizações públicas dos dezessete países que generosamente cederam seu tempo para participar da pesquisa, compartilhando conhecimento e experiência para tornar possível ao mundo uma visão empírica do grande desafio que as organizações públicas estão passando.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE EQUAÇÕES

LISTA DE ANEXOS

LISTA DE ABRAVIATURAS E SIGLAS

AIR	Artificial Intelligence Regulation framework
CNJ	Conselho Nacional de Justiça - Brasil
AI HLEG	High-Level Expert Group on Artificial - European Union
IA	Inteligência Artificial
IEEE	Institute of Electrical and Electronic Engineers
ISO	International Organization for Standardization
NIST	National Institute of Standards and Technology - USA
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ONU	Organização das Nações Unidas
QCA	Qualitative Quantitative Analysis
TCU	Tribunal de Contas da União - Brasil
UNESCO	Organização das Nações Unidas para Educação, Ciência e Cultura.
WEF	World Economic Forum
XAI	eXplainable Artificial Intelligence

RESUMO

A natureza ubíqua da inteligência artificial (IA) na sociedade acompanhada pelo aumento de alertas sobre riscos quando o uso e o desenvolvimento de sistemas de IA ocorrem sem que princípios éticos sejam considerados, têm desafiado pesquisadores. Adicionalmente, observam-se evoluções das discussões sobre regulação da IA no sentido de uma governança de IA. Imersa no desafio de regulação da IA, esta tese como objetivo aprofundar-se na multidisciplinaridade envolvida nas questões éticas dos sistemas de IA, levantar as recomendações para melhor implantar regulação e governança de IA, investigar como tais ações têm ocorrido e, o que pode ser feito para contribuir positivamente com tal desafio. O estudo se distribuiu em três etapas e materializadas sob a forma de artigos. Primeiramente, buscou-se na literatura especializada os conceitos e teorias relacionados à regulação da IA, assim como formas em que essa regulação pode ocorrer. Aprofundou-se na análise do *framework AIR-Artificial Intelligence Regulation,* o qual reúne principais instituições de um país em uma governança nacional de IA com protagonismo das instituições públicas do poder Executivo, Legislativo e Judiciário. A regulação proposta pelo AIR combina legislação, normativos governamentais e padrões internacionais para uso e desenvolvimento de sistemas de IA que atendam a princípios éticos. Em uma segunda fase, realizou-se pesquisa empírica de alcance global para investigar como organizações públicas têm incorporado as diretrizes apresentadas pela academia, pela legislação e pelos padrões internacionais, ao seu modelo de governança, de gestão e de desenvolvimento de sistemas de IA de maneira a considerar princípios éticos. Com objetivo exploratório e descritivo, em uma abordagem qualitativa e quantitativa, o estudo considerou vinte e oito organizações públicas, distribuídas em dezessete países, que possuíam sistemas de IA em operação. Questionário e roteiro de entrevista foram elaborados, validados e aplicados em toda a amostra. Por meio de *Qualitative Comparative Analysis* (QCA), nos modos *crisp-set* e *fuzzy*, acrescidos de análise de conteúdo das entrevistas e de documentos, foi possivel conhecer como processos e práticas, previstos na literatura de governança de IA, foram combinados pelas organizações, os problemas enfrentados e as soluções encontradas para superá-los. Os achados fundamentaram a elaboração de um *framework* de governança de IA em uma organização pública: AIGov4Gov. E, em uma etapa final, no terceiro artigo, analisou-se a compatibilidade entre o AIGov4Gov, resultante de casos reais em estágio mais avançado na implantação da governança de IA, e o AIR, modelo conceitual de governança de IA de um país. Os modelos foram considerados compatíveis, o que abre oportunidade para que se considere o foco na implantação da governança de IA em organizações públicas como possível estratégia para impulsionar a implantação de governança de IA nacionalmente. Como contribuições, a pesquisa apresentou formas de regulação e governança de IA de um país, como combinar processos e práticas para viabilizar a implantação da governança de IA de uma organização pública; e por fim, uma proposta de robustecimento da governança de IA de uma nação por meio de implantações bem-sucedidas de governança de IA nas organizações públicas.

Palavras-chave: Regulação da IA, Ética na IA, IA no Governo, IA confiável, Fuzzy QCA.

ABSTRACT

The ubiquitous nature of artificial intelligence (AI) in society, combined with increased awareness of risks when AI systems use and development occur without ethical principles being considered, challenge researchers. Additionally, there are evolutions in the discussions on AI regulation towards AI governance. Immersed in the challenge of AI regulation, this thesis aims to delve deeper into the multi-disciplinarity involved in the ethical issues of AI systems, raise recommendations to better implement AI regulation and AI governance, investigate how such actions have occurred and, what can be designed to make a positive contribution to such a challenge. The study was distributed in three stages and materialized in the form of articles. Firstly, one searched the specialized literature for the concepts and theories involved with AI regulation, as well as the reasons that motivate it and how to regulate the AI. An in-depth analysis of the AI regulation framework (AIR-Artificial Intelligence Regulation) was carried out, which brings together the main institutions of a country in a national AI governance with public institutions of the Executive, Legislative and Judiciary branches playing a leading role. The regulation proposed by AIR combines legislation, government policies and international standards for AI systems use and development that meet ethical principles. In a second phase, a global empirical research was carried out to investigate how public organizations have incorporated the recommendations presented by the academy, by legislation and by international standards, to their governance and management models, including AI systems development in a way to consider ethical principles. With an exploratory and descriptive objective, in a qualitative and quantitative approach, the study considered twenty-eight public organizations, distributed in seventeen countries, that had AI systems in operation. Questionnaire and interview script were elaborated, validated and applied to the entire sample. Through Qualitative Comparative Analysis (QCA), on crisp-set and fuzzy modes, as well as content analysis of interviews and documents, one knew how processes and practices, found in AI governance literature, were combined by organizations; the problems and solutions that they found to overcome such obstacles. The research´s findings supported the development of an AI governance framework in a public organization: AIGov4Gov. And, in a final step, through a third article, one analyzed the compatibility between AIGov4Gov, resulting from real cases at a more advanced stage in the implementation of AI governance, and the AIR framework, a conceptual model of AI governance in a country. The models were considered compatible, and thus, an opportunity opens up to consider the focus on the implementation of AI governance in public organizations as a possible strategy to boost the implementation of AI governance nationally. As contributions, the research presented ways of AI regulating and AI governing in a country, how to combine processes and practices to enable AI governance implementation in public organizations, and finally, a proposal to strengthen a national AI governance through successful deployments of AI governance in public organizations.

Key words: AI regulation, Ethics on AI, AI in government, Trustworthy AI, Fuzzy QCA.

1. **INTRODUÇÃO**

Como muitas tecnologias disruptivas, a inteligência artificial (IA) pode trazer impactos éticos profundos nas decisões e nas ações dela dependentes (Hopster 2021; Bonsón et al. 2021; Wirtz et al. 2022; Kale et al. 2022; Amigoni & Schiaffonati 2018; Schrader & Ghosh 2018; Boden et al. 2017).

A atenção aos dados dos sistemas de IA é apenas um dos passos no caminho complexo para que sejam minimizados riscos de efeitos indesejados com impactos na sociedade (Yapo & Weiss 2018; Jobin et al. 2019; Wirtz et al. 2022), nas instituições (Mika et al. 2019; Beltran 2020), e na humanidade (Jackson 2019b; Beltran 2020). O desafio é amplificado pela dificuldade em se conhecer todo o alcance que os algoritmos de IA podem atingir, adicionada às formas ainda desconhecidas de sua aplicação na vida das pessoas (Fjeld et al. 2020).

Nesse contexto, cresce o número de pesquisas sobre a regulação e governança de IA, fundamentadas na necessidade de se obter os benefícios que a tecnologia pode oferecer, preservando-se princípios éticos no uso e desenvolvimento de sistemas de IA (Cave et al. 2019; Floridi et al. 2020). Observa-se o reconhecimento de que o governo, em sentido amplo, possui grande influência na implantação da governança de IA de um país (Lenardon 2017; Zuiderwijk et al. 2021), fato que gera expectativas de se conhecer como as instâncias de governo têm atuado nesse propósito (Mäntymäki, et al. 2022a; Zuiderwijk et al. 2021).

A complexidade e multidisciplinaridade do estudo (Bonnemains et al. 2018) impuseram a esta tese uma trajetória que resultasse nas respostas às seguintes questões: O que significa regular a IA? Como regular a IA? Como organizações públicas estão adaptando ou implantando seu modelo de governança, de gestão e de desenvolvimento, para produzir sistemas de IA que considerem princípios éticos? Como essas organizações podem auxiliar à implantação da governança de IA de uma nação?

A busca pelas respostas de tais questões sustentaram os seguintes objetivos: a) identificar as formas de regular a IA; b) investigar como as organizações públicas têm incorporado as recomendações sobre governança de IA ao seu modelo de governança e de gestão na produção de sistemas de IA que considerem princípios éticos; c) verificar como organizações públicas podem impulsionar a governança de IA de um país.

O percurso para tais respostas requereu três fases de pesquisa organizadas em três artigos, um ensaio teórico e dois empíricos, sob os titulos: “Regulação da Inteligência Artificial: o papel das

instituições governamentais” , “Governança de Inteligência Artificial nas Organizações Públicas: *fuzzy* e *crisp-set* QCA aplicadas a processos e práticas que considerem princípios éticos”, e “Fortalecimento da Governança Nacional de Inteligência Artificial pelo Foco nas Organizações Públicas”.

---

2. REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL: O PAPEL DAS INSTITUIÇÕES GOVERNAMENTAIS

RESUMO

A posição de destaque que a inteligência artificial (IA) tem ocupado nas estratégias de governo e na indústria de tecnologia vem sendo acompanhada por pesquisas que alertam dos riscos às organizações, à sociedade e ao meio ambiente, quando princípios éticos não são considerados no uso e no desenvolvimento de sistemas de IA. Em consequência, a procura pela regulação da IA tem ocupado destaque na academia, na esfera legislativa e demais instâncias de governo, assim como na própria sociedade. A diversidade de *stakeholders* envolvidos no tema aumenta a complexidade de seu estudo e impõe clareza quanto aos conceitos envolvidos e às atribuições de cada ator no desafio de prover uma regulação sustentável para a IA. Imerso no cenário descrito, o presente estudo buscou aprofundamentos nos conceitos envolvidos quando se refere às questões éticas em sistemas de IA, assim como formas possíveis de regulação da tecnologia. A combinação entre legislação, recomendações governamentais e padrões internacionais foi encontrada como uma necessidade ao conjunto de processos entre *stakeholders* para a regulação de IA em proposta de implantar uma governança de IA nacional. O protagonismo do governo, distribuído nas atribuições próprias do poder Executivo, Legislativo e Judiciário, foi evidenciado durante os achados para maximização de benefícios e mitigação dos riscos da IA, o que coloca as organizações públicas como atores chave para o alcance do desafio de regulação da IA

Palavras-chave: Regulação da IA, Ética na IA, IA no Governo, IA confiável, Inteligência Artificial.

ABSTRACT

The prominent position that artificial intelligence (AI) occupies in government strategies and the technology industry has been monitored by research that warns about the risks to organizations, society and environment, when ethical principles are not considered in the use and development of AI systems. As a result, the search for AI regulation has been highlighted in the academy, in the legislative branch and in other government instances, as well as in society itself. The variety of *stakeholders* involved in such issue increases the complexity of its study and imposes precision on the concepts involved and on attributions of each actor in the challenge of providing sustainable AI regulation. Immersed in the scenario described, the present study sought to deepen the concepts involved when one refers to ethical issues in AI systems, as well as different ways of such technology regulation. The combination of legislation, government recommendations and international standards was found as a need for the set of processes among stakeholders for AI regulation in order to implement national AI governance. The role of the government, distributed

among the Executive, Legislative and Judiciary branches, was evidenced during the findings for maximizing the benefits and mitigating AI risks, which defines public organizations as key actors in achieving the challenge of AI regulation.

Key words: AI regulation, Ethics on AI, AI in Government, Trustworthy AI, Artificial Intelligence

2.1 Introdução

Cada vez mais frequente nas estratégias de governo (Ahn & Chen 2022; Holmström 2022), a inteligência artificial (IA) tem sido uma opção nas propostas de criação de valor público (Scupola & Mergel 2022; Misuraca & Van Noordt 2020).

Observa-se crescente a retórica dos gestores públicos em relação às expectativas da IA nos órgãos governamentais (Dwivedi et al. 2021; Medaglia et al. 2021) associadas à maior eficiência, produtividade (Dwivedi et al. 2021), melhorias na interação com o cidadão (Medaglia et al. 2021), à qualidade de vida das pessoas (Alshahrani et al. 2021; Eke et al. 2023a), ao desenvolvimento das nações e à sustentabilidade global do meio ambiente (Dwivedi et al. 2021).

Os benefícios da IA para as pessoas, as organizações e as nações têm ganhado visibilidade a cada dia (Cerka et al. 2015), e aberto caminhos para que a tecnologia habilite mudanças no comportamento da sociedade (Rahwan et al. 2019; Eyert et al., 2020). Simultaneamente ao aprofundamento das pesquisas nas formas de implantação da IA, crescem os alertas quanto a riscos, quando práticas que preservem princípios éticos não estejam sendo seguidas (Cave et al. 2019; Floridi et al. 2020).

Apesar dos riscos, a incorporação de serviços fundamentados em IA na sociedade é um caminho sem volta que nos impõe uma busca pelo equilíbrio entre a inovação e a regulação (Carter 2020). Nesse sentido, academia, organizações, governos e parlamentos têm se unido para encontrar formas de regular a IA, sem comprometer o ritmo acelerado do lançamento de novos produtos e serviços que possuem algoritmos baseados nessa tecnologia (Villaronga & Heldeweg 2018; Cerka et al. 2015; Larsson 2020; Holder et al. 2016a; Holder et al. 2016b). Diante do desafio exposto, a presente pesquisa objetiva responder à pergunta: “Como regular a inteligência artificial?”

2.2. Motivações para regulação da IA

No intuito de mitigar tais riscos, sistemas baseados em IA precisam ser confiáveis, justos, transparentes (Janssen et al. 2020; Buiten 2019) e seus resultados explicáveis pelos dados (Dwivedi et al. 2021; Andrews 2018).

2.2.1 Inteligência artificial no setor público

A materialização do nível de interesse do setor público no uso da IA pode ser percebida nos investimentos feitos pelos governos ao longo de todo o mundo. Em muitos casos, a própria estratégia de crescimento, desenvolvimento e sustentabilidade do país se confunde com a estratégia de transformação digital e estratégia de IA (Wirtz et al. 2018a).

O expressivo uso de serviços digitais interligados a vários tipos de bases de dados têm habilitado as organizações a considerarem IA como elemento chave de solução de problemas e de oportunidades para agregação de valor (Holmström et al. 2022). Nessa direção, observa-se a presença da IA nas estratégias de TI para atender iniciativas profundamente inovadoras das estratégias corporativas (Kitsios & Kamariotou 2021), assim como nas estratégias de transformação digital de governo, compondo posição de destaque (Holmström et al. 2022).

Uma vez decidindo-se pelo investimento em IA, observam-se mudanças no funcionamento do setor público, impondo sucessivas ondas de adaptações nos processos, nas tecnologias, na estruturas e nas pessoas, iniciando-se com os dois primeiros (Gong et al. 2020). Em consequência, a maturidade do processo de transformação de serviços ao modo digital requer consciência de que comportamento ético passa a ser premissa para a guarda e tratamento dos dados dos *stakeholders* (Saarikko et al. 2020); e assim, as questões éticas ganham relevância estratégica na transformação digital das organizações (Vial 2019).

O exponencial uso da IA nas organizações governamentais pode ser percebido em aplicações, por exemplo, para melhorias nos serviços de transporte (Prakken 2017; Li 2017; Hanna & Kimmel 2017; Ojo et al. 2019), educação (Sharma et al. 2020); saúde (Morley et al. 2020a; Rajkomar et al. 2018; Villaronga & Heldeweg 2018; Goenka & Tiwari 2021; Wang et al. 2016), segurança pública, segurança cibernética (Eggers et al. 2017); compreensão das necessidades e resposta ao cidadão (Ahn & Chen 022); para identificação de problemas dos cidadãos (Ojo et al. 2019); maior eficiência nas operações governamentais; pesquisa e desenvolvimento (Sharma et al. 2020), defesa nacional (Eggers et al. 2017), meio ambiente (Ojo et al. 2019), como exemplos.

Entre os benefícios viabilizados pela IA, destacam-se ganho de eficiência e desempenho, identificação de riscos, redução de custos, aumento da capacidade de processos de análise de dados, base para sustentação de diferentes processos decisórios, ampliação da oferta de serviços, geração de valor à sociedade, engajamento da sociedade, sustentabilidade (Zuiderwijk et al. 2021; Dwivedi

et al. 2021); aumento da qualidade e tempo de vida das pessoas (Morley et al. 2020), aumento da segurança física e cibernética (Eggers et al. 2017; Aaronson & Leblond 2018).

Simultaneamene às implantações de sistemas de IA, crescem as pesquisas e percepções da sociedade sobre a importância de questões ligadas à moralidade e à ética presentes nos serviços baseados em IA (Kazim & Koshiyma 2021; Ashok et al. 2022, Stahl et al. 2022a). Em pesquisa realizada com 566 servidores públicos nos Estados Unidos, Ahn e Chen (2022) constataram que funcionários confiantes de um julgamento ético e moral dos sistemas de IA estavam mais dispostos a adotar a IA no governo. Urge, portanto, necessidade de avanço das organizações públicas em conhecer melhor a inteligência artificial para que possam fazer melhor uso dela (Awad et al. 2020; Ma et al. 2018).

Desde 1956, quando cunhado o termo "Inteligência Artificial", várias conotações têm sido atribuídas a ele (Cerka et al. 2017; Jackson 2019a) para referenciar sistemas que se propõem a "pensar" e agir como humanos, e a sistemas que "pensam" e agem racionalmente (Cerka et al. 2015; Russell & Norvig 1995). Considerando a longa relação de nomes utilizados pela literatura quando se refere à tecnologia de inteligência artificial (IA) – robôs, sistemas inteligentes, agentes inteligentes, agentes de IA, algoritmos de IA, algoritmos inteligentes, sistemas autônomos, entre outros; para o presente estudo, é considerado o conceito estabelecido pela Comissão Europeia, por meio do Grupo de Especialistas em IA (AI HLEG 2019a p.1) para sistemas de IA: "*software (e possivelmente hardware) projetados por humanos que, dado um objetivo complexo, agem na dimensão física ou digital percebendo seu ambiente por meio da aquisição de dados, interpretando os dados coletados, estruturados ou não, raciocinando sobre o conhecimento, ou processando as informações derivadas desses dados, e decidindo pelas melhores ações a serem tomadas para atingir o objetivo determinado. Os sistemas de IA podem usar regras simbólicas ou aprender um modelo numérico, adaptando seu comportamento e analisando como o ambiente é afetado por suas ações anteriores*."

2.2.2 Riscos associados ao uso e provimento de sistemas de IA

Na medida em que se avança na aplicação da IA em diferentes dimensões e por meio de desenvolvimento em novas abordagens, percebe-se o seu poder de transformação social (Boyd & Holton 2018) e a sua natureza ubíqua na sociedade (Zuiderwijk et al. 2021; Rahwan et al. 2019,

Liang & Acuna 2020; Ma et al. 2018; Rahwan et al. 2019; Morley et al. 2020a; Rajkomar et al. 2018).

Paralelamente, crescem as identificações de riscos associados ao uso, ao desenvolvimento e à implantação inadequados (Conitzer et al. 2017; Firth-Butterfield 2017; Wirtz et al. 2022), como decisões enviesadas (Buiten 2019; Benjamins & Garcia 2020), discriminação (Yapo & Weiss 2018; Borgesius 2018; Jobin et al. 2019), perda de privacidade (Jackson 2019b; Jobin et al. 2019), perda de autonomia (Beltran 2020), danos (psicológicos, financeiros e físicos) (Beltran 2020; Jobin et al. 2019), perda de controle (Wirtz et al. 2022), perda ou decréscimo de direitos humanos (Beltran 2020), desemprego em massa (Benjamins & Garcia 2020; Wirtz et al. 2022), erros de julgamentos em processos judiciais (Yapo & Weiss 2018; Jackson 2019b), redução da democracia (Mika et al. 2019); aumento da concentração de poder e riquezas (Wirtz et al. 2022), terrorismo (Beltran 2020), crimes digitais (Jobin et al. 2019), manipulação da opinião das pessoas (Yapo & Weiss 2018), diagnósticos médicos errados (Morley et al. 2020a; Rajkomar et al. 2018), aumento da potência de armamento bélico e do poder de destruição nas guerras (Beltran 2020; Chmielewski 2018), entre outros.

Um dos sutis riscos refere-se à manipulação da opinião de pessoas por meio de *fake news (*Choraś et al. 2021; Zhang & Ghorbani 2020), *deep fakes* (Agarwal & Farid 2019; Stuurman & Lachaud 2022), atividades automatizadas em redes sociais (Wirtz & Müller 2018b) ou censura automatizada (Doneda et al. 2016). Adicionalmente à perda de privacidade no mundo conectado (Liu et al. 2022b), os tênues limites da privacidade digital têm sido continuamente fragilizados com as potentes análises realizadas nas redes sociais, nos dados dos serviços comerciais e até dados pessoais em organizações públicas (Roorda 2021; Liu et al. 2022b).

A percepção de insegurança em relação a algo ainda não completamente dominado e em plena evolução, alinha-se à percepção de que um sistema autônomo, inevitavelmente, encontrar-se-á em uma situação na qual, além das regras pré-estabelecidas, precisará tomar decisões complexas na dimensão ética (Dennis et al. 2016).

2.2.3 Vieses

Antecipando-se ao foco em vieses na tomada de decisões, Herbert Simon (1955; 1956) argumentou que erros sistemáticos ocorrem na tentativa de simplificar situações reais em modelos teóricos, e de que tais ações dependem do nível de conhecimento do tomador de decisões.

Seguindo caminho semelhante, Kahneman et al. (1982) sugeriram estudos mais profundos sobre como os métodos estatísticos poderiam lidar com o ambiente complexo e as limitações do raciocínio indutivo humano. Posteriormente, Kahneman (2003) pesquisou diversas situações de limitações do pensamento intuitivo em julgamentos embutidos nas tomadas de decisões, e como os resultados poderiam se alterar em grupos sociais diferentes.

Em uma dimensão estatística, erros em sistemas preditivos podem ocorrer, haja vista serem a diferença entre os valores previstos como resultado do modelo e o valor real das variáveis consideradas na amostra. Quando o erro ocorre sistematicamente em uma direção ou para um subconjunto de dados ou subpopulação específica, pode-se identificar uma situação de viés no tratamento dos dados (González et al. 2020; Campitelli & Gobet 2010); portanto, não se trata de fenômeno exclusivo de sistemas de IA (Ntoutsi et al. 2020). Em uma abordagem mais ampla, o viés é uma sistemática diferença no tratamento de objetos, de pessoas, ou de grupos em comparação com outros (ISO 2021a); como consequência, ocorre um desequilíbrio na distribuição dos dados em classes (Vetrò et al. 2021).

A aprendizagem dos sistemas de IA com os dados do mundo real gera a possibilidade de que os modelos de *machine learning* possam aprender, ou mesmo ampliar vieses existentes (Borgesius 2018; Vetrò et al. 2021; Ntoutsi et al. 2020).

Vieses em sistemas de IA nascem de vieses cognitivos humanos, ou de características dos dados utilizados (ISO 2021a). Vieses cognitivos são erros sistemáticos em julgamentos ou decisões comuns aos seres humanos devido às limitações cognitivas, fatores motivacionais e/ou adaptações ao ambiente natural (Kliegr et al. 2021). Conscientes ou não, os seres humanos acumulam vieses cognitivos ao longo da vida (Kliegr et al. 2021), apresentando-se de várias formas: a) viés de automação, quando há uma valorização das conclusões por algoritmos superior às análises feitas por humanos (Hickman 2020, Strauß 2021; Jahn & Kordyaka 2019); b) viés de atribuição de grupo, assumindo tudo que ocorre para um indivíduo como verdade para todos; c) viés implícito, em que humanos associam situações ao seu próprio modelo mental de representar aquela situação (Lin et al. 2021); d) favoritismo intragrupo, com parcialidade aos aspectos existentes ao grupo a que se pertence (Jahn & Kordyaka 2019); e) favoritismo a grupos externos ao grupo a que se pertence; f) viés social, quando muitos indivíduos de uma sociedade ou comunidade compartilham do mesmo viés; h) viés de regras e sistemas, quando desenvolvedores habituados com algumas regras embutidas nos sistemas, tentam reproduzir as mesmas regras para representar outras situações; i)

viés de requisito, com a assunção de que todas as pessoas ou situações estão aptas ou se enquadram em mesmos requisitos técnicos (hardware e software) (ISO 2021a).

Os vieses de dados são mais frequentes em situações associadas a *design* e restrições técnicas; podendo ser encontradas não somente em sistemas de IA, apesar de fortemente ampliada nesses casos (ISO 2021a). Um dos maiores grupos são os vieses estatísticos, podendo ocorrer na seleção dos dados - viés de seleção (Baeza-Yates 2018; Mehrabi et al. 2019); imprecisão de dados devido a falhas na entrada manual de dados (Roselli et al 2019); dados obsoletos, geralmente quando grandes volumes são carregados localmente para prover rapidez ao acesso (Roselli et al 2019); na escolha de variáveis confusas, que influenciam variáveis dependentes e independentes (ISO 2021a); na correspondência inadequada entre variáveis desejadas e variáveis disponíveis (Roselli et al 2019), e na classificação inadequada de normalidade a amostras que não o são. O viés de seleção pode ser um viés de amostragem - dados não foram selecionados randomicamente; viés de cobertura ocorre quando uma população selecionada não corresponde à intenção de predição; e o viés de participação, quando pessoas de certos grupos decidem não participar da amostra (ISO 2021a).

Ainda entre os vieses de dados, encontram-se: viés na rotulação dos dados (Lloyd 2018); viés de amostras não representativas durante a seleção de dados para treinamento (Silberg & Manyika 2019); dados omissos de um específico grupo (Mehrabi et al. 2019); processamento inadequado de dados tentando corrigir dados omissos, correção de *outliers*, correção de erros, paradoxo de Simpson, no qual comportamentos de cada grupo de dados desaparece quando os grupos são combinados; agregação de dados de grupos com distribuições diferentes; e treinamento de dados distribuídos (Ashokan & Haas 2021).

Há ainda, um conjunto de vieses associados à engenharia de software para o sistema de IA, os quais têm, em parte, origem em vieses de cognição humana, e outra parte, dos próprios dados, como: engenharia usada para algumas funcionalidades - conversão de dados e redução de dimensões; algumas combinações de algoritmos (submodelos que utilizem árvores de regressão rasa, algumas construções de redes neurais, por exemplo) (Baeza-Yates 2018; ISO 2021a).

A busca por elementos causadores de vieses geralmente está associada à busca por elementos causadores de decisões injustas (Silberg & Manyika 2019). No contexto de um sistema de IA, a caracterização de justiça de uma decisão está diretamente relacionada aos impactos que gera nos indivíduos, nos grupos, nas organizações e na sociedade (ISO 2021a). Convém o registro de que

vieses nem sempre resultam em predições ou decisões injustas, e decisões injustas nem sempre são causadas por vieses (ISO 2021a). Contudo, o fato de não haver uma percepção universal de justiça, a complexidade da análise requer ser inserida em contextos que considerem a cultura, a geografia, a política, a temporalidade, entre outros fatores (ISO 2021a; Wirtz et al. 2022).

Admitindo-se que nem sempre é possível a eliminação plena de vieses (Leavy et al. 2020), e considerando que a minimização de vieses não é uma tarefa apenas tecnológica (Strauß 2021); a complexidade das soluções de sua minimização requer envolvimento de pessoas com habilidades e conhecimentos diferentes em diversos momentos do ciclo de vida de um sistema de IA.

Diante da grande variedade de riscos e impactos, e da amplitude que eles podem alcançar, estratégias convencionais de gestão de riscos das organizações passam a requerer revisão de maneira a poder captar novas variáveis, como por exemplo, aquelas associadas à ética (Aiken 2021; Roorda 2021), e seus impactos na sociedade (ISO 2021a). Nesse sentido, ampla discussão tem sido levantada na retórica de construção de uma "Boa IA", tendo-se como base a ética "*by design*" (Smuha 2021; Kazim & Koshiyma 2021), cujas ações requerem mais clareza nas causas e impactos (Aiken 2021) em todo o ciclo de vida dos sistemas de IA (Vetró et al 2021).

2.2.4 Moralidade, Valores e Ética

A moralidade é o conjunto de modelos comportamentais, valores, regras e normas sociais aceito dentro de um grupo ou da sociedade (Stahl 2008; Besio & Pronzini 2014). Julgamentos morais podem ser aplicados a pessoas, organizações e nações quanto às suas ações; sendo algumas vezes, motivação para a busca de normativos institucionais (Besio & Pronzini 2014).

A ética é a reflexão e justificativa que sustentam essas regras e normas (Stahl 2008; Besio & Pronzini 2014), e ocorre em uma atitude mais consciente das escolhas e, menos volátil (Weiss 1942). Uma visão mais ampla do conceito de ética requer análise de uma estrutura distribuída em camadas de conceitos relacionados, na qual a base, em um nível mais raso, é representada pela intuição moral do que consideramos certo ou errado. Quando se procura entender as razões de tais intuições morais, analisam-se as convicções morais, de posse das quais argumentamos que devemos sempre fazer isso ou aquilo. Ainda requerendo mais respostas, em uma camada superior, encontram-se as justificativas éticas para as convicções sustentadas. E, no nível de maior abstração, residem as reflexões gerais sobre as razões de uma justificativa ser aceita (Stahl 2008).

Então, apesar da distinção entre moralidade e ética, tais conceitos não são excludentes (Weiss 1942), tendo, inclusive, sido objeto de estratégias e pesquisas para que sejam encontradas ou perseguidas juntas (Besio & Pronzini 2014).

Originária do Latim, a palavra "valor" significa ser forte ou ser valioso. Provavelmente, seja esta a razão de geralmente se atribuir "valor" ao que é desejável possuir. No entanto, valores surgem de fenômenos de percepção (Kelly 2011); logo, admitem axiomas positivos e negativos. A busca por identificar valores em produtos e serviços utiliza-se de teorias éticas, visto estas serem *insights* sobre a forma de julgamento das pessoas em relação a um assunto ou fato dentro de um padrão comportamental geral ou universal (IEEE 2021a).

Entre as teorias éticas mais utilizadas na análise de sistemas de IA, destacam-se: a ética Teleológica associada ao consequencialismo, busca maximizar benefícios e minimizar prejuízos) (Baumane-Vitonina 2016; Thimbleby 2008); a ética Deontológica, onde as ações são guiadas pelo dever, não sendo desconsiderada a remuneração como um fator motivador (Baumane-Vitonina 2016); e a ética das Virtudes na qual características específicas de uma pessoa garantem a escolha correta diante de dilemas morais (Schrader & Ghosh 2018; Baumane-Vitonina 2016).

Em um movimento mais recente, a Teoria Comunitarista, fundamenta-se na influência de grupos da sociedade, comunidades, nos valores dos indivíduos. Nesse sentido, os valores comunitários modelam a base das obrigações de responsabilidades dos cidadãos (Schrader & Ghosh 2018). Igualmente recente, a Ética da Florescência argumenta que o ser humano necessita de autonomia de pensamento e deliberação, de convivência em comunidades, e que era necessário resolver anomalias identificadas nas teorias éticas anteriores (cada teoria rejeita as demais), situações não respondidas por algumas teorias, e a dificuldade de se aplicar os princípios éticos em agentes não humanos (Bynum 2006).

Em uma linha paralela, nascida de pesquisadores envolvidos com a ciência da computação, James Moor (1985, 1998, 1999) também buscava elementos para uma ética aplicável ao que estava sendo produzido computacionalmente. Buscava analisar impactos sociais e éticos da tecnologia da computação e formular políticas para o uso ético da tecnologia da computação. Nos resultados de Moor, destacam-se princípios similares aos de Liberdade e Benevolência (Bynum 2006). E, muito recentemente, com uma proposta de Ética da Florescência Geral, Floridi (2003, 2006) defende uma teoria ética para a era da Informação – a Teoria da Informação, na qual coexistem agentes humanos, agentes da natureza e agentes não humanos.

2.2.5 Temas abordados por princípios éticos aplicados aos sistemas de IA

Princípios são declarações abstratas sobre ação ou comportamento, que representam resumidamente um consenso. Quando guiados por princípios éticos, sistemas promovem comportamentos mais aceitos para o contexto projetado (Anderson & Anderson 2018).

Na crença de que os benefícios da IA merecem os esforços para mitigar os seus riscos, crescem os estudos na defesa de uma IA confiável (Kale et al. 2022; Medgalia et al. 2021; AI HLEG 2019), cujas bases estejam em princípios éticos (Amigoni & Schiaffonati 2018; Schrader & Ghosh 2018; Boden et al. 2017; Wright & Schultz 2018; Donahoe & Metzger 2019; Millar 2016; Bonnemains et al. 2018; Yeung et al. 2019). Nessa direção, governos e organizações têm se articulado na elaboração de guias éticos a serem aplicados em um projeto ou serviço que utilize a IA (tabela 1).

Tabela 1: Exemplos de iniciativas de definição de princípios éticos para IA.

Princípios Éticos para Sistemas de IA

Governo - princípios para todo o governo ou de uma organizaão pública	Alemanha	German Federal Government (2017)
	Australia	Australian Government (2019b)
	Brasil	Conselho Nacional de Justiça (2020) LIAA-3R (2022)
	Canada	Toronto (2020) Government of Canada (2022a) University of Montreal (2018)
	Emirados Árabes Unidos	Dubai Government (2019)
	Chile	Gobierno de Chile (2020)
	China	China Academy (2021)
	Comissão Europeia	AI HLEG (2019b)
	Dinamarca	Danish Government (2019)
	Espanha	Government of Spain (2020)
	Estados Unidos	White House (2020) US DoD (2018)
	Finlandia	Vero (2019)
	Hungria	Ministry for Innovation and Technology (2020)
	india	NITI Aayog (2021) NITI Aayog (2022)
	Japão	Japanese Cabinet Office (2019)
	Noruega	Norwegian Ministry of Local Government and Modernisation (2020)
	Reino Unido	Government of United Kingdom (2020a)
	Singapura	Monetary Authority of Singapore (2019)
	Suíça	Switzerland Federal Council (2020)
Instituições intergovernamentais	Continente europeu	Council of Europe (2018)
	OCDE	OECD (2019)
	UNESCO	UNESCO (2019, 2020, 2021)
	World Economic Forum	World Economic Forum (2020d)
Associações de Profissionais e Organizações	AI4People	AI4People (2018)
	Future of Life Institute	Future of Life (2019)
	Partnership on AI	Partnership on AI (2016)

Fonte: Elaboração própria.

Como cada uma das iniciativas citadas apresenta seu próprio portfólio de princípios, pesquisas identificaram um conjunto comum de temas que agrupam guias de princípios éticos, à luz do que a academia, os órgãos reguladores e padrões internacionais têm pesquisado e discutido. Seguem alguns temas mais comuns nas dezenas de guias estudados por Fjeld et al. (2020).

2.2.5.1 Privacidade

Constante em grande parte das leis que tratam de direitos humanos, a privacidade é fortemente impactada pela IA (Janssen et al. 2020; Stahl et al. 2022a) em todas as fases do ciclo de vida do sistema de IA (European Commission 2021a), muitas vezes associada à segurança digital (Kuziemski & Misuraca 2020; Alshahrani 2021; Rhahla et al. 2021). Como consequência, torna-se necessária a existência de mecanismos que viabilizem o controle nos dados pessoais, assim como capacitação nas organizações para fazer os tratamentos dos dados, respeitando a privacidade das pessoas (Stahl et al. 2022a; Rhahla et al. 2021; Jackson 2019b). E, por fim, a existência de lei específica para proteção dos dados pessoais (Fjeld et al. 2020; EU GDPR 2016; Ruijer 2021).

2.2.5.2 *Accountability*

Em abordagem mais ampla, no contexto da IA, *accountability* trata das estruturas de governança, procedimentos e ferramentas utilizadas para avaliar e responsabilizar pelos sistemas de IA (Sikstrom et al. 2022).

Sob o tema *accountability*, estão os princípios associados ao estabelecimento claro de que as responsabilidades por ações e decisões anteriormente feitas apenas por humanos, transferidas para sistemas de IA, (Wirtz et al. 2022; Zuiderwijk et al. 2021), com destaque a se viabilizar condições de que possam ser produzidos os mesmos efeitos, quando submetidos às mesmas condições, com vistas a serem avaliados e auditados (Fjeld et al. 2020). A autoavaliação é sempre o primeiro passo, contudo, dada a previsibilidade de conflitos internos, a auditoria externa torna-se necessária (Zicari et al. 2021). Auditoria de IA tem sido proposta como uma das ferramentas para operacionalização e avaliação da Governança da IA (Raji et al. 2020; González et al. 2020; Wirtz et al. 2022). No entanto, as práticas de auditoria de IA estão em estágio embrionário, ainda carente de definições de escopo, de métricas abrangentes aos princípios éticos e de robustez estabelecidos e, modelos que permitam uma padronização (Mökander & Foridi 2021).

Como os sistemas de IA não têm reponsabilidade por seus atos; sendo essa atribuição dos humanos que fazem parte de sua cadeia produtiva (Boden et al. 2017); um dos primeiros requisitos é o conhecimento da toda cadeia produtiva do produto ou serviço de IA, da concepção até estar disponível aos usuários (González et al. 2020). Como regra geral, a responsabilidade deve ser compartilhada por todos (Nevejans 2016; Jackson 2019a); contudo, somente a rastreabilidade permitirá identificar os exatos fatores causadores dos problemas, distinguindo erros de projetos, erros de execução, erros descobertos e não notificados, ou até erros notificados e não considerados pelos tomadores de decisão (Zuiderwijk et al. 2021; Gasser & Schmitt 2019).

2.2.5.3 Segurança

A segurança tem sido abordada sob duas perspectivas: a necessidade de evitar falhas que possam causar danos; e a necessidade de proteger os sistemas de IA dos ataques externos (Xue et al 2020; European Union Agency for Cyber Security 2021; AI HLEG 2019b). Para o primeiro caso, procura-se a mitigação de danos físicos, psicológicos, sociais e materiais às pessoas e ao meio ambiente (AI HLEG 2019b). Na segunda abordagem, a defesa de ataques aos sistemas de IA impõe ciclo completo de gestão da segurança, amparado por normativos, que considerem desde a prevenção e recuperação até pleno restabelecimento, impondo robustos processos de gestão de riscos (NIST 2022; Jing et al. 2021; Breier et al. 2020). Além dos tradicionais processos e mecanismos para segurança digital, observam-se novos tipos de vulnerabilidades próprias dos sistemas de IA, o que inclui os dados por eles utilizados (Xue et al. 2020, European Union Agency for Cyber Security 2021; Rhahla et al. 2021; Eggers & Sample 2020; McGraw et al. 2020).

2.2.5.4 Transparência

Sob a perspectiva da governança, um dos maiores desafios é suplantar a opacidade dos algoritmos de IA (Tutt 2017; Butterworth 2018; Buiten 2019). A "explicabilidade" das decisões passa a ser requisito a ser buscado em diferentes e complementares caminhos (Zuiderwijk et al. 2021); desde a clareza do propósito dos algoritmos, parâmetros utilizados, a abertura dos dados utilizados (González et al 2020), a priorização por algoritmos de código aberto com as vantagens de comunidades de colaboração (De Silva & Alahakoon 2022), até mesmo os processos de contratação por meio de editais públicos (Fjeld et al. 2020). Os instrumentos utilizados devem, dentro do possível, permitir o rastreamento de todo o ciclo de vida do dado, e das decisões dos

sistemas de IA (AI HLEG 2019b; Kozuka 2019; González et al 2020), de maneira a viabilizar efetivamente possíveis contestações das decisões baseadas na IA (AI HLEG 2019b).

Além da transparência do código e dos dados utilizados, faz-se importante uma forma de representação dos valores morais utilizados nas regras de decisão implantadas no sistema de IA (Rubicondo & Rosato 2022), assim como dos dilemas éticos envolvidos nas variáveis consideradas (Bonnemains et al. 2018; Anderson & Anderson 2018; Bench-Capon & Modgil 2017).

Entre os mecanismos associados à codificação para reduzir a opacidade algorítmica, o padrão *eXplainable Artificial Intelligence* (XAI) ganhou destaque como um conjunto de ferramentas, algoritmos e técnicas de *machine learning* com proposta de explicar as decisões e conclusões efetuadas por um sistema de IA (Das 2020; Dazeley et al 2021; Adadi & Berrada 2018; Phillips et al. 2021) para habilitar humanos a conhecerem e efetivamente gerenciarem a geração de novos padrões de IA (Arrieta et al 2020). O XAI se sustenta nos princípios de: explicação (o sistema deve gerar evidências ou razões de seus resultados); significado (explicações devem ser compreensíveis para os seus usuários), acurácia explicada; e limites de conhecimento (somente deve operar nas condições para as quais foi projetado) (Phillips et al. 2021).

2.2.5.5 Justiça, diversidade e não discriminação

A justiça, no escopo de IA, pode ser sustentada pelos pilares: transparência, imparcialidade e inclusão (Silkstrom et al. 2022). Decisões justas e não discriminatórias pressupõem dados e algoritmos isentos de vieses (Ashokan & Haas 2021; Silkstrom et al. 2022; Domnich & Anbarjafari 2021). Tanto os dados representativos e de boa qualidade (Leavy et al. 2020, Ntoutsi et al. 2020), quanto regras algorítmicas justas, passam a ser imperativas (Strauß 2021). O desafio de tais princípios se amplia na medida em que ocorrem não somente falhas em procedimentos e dados; mas também, vieses que diversos atores possuem e nunca foram percebidos por eles ou por quem toma decisões na organização (Kliegr et al. 2021; ISO 2021a).

Normativos, processos, procedimentos técnicos e de gestão passam a ser cruciais para minimizar ou eliminar vieses ou qualquer tratamento injusto (Ashokan & Haas 2021; Oneto & Chiappa 2020; González et al. 2020, Tomalin et al. 2021; Liu et al. 2022a). Para tal, o recrutamento de equipes multidisciplinares e com diversidade de perfis (etnia, idade, raça, cor, religião, gênero, preferência sexual, preferência política, nacionalidades) torna-se ação basilar (Dignum 2022).

2.2.5.6 Controle humano

Do reconhecimento de que a academia e a ciência ainda não têm domínio sobre todas as possibilidades de uso e impactos de sistemas de IA (Fjeld et al. 2020; Scherer 2016; Wirtz & Müller 2018b); surge a necessidade de se preservar a autonomia humana com princípios para garantir que, em situações nas quais as decisões não estiverem adequadas, ainda exista a possibilidade do humano responsável pelas decisões possa ter controle sobre a tecnologia, ou mesmo rever as decisões feitas por ela (Wirtz et al. 2022; Jobin et al. 2020; Jackson 2019b, Hickman 2020).

A supervisão humana pode ser implantada por meio de mecanismos como *human-in-the loop* (HITL), *human-on-the-loop* (HOTL) ou *human-in-command* (HIC) (Hickman 2020; Rahwan 2017). No modo HITL, um humano faz a intermediação de todas as decisões feitas pelo sistema, o que nem sempre é desejável ou possível. No HOTL, apesar do humano poder intervir durante o projeto, após entrar na fase funcional, o humano passa a monitorar seu funcionamento e suas decisões (Awad et al. 2020). Em uma atuação mais holística, no HIC a supervisão humana transcende o funcionamento do sistema de IA e se estende a impactos econômicos, sociais, legais e éticos (Wirtz et al 2022). Nesse último caso, pode-se acrescentar *feedback* direto da sociedade – *society-in-the-loop*, modelo defendido para os canais entre governo e cidadãos, no qual a sociedade avalia os serviços públicos baseados em IA, cujo reporte permite o conhecimento de seus valores e da distância entre eles e os valores transmitidos pelos serviços públicos com IA embutida (Rahwan 2017; Awad et al. 2022; de Almeida et al. 2021).

2.2.5.7 Promoção dos valores humanos

O grande alcance da IA em praticamente todos os setores da atividade humana destaca o quanto crucial será para construção de novos modelos de análise, de julgamento, nova jurisprudência e de pensamento geral sobre a vida (AI HLEG 2019b). Torna-se, portanto, imprescindível que a IA seja utilizada de maneira a promover valores humanos, como a dignidade e a autonomia humana com a promoção do benefício da sociedade (Smuha 2021; Kazim & Koshiyma 2021; Mantelero 2018; Donahoe & Metzger 2019). De maneira pragmática, tais princípios priorizam o espírito público considerando fatores sociais, políticos e econômicos. Respeito à dignidade humana, à democracia, às leis, à cidadania, ao voto, a uma correta administração de recursos públicos, como exemplos (Fjeld et al. 2020; European Commission 2021b).

2.2.6 Dilemas éticos em sistemas de IA

Como os resultados dos sistemas de IA embutem valores e estes variam com os grupos sociais (Kelly 2011), o provimento de serviços digitais baseados em IA requer profundo conhecimento do contexto em que se inserem (Dignum 2022). Variáveis associadas à cultura, à sociedade, à geografia, à temporalidade, influenciam tais conceitos e sugerem uma abordagem relacional aos serviços e produtos com IA em um mundo real (Eke et al. 2023b; Dignum 2022; Hildebrandt 2018).

Quando se desce dos níveis das definições de princípios para os processos, práticas e o desenvolvimento e sustentação dos sistemas de IA, podem surgir situações de conflito ou dúvidas sobre a forma de implementar um princípio sem prejudicar outros (Locher & Bolander 2019; Awad et. Al 2022; Ma et al. 2018).

Um dilema ético é uma situação, entre todas as alternativas possíveis, em que não existe uma única resposta considerada correta para a tomada de decisão, sem que se viole um princípio moral (Bonnemains et al. 2018; Poel 2016). Muitas vezes, é necessário definir quais variáveis são mais relevantes em determinado contexto, antes de se iniciar o processo decisório (Locher & Bolander 2019). Como os princípios éticos devem ser traduzidos para um entendimento lógico geral (Ashok et al. 2022), a inserção de regras para orientar decisões em sistemas autônomos impõe a previsão de dilemas éticos nos processos decisórios embutidos (Bench-Capon & Modgil 2017).

Em Bonnemains et al. (2018), com alcance na Ética Utilitária e na Ética Deontológica, o *framework* proposto se fundamenta em operações lógicas aplicadas aos dilemas utilizando as classes "decisão", "evento" e "efeito" para três possíveis resultados: aceitável ($\top$), inaceitável ($\bot$), indeterminado (?). Em uma abordagem já transformada em sistema, Anderson & Anderson (2018) criaram o GenEth, baseado na lógica indutiva do Teste Alan Turing (Copeland 2000; Oppy & Dowe 2011) para analisar dilemas éticos. O modelo requer que se identifique primeiramente as ações que se constituem um dilema (fazer ou não fazer algo, por exemplo). Em seguida, identificam-se os princípios éticos envolvidos, e as características eticamente relevantes à análise, ou seja, quais as circunstâncias afetam a ação. Outras estratégias de tratamento de dilemas, sem explicitar linguagens formais, foram propostas por Awad et al. (2022) por meio de *crowdsourcing*, assim como por Wright et al. (2014) por meio da análise de cenários.

Nas etapas em que os dilemas são discutidos e reavaliados, torna-se imperativa a participação de perfis capacitados a lidar com os clássicos dilemas éticos, assim como pessoas sintonizadas com

padrões de comportamento na sociedade ao longo do tempo (Vanhée & Borit 2022; Awad et al. 2022). E, em situações nas quais não foram encontradas alternativas de tratar os conflitos éticos, o projeto deve ser interrompido com claro registro das motivações. Por fim, em qualquer situação, os tomadores de decisão são responsáveis pelas escolhas feitas para cada dilema ético, ou pela descontinuidade do projeto (AI HLEG 2019b).

2.3. Regulação da IA

Apesar de ser preliminarmente associado à legislação (Hildebrandt 2018), o ato de regular, em sentido geral, objetiva tentar modificar comportamentos de acordo com padrões definidos ou propostos, com o intuito de produzir resultados desejados (Black 2002). Quando questões éticas são envolvidas, exigindo alta complexidade em aplicação no mundo real, somente a lei não é suficiente, sendo necessárias alternativas de regular comportamentos diversos na sociedade (Hagendorff 2019). Com adicional complexidade, regular tecnologias "disrruptivas" exige modelos distintos dos tradicionais, haja vista a imprevisibilidade de alternativas de implantação e de impactos (Kaal & Vermeulen 2017).

Uma das primeiras barreiras são os diferentes entendimentos sobre o termo "inteligência artificial" em suas diversas formas de abordagem e de aplicação na sociedade (Firth-Butterfield 2017; Buiten 2019), dificultando legisladores a acompanharem o ritmo em que a tecnologia evolui e se modifica; o que reforça a necessidade de um modelo de construção legislativa inovador (Villaronga & Heldeweg 2018; Cerka et al. 2015; Larsson 2020; Taeihagh 2021).

2.3.1 Legislação sobre IA – *Hard law*

No contexto global, observa-se grande distância entre os países na maturidade de discussões sobre IA na esfera legislativa. Contudo, antecipando um cenário próximo, Ruttkamp-Bloem (2023) alerta para o risco de empresas multinacionais de tecnologias migrarem para países com fraca ou inexistente legislação. No escopo das leis existentes (*hard law*), raras leis já foram aprovadas para regular a inteligência artificial considerando as questões éticas, enquanto muitas discussões ainda ocorrem nas casas legislativas nacionais, supranacionais, e subnacionais, como pode ser acompanhado pelo observatório da OCDE (OECD 2022c).

Entre as poucas leis, a maior abordagem ocorre para regulação de aplicações específicas, como veículos autônomos, reconhecimento facial (Tangerding 2021; Husch & Teiden 2017; Imai 2019;

Chen 2021), por exemplo, com focos menos abrangentes sobre princípios éticos (tabela 2). Consistem em abordagens dirigidas aos principais riscos daquela tecnologia, e nem sempre se estendem aos aspectos éticos. Casos de legislação para veículos autônomos geralmente se aplicam a ajustar a legislação de tráfego do país para aceitar os testes com tais veículos em níveis de autonomia mais baixos. Depois, evoluem para permitir produção de tais veículos. São casos com maior foco nos temas segurança, transparência e autonomia humana. Para os casos de legislação para uso de biometria, amplia-se a mais princípios como o caso do direito à privacidade.

Tabela 2: Exemplos de leis aprovadas que fazem algum tipo de regulação sobre uso da IA.

Tema	Abordagem	País/Estado	Referência
Ética em dados	Alteração de lei para exigência de política sobre ética em dados nas organizações	Dinamarca	Danish Government (2020)
Veículos autônomos	Alteração de lei para definir condições de permissão de veículos autônomos nas estradas e sobre a guarda dos dados por eles gerados	Alemanha	German Federal Government (2021)
	Autorização para teste e/ou autorização para uso de veículos autônomous em algumas modalidades.	Alabama, Arizona, Arkansas, California, Colorado, Connecticut, District of Columbia, Florida, Georgia, Hawaii, Illinois , Iowa, Kansas, Louisiana, Maine, Massachusetts , Michigan, Nebraska, Nevada, New Hampshire, New Mexico, New York , North Carolina, North Dakota, Ohio, Tennessee, Texas, Utah, Vermont , Virginia , Washington, West Virginia	Insurance Institute for Highway Safety (2022).
	Novas regras para algumas categorias de veículos autônomos	União Europeia	European Commission (2022b)
	Revisão da lei de tráfego para regulação de algumas modalidades de veículos automatizados.	Japão	National Policy Agency of Japan (2020)
	Regulação de algumas modalidades de veículos autônomos	Coreia do Sul	Kim et al. (2022) BLK (2020)
Informações biométricas	Proíbe instituições privadas de coleta e uso de informações biométricas	Illinois	Illinois General Assembly (2008)
	Regula o uso de reconhecimento facial	California Washingtom, DC	California General Assembly (2020), OECD (2022b)
	Proíbe o uso de reconhecimento facial em organizações públicas	Northampton, MA ; Brookline, MA; Cambridge, MA ; Springfield, MA ;	OECD (2022b)
	Regulação das tecnologias para vigilância pública	New York	OECD (2022b)

Fonte: Elaboração própria

A complexidade de aprovação de uma lei transversal que contemple um amplo conjunto de princípios éticos e que seja aplicada com a necessária amplitude de modelos de desenvolvimento e tipos de comercialização e utilização, tem feito discussões se estenderem nos últimos anos (OECD 2022c).

As discussões no âmbito legislativo em torno de uma lei abrangente geralmente se iniciam com a definição de inteligência artificial (AI-HLEG 2019a), que por si, já gera percepções diferenciadas até mesmo entre os especialistas (Firth-Butterfield 2017; Larsson 2020). A variedade de dimensões abordadas pela proposta de se regular a IA tem movimentado as comissões nos parlamentos com a participação de especialistas no tema (AI-HLEG 2019a; AI-HLEG 2019b; House of Lords 2018; German Federal Government 2019).

Um dos projetos de lei mais debatidos tem sido o Ato Europeu para IA (European Commission 2021a), fundamentado no trabalho do Grupo de Especialistas para IA formado para esse fim (AI-HLEG 2019b). O Modelo regulatório defendido pelo Parlamento Europeu é constituído de um instrumento horizontal em toda a União Europeia, com abordagem baseada em risco proporcional e códigos de conduta para sistemas de IA. A obrigatoriedade das ações se aplica para situações classificadas como risco elevado; permitindo, portanto, aos demais casos, apenas o seguimento de um código de conduta de maneira voluntária, o que tem sido questionado por Djeffal (2022).

Além da classificação em risco alto e risco baixo ou mínimo, a proposta europeia também criou uma categoria de riscos inaceitáveis para representar as práticas proibidas, definidas como aquelas com potencial para manipulação de pessoas por meio de técnicas subliminares, para exploração de grupos vulneráveis, para causar danos psicológicos ou físicos, e de atribuição de rótulos a pessoas. Destaca-se como prática proibida a utilização de sistemas de identificação biométrica à distância em tempo real para espaços públicos, permitindo poucas exceções. Ainda para sistemas de risco elevado, entre as obrigações associadas à transparência, destaca-se obrigatoriedade de se informar as pessoas, quando elas estão interagindo diretamente com um sistema de IA, ou quando suas emoções ou características estão sendo reconhecidas por meios automatizados (European Commission 2021a).

O grande número de discussões e engajamento promovido pelo Parlamento Europeu na proposta de Ato para IA permitiu avaliações profundas para sua evolução (European Parliament 2022a; Georgieva et al. 2022; Muller et al. 2022a, 2022b, 2022c; Bogucki et al. 2022) em diversas dimensões que ainda estão em andamento. Destacam-se as recomendações por harmonização com vários artigos da lei europeia para proteção de dados pessoais (EU GDPR 2016), demonstrando a estreita relação existente entre a regulação de sistemas de IA e a de proteção de dados pessoais (Bogucki et al. 2022), fato antecipado por Firth-Butterfield (2017) e

Butterworth (2018). Similarmente, foram identificadas relações com outras leis como a legislação sobre segurança cibernética, serviços digitais, marketing digital, governança de dados (Bogucki et al. 2022; Muller et al. 2022a).

2.3.2 *Soft law* para a IA

O rito necessário ao trâmite legislativo impõe uma velocidade na legislação tradicional muito menor ao necessário para acompanhar a evolução da IA, seja em sua concepção, seja em sua aplicação na sociedade (Marchant 2019; Villaronga & Heldeweg 2018; Tokmakov 2019)**.** Este descompasso é o primeiro motivador para adoção de alternativas a *hard law* para regulação da IA. A segunda causa reside no fato de que a IA envolve aplicações transversais a múltiplos setores e grupos culturais de *stakeholders*, resultando em um sistema de equações de complexa solução. Portanto, seria insustentável tentar resolver os desafios regulatórios da IA apenas com *hard law* (Marchant 2019; Gutierrez & Marchant 2021).

A percepção de que a legislação não consegue cobrir todos os aspectos da responsabilidade sobre algoritmos de IA (Mäntymäki et al. 2022a) reforça o entendimento de que a adoção de *soft laws*, encaixa-se perfeitamente às necessidades de normatização de tecnologias inovadoras, e, em especial, à inteligência artificial (Marchant 2019; Wallach & Marchant 2019); minimizando os riscos de se partir para adoção de uma legislação sem a maturidade necessária (Scherer 2016; Hopster & Maas 2022).

Soft laws são normas flexíveis, não obrigatórias, que requerem uma governança para minimizar o risco de perder legitimidade. Por isso, sua implantação geralmente é embutida em programas de fomento, nos quais são estruturados processos com forte preocupação em engajamento. Em alguns casos, *soft law* é a própria estratégia de governança. Geralmente, são utilizadas para princípios, códigos de conduta, guias técnicos, recomendações, guias para certificação, estratégias e padrões (Senden 2005; Nolan 2013; Gutierrrez & Marchant 2021; Marchant 2019; Cogliaanese 2020; IEEE 2019b). A busca pelo estabelecimento de princípios éticos para a IA tem resultado em muitas iniciativas de *soft law* pelas organizações governamentais de todo o mundo, embutidas em estratégias de transformação digital, estratégias de IA, ou apenas focadas na regulação da IA (Fjeld et al. 2020; Jobin et al. 2020; IEEE 2020a).

Adicionalmente, merecem destaque algumas iniciativas de grande repercussão promovidas por associações de pesquisadores, pela academia e pelo setor privado como a *Partnership on AI to*

Benefit People and Society (2016), AI4People (2018*), Future of Life Institute* (2019), o *Institute of Electrical and Electronics Engineers* (2022b), Microsoft (2020). Também possível a associação em instituições públicas, em organizações que reúnem governos como a União Europeia, e os intergovernamentais, como a união do Conselho Europeu, da Agência da União Europeia para Direitos Fundamentais, do Banco de Desenvolvimento Interamericano, da Organização para Cooperação e Desenvolvimento Econômico, das Nações Unidas, da Unesco e do Banco Mundial no GlobalPolicy.AI (2021).

Uma vez que padrões são acordos baseados em especificações mínimas necessárias para o ciclo de vida de produtos e serviços (Dignum 2022), várias organizações especializadas no estabelecimento de padrões ligados à tecnologia têm investido neste tipo de *soft law* sob diversas abordagens, desde modelos organizacionais, processos produtivos e gerenciais, até modelos algorítmicos (ISO 2021a; ISO 2021b; ISO 2021c; ISO 2022a; ISO 2022b; IEEE 2019a; IEEE 2020b; IEEE 2021a; IEEE 2021b; IEEE 2021c; IEEE 2021d; IEEE 2022a; Taddeo & Floridi 2018; Adadi & Berrada 2018; Neznamov 2020).

E, por fim, *soft laws* utilizadas para banir mau uso da IA se constituem em uma forma mais direta de alertar para usos extremos (ex: armas autônomas letais), focando no que é proibido (Future of Life 2017; Stop Killer Robot Coallision 2020; University of Ottawa 2017).

O uso de *soft law* pode ser feito como uma estratégia de comunicação e preparação para instrumentos legais mais fortes, muito comum na forma de "*white papers*" e programas de ação. Contudo, também pode ser utilizado como um instrumento interpretativo para uma comunidade jurídica sobre um tema ou lei existente. Na primeira abordagem, *soft law* apresenta natureza transitória, na segunda, permanente (Senden 2005). Quando adotada a combinação de *soft law* com *hard law*, obtém-se maior efetividade (Amigoni & Schiaffonati 2018; Scherer 2016; Rahwan 2017; Boden et al. 2017; Wallach & Marchant 2018). Em ambos os casos, a aplicação de *soft law* requer clareza dos papéis e responsabilidades, portanto, de um modelo de governança (Nolan 2013).

Em estudo sistematizado, Gutierrez e Marchant (2021) identificaram 634 organizações e associações em todo o mundo que prestam serviços de criação, fomento ou certificação de *soft law*s dirigidas a IA. Fazem parte dessa amostra, empresas do setor privado, organizações públicas, academia, e associações sem fins lucrativos que reúnem pesquisadores, patrocinadores influentes, atuando em diferentes formatos. Como resultados, foram apresentados padrões de processos e

práticas, guias de princípios éticos e programas de certificação para produção de sistema de IA, sendo a maior parcela provida diretamente ou em parceria com instituições de governo.

2.3.3 Modelo híbrido –*Hard law* e *Soft law*

Dado o desafio exposto, o conceito de regulação dinâmica se aproxima da solução para a regulação da IA na medida em que se fundamenta no estabelecimento de um processo contínuo de aprendizagem relacionando-se às descobertas tecnológicas com a regulação, em uma alimentação mútua (Lewis & Yildirim 2002; Kaal & Vermeulen 2017). Com fito em superar tais desafios, pesquisadores estudam como *hard laws* e *soft laws* poderiam ser efetivas na regulação da IA (AI HLEG 2019b; Marchant 2019; Gutierrez & Marchant 2021).

Em uma revisão sistemática da literatura sobre a regulação e governança da IA de 2009 a 2020, de Almeida et al. (2021) buscaram unir as principais propostas de 21 *frameworks* apresentados para regular a IA. Construíram em uma visão holística um f*ramework* que representasse os macroprocessos necessários entre os principais *stakeholders* envolvidos na regulação da IA de uma nação: governo - legislativo, executivo e judiciário (Maluf 1995), sociedade, academia e provedores de sistemas de IA. O estudo resultou no framework AIR – *Artificial Intelligence Regulation*, como proposta de modelo de governança da IA de um país no qual ocorrem simultaneamente e necessariamente, processos para elaboração de *soft laws* e de *hard laws*, coadunando com as percepções de Stuurman e Lachaud (2022), e de Gutierrez e Marchant (2021).

2.4. Papel do Governo na governança de IA de um país

O ecossistema de governança e regulação representado no AIR (de Almeida et al. 2021) apresenta, de maneira sintetizada, esforços em dimensões diferentes, mas, integrados, na tentativa de demonstrar a multidisciplinaridade (Bonnemais et al. 2018) necessária ao contexto social e político que a IA se encontra (Leitner & Stiefmueller 2019). Não obstante a previsão da sociedade, da academia e da indústria envolvida em produtos e serviços com IA embutida, o papel do Governo se apresenta como protagonista da governança de IA de um país, respeitando a missão de cada esfera de governo – Executivo, Legislativo e Judiciário.

Em uma posição de protagonismo no poder Executivo, propõe-se a criação de uma Agência Reguladora como uma instância desse poder para assumir atribuições de pesquisa, construção de padrões, certificação e auditoria. Uma vez criada, a Agência Reguladora inicia processo para

criação de estudos e testes para dar subsídio à elaboração das leis, enquanto também avalia propostas de leis em discussão no Legislativo; e normativos são elaborados de maneira incremental, acompanhando a evolução da tecnologia e das leis criadas e/ou ajustadas. Como atribuições para a Agência Reguladora foram previstas: análise de impacto da legislação nos sistemas de IA, elaboração de padrões técnicos, certificações de organizações e de seus sistemas de IA, além de auditoria. Tais ações se aplicam ao escopo de: representação formal de regras e dilemas éticos, análise de impacto dos sistemas de IA nos *stakeholders*, avaliação da governança de dados, avaliação do processo de desenvolvimento de sistemas de IA, identificação de vieses, avaliação das ações utilizadas para mitigação de riscos. Em paralelo, ocorrem interações com órgãos de padronização internacionais em torno das boas práticas para uso e desenvolvimento de sistemas de IA. Admite-se a possibilidade de segmentação das atribuições da entidade "Agência Reguladora" a várias agências a depender da formatação que os Governos possuírem.

Responsável pela legislação, o Legislativo, em seu papel representativo, por meio de canais voltados aos cidadãos, mantém-se aberto a receber contribuições às propostas de leis sobre a regulação da IA. Como parte de um processo contínuo entre Legislativo e Executivo, Agência repassa ao Legislativo documentos e análises sobre *soft laws* que fundamentem a legislação em construção, o que pode ocorrer pelas comissões legislativas nos parlamentos e assembleias. E o Legislativo, por sua vez, repassa à Agência as versões de propostas de leis em discussão. Em sucessivas iterações, os fluxos citados permitem uma construção legislativa incremental e compatível à filosofia Ágil (Nerur, S. & Balijepally 2007; Mohammad 2017), até que exista maturidade para aprovação ou evolução das leis.

Observa-se movimento iterativo semelhante na construção do Ato Europeu para IA (European Commission 2021a), quando o grupo de especialistas contratados pela Comissão Europeia fez e discutiu os estudos com o Parlamento Europeu (AI HLEG 2019b; European Parliament 2022a). Algumas atribuições previstas no AIR para a Agência Reguladora correspondem ao funcionamento de uma *sandbox* - ambiente controlado por organizações regulatórias no qual são realizadas pesquisas e testes com tecnologias para comprovação de funcionalidades e análises de riscos (European Parliament 2022c). Confirmando as previsões, uma primeira *sandbox* regulatória para IA foi apresentada pela Comissão Europeia para estabelecer padrões no desenvolvimento de sistemas de IA (European Commission (2022c).

Em meio às discussões sobre emendas legislativas ao referido ato, e na proximidade dele ser transformado em lei, Stahl et al. (2022b) avaliaram as atribuições previstas para o Conselho de IA Europeu, e apresentaram uma proposta de criação e uma Agência de IA com estrutura mais robusta à altura das necessidades regulação e de governança de IA ao bloco Europeu, e à previsão de *sandbox* pelo Ato de IA Europeu (European Parliament 2022c; European Commission 2021a). Na mesma direção, o governo da Noruega estabeleceu uma *sandbox* regulatória para IA (Norwegian Ministry of Local Government and Modernisation 2020), e o Reino Unido, estabeleceu *sandbox* para novas tecnologias que envolvam sistemas para área de saúde, realidade aumentada e reconhecimento facial (Information Commissioner's Office 2022).

Por sua vez, o Judiciário, na proposta do AIR, recebe da Agência Reguladora de IA documentos sobre as organizações, produtos e serviços certificados, e encaminha os resultados de julgamentos efetuados em incidentes ou outros litígios envolvendo a IA, requerendo, para isso, transparência na produção de sistemas de IA que garantam compreensão para os julgamentos.

As propostas e ações dirigidas a uma convivência entre *hard law* e *soft law*, materializadas em legislação, recomendações governamentais e padrões internacionais para governança de IA (Dignum 2022) geram expectativas na sociedade, na academia e no próprio governo de que organizações públicas adiram ao movimento dentro de sua esfera de poder (figura 1), e implantem práticas e processos para que o uso e o desenvolvimento de sistemas de IA sob sua responsabilidade considerem as questões éticas (Mäntymäki et al. 2022a).

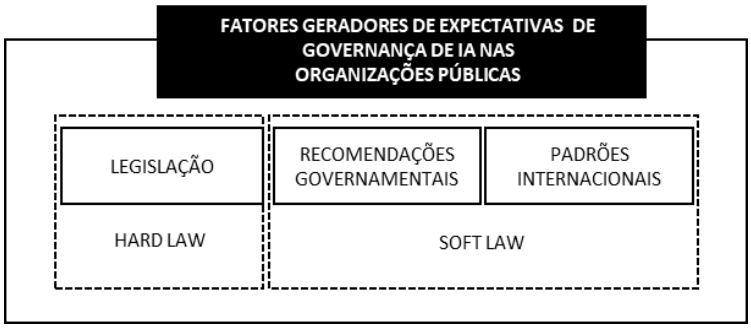

Figura 1: Fatores geradores de expectativas de iniciativas para governança de IA nas organizações públicas.
Fonte: Elaboração própria.

2.5. Limitações, Contribuições e Agenda

A presente pesquisa não foi exaustiva em análise de projetos de lei em discussão sobre IA nas casas legislativas, uma vez que o conteúdo e forma dessas proposições são suscetíveis a muitas e frequentes mudanças. O caso do Ato Europeu (European Commission 2021a) foi abordado em razão do longo e transparente processo em torno de sua construção, com amplo envolvimento da academia e governos, o que coloca o trabalho do Parlamento Europeu como referência para outros parlamentos. Quanto a leis sancionadas para IA, apesar dos esforços com parlamentos e no observatório de IA da OCDE (OECD 2022c) o levantamento realizado não pode ser considerado completo, visto que a cada dia uma proposta de lei em discussão pode ter sido aprovada e sancionada.

O estudo contribui com maior clareza sobre as razões para a regulação da IA, um aprofundamento nos conceitos envolvidos quando se refere à existência de questões éticas no uso e desenvolvimento de IA - valor, moral, ética, teorias éticas, dilemas éticos, vieses, e temas abordados nos guias de princípios éticos. Apresentou as formas de regular a IA, e expôs porque a necessidade de regulação da IA evolui para a governança de IA em um país, com dependência das ações dos órgãos de governo para tal fim. A pesquisa também apresentou casos reais que implantaram recentemente atribuições previstas no *framework* AIR (de Almeida et al. 2021), confirmando a combinação *hard law* e *soft law* para regular a IA, e a importância das instituições públicas na regulação da IA.

Abstraindo-se da missão de cada esfera de governo nos processos envolvidos para regulação de IA de um país, cada organização pública que produz e utiliza sistemas de IA, é uma entidade que precisa se inserir nas determinações legislativas e nas *soft laws* de seu país. Posto isso, propõe-se uma agenda de investigação de como as organizações públicas, independentemente de sua esfera de governo, estão interpretando a necessidade de considerar as questões éticas ao produzirem sistemas de IA.

3. GOVERNANÇA DE INTELIGÊNCIA ARTIFICIAL NAS ORGANIZAÇÕES PÚBLICAS: FUZZY E CRISP-SET QCA APLICADAS A PROCESSOS E PRÁTICAS QUE CONSIDEREM PRINCÍPIOS ÉTICOS

RESUMO

A corrida pela regulação e governança global da inteligência artificial (IA) tem desafiado pesquisadores, gestores, juristas, legisladores e a própria sociedade. Enquanto observam esse movimento, organizações públicas se deparam com a necessidade de se estruturarem para que seus sistemas de IA considerem princípios éticos. A presente pesquisa teve como objetivo investigar como as organizações públicas têm incorporado as diretrizes apresentadas pela academia, pela legislação e pelos padrões internacionais, ao seu modelo de governança, de gestão e de desenvolvimento de sistemas de IA de maneira a considerar princípios éticos. Foram elaboradas proposições sobre os processos e práticas recomendados pela literatura especializada na implantação da governança de IA. Questionário e roteiro de entrevista foram construídos e submetidos à avaliação por uma banca de juízes, seguida de pré-testes. Procedeu-se uma busca em organizações públicas que tivessem sistemas de IA em operação, resultando em amostra composta de vinte e oito organizações públicas, distribuídas em dezessete países. Com objetivo exploratório e descritivo, em uma abordagem qualitativa e quantitativa, foi utilizado o método *Qualitative Comparative Analysis* (QCA), nos modos *crisp-set* e *fuzzy*, a partir das respostas de questionários, cujos resultados foram complementados com análise de conteúdo das entrevistas e documentos disponibilizados. Os resultados permitiram identificar como processos e práticas dirigidos à aplicação de princípios éticos na produção de sistemas de IA têm sido combinados e internalizados aos modelos de governança, de gestão e de desenvolvimento de sistemas de IA nessas organizações públicas; como utilizaram habilitadores no auxílio da implantação da governança de IA; e como enfrentaram as dificuldades nessa jornada. Os resultados também fundamentaram a elaboração do *framework* AIGov4Gov que apresenta como os processos e práticas se distribuem em uma organização pública para implantar sua própria governança de IA, unindo unidades de negócio e unidade de TI. E no propósito da governança de IA, como organizações públicas se relacionam com agências de governo para padronização, assim como interagem com organizações parceiras e terceirizadas para a produção de sistemas de IA em atendimento a princípios éticos. Como contribuição aos gestores e pesquisadores, além do conhecimento de como as organizações se estruturaram para atender às questões éticas envolvidas na produção de sistemas de IA, a pesquisa apresentou um modelo integrado de processos e práticas de governança, de gestão, e de desenvolvimento de sistemas de IA, que se estendem do nível estratégico até o nível operacional de uma organização pública para implantação da sua governança de IA, de maneira que ela possa se inserir na governança de IA de seu país, e, assim, somar ao movimento global de governança de IA.

Palavras-chave: Governança de IA, Regulação de IA, IA responsável, Ética na IA, IA nas organizações públicas, Fuzzy QCA.

ABSTRACT

The race for artificial intelligence (AI) regulation and global governance has challenged researchers, managers, jurists, legislators and society itself. While observing such movement, public organizations are faced with the need for structuring themselves so that their AI systems consider ethical principles. This research investigated how public organizations have incorporated the recommendations presented by academia, legislation and international standards, to their governance, management and AI systems development models, in order to consider ethical principles. Propositions were elaborated on the processes and practices recommended by the AI governance literature. A questionnaire and interview script were elaborated and submitted to evaluation by a judges' board, followed by pre-tests. A search was carried out in public organizations that had AI systems in operation, resulting in a sample composed of twenty-eight public organizations, distributed in seventeen countries. With an exploratory and descriptive objective, through a qualitative and quantitative approach, the Qualitative Comparative Analysis (QCA) method was used, in crisp-set and fuzzy modes, based on questionnaire responses, whose results were complemented with content analysis of interviews and shared documents. As results, one could know how those organizations have combined processes and practices for applying ethical principals in such way to incorporate them to their governance, management and AI systems development models; how they used enablers to support their AI governance implementation; and how they faced difficulties on this journey. The results also supported the elaboration of the AIGov4Gov framework, which presents how processes and practices are distributed in a public organization to implement its own AI governance, joining business units and IT units. In addition, how public organizations relate to government agencies for standardization, as well as interact with partner and third-party organizations for AI systems development in compliance with ethical principles. As a contribution to managers and researchers, in addition to knowledge on how organizations are structured to meet the ethical issues involved in AI systems development, the research presented an integrated model of governance and management processes and practices, which extend from the strategic level down to the operational level of a public organization to implement its AI governance, so that it can be part of its country's AI governance, and thus join the global AI governance movement.

Key words: AI governance, AI regulation, Responsible AI, Ethical AI, AI in public organizations, Fuzzy QCA

3.1. Introdução

O movimento pela regulação e governança da inteligência artificial (IA) tem envolvido uma variedade de *stakeholders*, com destaque aos governos, nas esferas executiva, legislativa e judiciária, a academia, e órgãos internacionais de padronização (Gutierrez & Marchant 2021; OECD 2022c).

Não obstante a literatura especializada em regulação e governança de IA ter se apresentado com expressivo número de ensaios teóricos na tentativa de explicar classificações de modelos conceituais em diferentes abordagens (de Almeida et al. 2021; Stix 2021), a governança de IA ainda é uma área de pesquisa subdesenvolvida (Taeihagh 2021), requerendo uma maior compreensão de como as organizações têm interpretado e incorporados princípios éticos em suas práticas, processos e estrutura para produção de sistemas de IA (Mäntymäki et al. 2022a; Mikalef et al. 2022).

Em pesquisa sistemática sobre a inteligência artificial na governança pública, Zuiderwijk et al. (2021) identificaram lacunas tanto nos processos, quanto no conteúdo das pesquisas. Para atender à primeira lacuna, os autores recomendaram adoção de múltiplos métodos orientados a dados, para se obter com pesquisas empíricas, abordagens exploratórias que também permitam análises qualitativa-quantitativas, de maneira a se aprofundar nos aspectos próprios da governança no setor público em relação à produção de sistemas de IA, com destaque para as questões éticas que a IA pode gerar. Entre as lacunas quanto ao conteúdo, os autores destacaram a necessidade de se investigar: planos e estratégias dos órgãos governamentais para implantação de sistemas de IA; práticas de gestão de riscos da IA no setor público, com ênfase nas questões éticas; e os possíveis modelos de governança de IA no setor público.

O cenário atual do ecossistema de regulação da IA, formado pelo conjunto de *stakeholders,* processos e produtos dirigidos a sistemas de IA responsáveis (Minkkinen et al. 2022) é caracterizado por:

a) Raros casos de legislação aprovada, com sinais de muitas evoluções nos próximos anos (OECD 2022c);
b) Muitas propostas de leis tramitando em Casas Legislativas nacionais, supranacionais e subnacionais de vários países, tanto com abordagem geral para a IA, quanto para serviço específico sustentado pela IA (OECD 2022c; European Commission 2021a);

c) Variados instrumentos normativos, em diversas abordagens, contemplando princípios éticos, recomendações, políticas e estratégias elaborados por organizações em todo o mundo com boas práticas de uso e desenvolvimento de sistemas de IA (Fjeld 2020; Gutierrez & Marchant 2021).

Considerando a importância que as instâncias governamentais exercem nos esforços de regulação e de governança de IA (de Almeida et al. 2021, Stix 2021), os potenciais benefícios e riscos que a IA pode trazer à sociedade quando suporta organizações do setor público (Zuiderwijk et al. 2021), e o desafio de evitar a perda de confiança no governo e nas decisões governamentais (Zuiderwijk et al. 2021) suportadas por sistemas de IA, torna-se crucial entender: como organizações públicas estão adaptando ou implantando seu modelo de governança, de gestão e de desenvolvimento, para produzir sistemas de inteligência artificial (IA) que considerem princípios éticos?

3.2 Da Governança corporativa pública até a governança de IA

Unindo conceitos de diversas abordagens Matei & Drumasu (2015 p. 497) definiram governança corporativa como "*forma como uma organização (pública ou privada) é liderada e controlada, com o propósito de obter desempenho/cumprimento de suas responsabilidades com sucesso e agregação de valor, bem como utilizar eficientemente recursos financeiros, humanos materiais e informacionais, respeitando aos direitos e obrigações de todas as partes envolvidas (acionistas/investidores, Conselho de Administração, administradores, funcionários, Estado, fornecedores, clientes e demais pessoas com interesse direto)*".

3.2.1 Governança pública

Considerada como precursora da governança corporativa (Heracleois 2012), a Teoria da Agência descreve as relações entre o Principal que delega ações executivas ao Agente, nas quais o proprietário é afetado pelas escolhas do Agente (Jensen & Meckling 1976). Trata-se de uma importante ferramenta analítica para situações que envolvam delegações, independentemente do contexto institucional, como aquelas existentes em uma governança (Wiseman et al. 2012).

Para o propósito da presente pesquisa, cujo *locus* são organizações públicas, o conceito de governança requer uma conotação mais próxima dos serviços públicos na qual destaca-se o relacionamento entre os formuladores de políticas e / ou administradores de organizações públicas

e os gerentes seniores, dada a tarefa de tornar essas políticas em realidade (Cornforth 2003); mas, também considerando a relação da sociedade na formulação de políticas públicas (Leftwich 1993; Rhodes 1997). Nesse sentido, o conceito coaduna com o de "Boa Governança", cunhado pelo Banco Mundial (IFAD 1999), e com o da "Nova Governança Pública" (Osborne 1997). Pilar da "Nova Governança Pública" e da "Boa Governança", a *accountability* é compreendida como uma relação social na qual um ator sente uma obrigação para explicar e justificar sua conduta para algum outro ator (Bovens 2009). A *accountability* pública é aquela dirigida ao objeto público, gastos públicos, no exercício do poder público ou na condução de instituições públicas (Scott 2000).

3.2.2 Governança de TI

De acordo com o *IT Governance Institute* (2003 p. 19), a Governança de TI é "*parte integrante da governança corporativa e consiste na liderança e nas estruturas e processos organizacionais que garantam que a TI da organização sustente e amplie a estratégia e os objetivos da organização".*

Dada a importância estratégica que a Tecnologia da Informação (TI) tem exercido para as organizações, a governança de TI tem ganhado destaque para garantir o alinhamento da TI com os objetivos das organizações (De Haes & Gremberg 2004; Weill & Ross 2004), sendo considerada parte integrante da governança corporativa (Correia & Água 2021).

Integrando liderança, estrutura, comitê e processos organizacionais nos níveis estratégico e tático (Simonsson & Johnson 2006), a governança de TI tem sido implantada com um foco na fusão entre TI e negócio com deliberações do conselho executivo tendo a participação da TI (Aasi et al. 2014; Grembergen 2002; Correia & Água 2021), de maneira a contemplar os riscos e o desempenho associados às estratégias corporativa e de TI (De Haes & Gremberg 2004).

Para realizar a governança de TI, faz-se necessária uma estrutura composta de comitê de TI, processos e mecanismos, utilizando-se de recursos humanos envolvidos em uma cultura dirigida às necessidades e objetivos da organização (Aasi et al. 2014). Atuando tanto em nível estratégico, quanto tático (Simonsson & Johnson 2006), os processos da governança de TI atuam no contexto do alinhamento estratégico, dos riscos, da gestão de recursos, da distribuição de valor e do desempenho da organização (Aasi et al. 2014).

3.2.3 Governança de IA

O propósito de uma IA que considere princípios éticos traz novas obrigações para as organizações produzindo um novo ecossistema a ser governando e gerido de forma sustentável (Hickman 2020). Tais implicações refletem-se nos processos e na estrutura de governança corporativa (ISO 2022a) cujos ajustes precisam manter serviços digitais confiáveis (Zuiderwijk et al. 2021; ISO 2022a). Contudo, os tradicionais processos e estruturas de governança existentes parecem não ser suficientes a atender os desafios da Governança de IA (Taeihagh 2021).

Nas instituições públicas, a responsabilidade e o desafio são maiores, porque, além de precisarem governar sua IA, recai o peso de um cenário no qual os cidadãos não escolhem os produtos de IA, mas, são obrigados a consumi-los na medida em que são embutidos nos serviços públicos (Zuiderwijk et al. 2021). Isso ocorre porque o ciclo de vida de um sistema de IA de uma organização pública não existe de forma isolada, ele interseciona com o ciclo de vida da política pública. É inevitável não haver impacto na política em decorrência de êxitos, falhas ou mesmo danos causados pelo sistema de IA (González et al. 2020). No entanto, também parece ser consenso a crença de que uma governança efetiva da IA aumentaria as chances de segurança, de *accountability*, de mais responsabilidade na pesquisa, desenvolvimento e aplicação dessa tecnologia (Cihon et al. 2020).

Na prática, identificam-se organizações vivenciando um clássico "problema de tradução", caracterizado pela proliferação de guias com princípios éticos abstratos e distantes de sua implementação por práticas e processos associados à governança da IA (Morley et al. 2020b). Como consequência, passam a ser urgentes novas ferramentas para operacionalização dos requisitos éticos nas organizações (Mäntymäki, et al. 2022a). Em meio a tantos caminhos a trilhar, um ponto é destaque nas discussões: a necessidade de uma governança de dados que estabeleça processos e condições de gestão do ciclo completo dos dados como condição básica à governança da IA (Taeihagh 2021). Outro aspecto comum às discussões sobre Governança de IA e Governança de Dados, é a necessidade de se convergir interesses e benefícios de muitos *stakeholders* (de Almeida et al. 2021; Wright & Schultz 2018, Abraham et al. 2018), o que naturalmente remete à discussão para a Teoria dos *Stakeholders*, fundamentada na tríade: união de interesses, postura estratégica cooperativa, e rejeição a uma visão estritamente econômica.

O modelo parece conter elementos cruciais para resolver problemas interdisciplinares (Shah 2022), como o caso da governança da IA em organizações públicas, na qual os aspectos de regulação impactam potencial distinção de valores entre alguns *stakeholders* (Rose et al. 2018; Vial 2019). O propósito da aplicação de princípios éticos para maximizar benefícios e minimizar riscos da IA apresenta-se para viabilizar a colaboração dos *stakeholders* nas organizações e nações no contexto da regulação e da governança da IA. Esse esforço já é evidente desde o início de um projeto de IA, onde a análise de impacto nos *stakeholders* (Wright & Schultz 2018; AI HLEG 2019b; Government of Canada 2020a) é realizada para ser estrutura basilar a todas as ações de mitigação de riscos no ciclo de vida do sistema de IA (Wirtz et al. 2022).

Pesquisando como a Teoria dos *Stakehoders* se aplicaria nas iniciativas de governo eletrônico, Rose et al. (2018) destacam que os normativos das organizações públicas embutem princípios éticos filosoficamente sustentados, que precisam observar todos envolvidos no novo modelo de prestação de serviços. Preocupação semelhante foi resultado da pesquisa de Vial (2019) sobre transformações digitais em organizações, chamando a atenção para o momento de maior diálogo das organizações com seus *stakeholders* e sugerindo estudos sobre Estratégias de TI que considerem teorias éticas. Em estudos mais recentes, dirigidos ao engajamento de *stakeholders* em organizações de TI, Shah (2022) identificou necessidade de maior empenho das áreas de TI reconhecerem mais *stakeholders* que os habituais, na medida em que elas caminham para áreas de inovação, e precisam estar muito próximas das áreas de negócio (Wirtz et al. 2022).

Considerando o foco nos impactos gerados na sociedade como uma premissa de um modelo de Governança de IA (Djeffal 2018), e seguindo as diretrizes da "Nova Governança Pública" e da "Boa Governança", alguns pesquisadores apontam na Governança da IA, a necessidade de integração da Teoria dos *Stakeholders* com a Teoria do Contrato Social (Bonsón et al. 2021; Wright & Schultz 2018), haja vista a necessidade de se obter as percepções da sociedade em relação aos serviços prestados, assim como às percepções dos valores envolvidos nas decisões dos sistemas de IA. Seguindo a inspiração de Rousseau (2016) de que o Estado deve zelar pela segurança e bem-estar dos cidadãos, e permitindo que estes sejam ouvidos, Rahwan (2017) baseou-se numa versão ampliada do *"human-on-the-loop* (HITL)" para "*society-in-the-loop* (SITL)", na tentativa de inserir a sociedade na tarefa de monitoração constante do comportamento do sistema com IA, o que Hickman (2020) resume como *human-in-the-command*, sustentando-se na argumentação de

que somente uma monitoração coletiva seria capaz de trazer diferentes pontos de vista para algo que requer olhar multidisciplinar (Rahwan et al. 2019; Dignum 2019).

Respeitando sua subordinação à governança de TI, na busca por uma efetiva implantação de governança de IA, Mäntymäki et al. (2022a) e Gasser & Almeida (2017) defendem a necessidade de um modelo em múltiplas camadas de maneira a ter na camada superior, mecanismos de captação dos requisitos de normativos regulatórios superiores e da legislação, traduzindo-os para o contexto organizacional por meio de normativos que estabeleçam os princípios éticos e condições para sua aplicação por meio de processos e práticas.

3.3. Modelo de Pesquisa

A pesquisa realizada na literatura especializada apresentada nas seções anteriores permitiu o desenho de um modelo conceitual para o estudo realizado, como ilustrado na figura 2, na qual observam-se no ecossistema de regulação da IA, fatores geradores de expectativas na sociedade de que as organizações públicas considerem princípios éticos quando produzirem sistemas de IA.

Para atender a tais expectativas, no modelo construído, as organizações públicas adequam seu modelo de governança para implantar sua própria governança de IA considerando os princípios éticos. Nessa direção, a organização define estratégia, política e princípios éticos para sistemas de IA, e implanta processos e práticas envolvendo dados, mitigação de riscos, segurança e auditoria. Abrangendo os níveis tático e operacional, são introduzidas práticas voltadas a princípios éticos no seu processo de desenvolvimento de sistemas de IA. Como habilitadores das ações citadas, a organização treina alguns *stakeholders*, podendo também fazer uso de contratações e parcerias para tais práticas; assim como de todo o desenvolvimento dos sistemas de IA. A identificação e compreensão dos fenômenos apresentados no modelo de pesquisa requer a investigação de algumas proposições como segue.

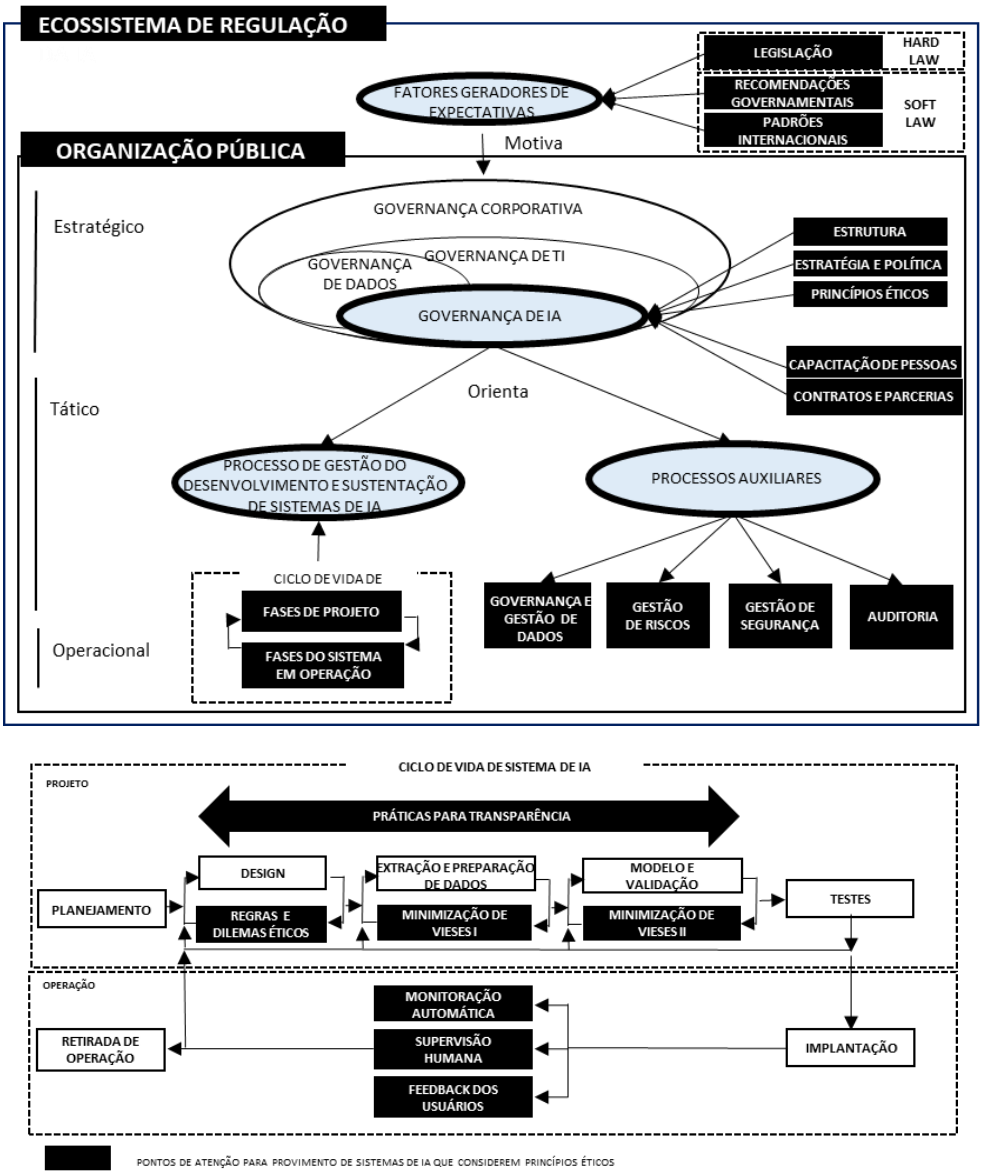

Figura 2: Modelo conceitual do objeto de pesquisa
Fonte: Elaboração propria

3.4. Formulação de proposições

3.4.1. Fatores de regulação na motivação para a governança de IA

Considerando as diversas iniciativas em se regular a inteligência artificial por meio de legislação, por políticas/recomendações governamentais, e/ou padrões internacionais (Marchant 2019; Wallach & Marchant 2019), faz necessário compreender como tais fatores têm influenciado a motivação das organizações públicas a implantarem sua própria governança de IA.

A partir do exposto, conjectura-se a **proposição 1** desta pesquisa: A combinação entre legislação e normativos governamentais (*hard* law e *soft law*) para regulação da IA são fatores geradores de expectativas de que organizações públicas implantem sua própria governança de IA.

3.4.2 Governança de IA nas organizações públicas – nível estratégico

A relação entre os contextos das governanças definidos por Mäntymäki et al. (2022a), na qual a organização corporativa contém a governança de TI, que, por sua vez, contém a governança de IA, permite que se atribua à governança de IA características herdadas da governança de TI. Portanto, analogamente às ações que formalizam uma governança de TI (Aasi et al. 2014; Simonsson & Johnson 2006), considera-se o grupo de ações em nível estratégico voltadas à formalização da governança de IA, composto pela criação de uma estratégia para IA, de uma política para sistemas de IA, de um comitê ou unidade ou pessoa responsável pela governança de IA, a criação de um processo de governança de IA, e de princípios éticos para aplicação em sistemas de IA. No intuito de viabilizar e dar sustentação ao nível estratégico de uma governança de IA, mecanismos são criados para estabelecer uma governança de dados (Janssen et al. 2020; Rhahla et al. 2021; Haneem et al. 2019), processo de gestão de segurança de sistemas de IA (Breier et al 2020; Xue et al 2020), processos e práticas voltados a mitigação de riscos (NIST 2022; Wirtz et al 2022; Breier et al 2020), processo de auditoria de sistemas de IA, e práticas do processo de desenvolvimento de sistemas de IA dirigidas às questões éticas (De Silva & Alahakoon 2022; González et al 2020; Laato et al. 2022).

A partir do exposto, conjectura-se a **proposição 2** desta pesquisa: Processos e práticas na dimensão "dados", "riscos", "segurança", e "desenvolvimento" devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

Considerando o fato de que cada dimensão abordada na proposição 2 pode ser decomposta em outras ações, para maior precisão do fenômeno, proposições derivadas foram criadas para cada dimensão.

3.4.3 Diretrizes para processos e práticas no domínio “dados”

Dada a dependência da IA em relação aos dados, estabelece-se a forte ligação entre a qualidade e integridade dos dados e o resultado dos serviços digitais baseados em IA (Dwivedi et al. 2021;

Eggers et al. 2017; Janssen et al 2020). Isso decorre porque os modelos preditivos que fundamentam os algoritmos de IA – conjunto de regras a serem seguidas no cálculo ou solução de problema (Mergel et al. 2016) - são mais precisos à medida que se aumenta o volume de dados a ser utilizado no treinamento, e se eleva o nível de confiabilidade desses dados (Janssen et al. 2020). Por esta razão, alguns riscos decorrentes do uso e desenvolvimento inadequados da IA também podem ser mitigados por meio de um modelo de governança de dados que contemple políticas, regulamentos e processos (Vining et al. 2022; Haneem et al. 2019) e que direcione a instituição à cultura e às ações de igualdade, de justiça, de transparência (Vetrò 2021), e de princípios eticamente aceitos, assim como a proteção de dados sensíveis e pessoais (Janssen et al. 2020).

A governança de dados pode ser definida como o exercício da autoridade e controle sobre a gestão dos dados, por meio de uma proposta de maximizar o valor dos dados da organização e gestão dos riscos associados aos dados (Abraham et al. 2018). Em seu domínio, distribuídos em vários níveis da organização, são definidos processos como por exemplo, para gestão da qualidade de dados (Haneen et al. 2019; Khatri 2016), e para a gestão da proteção de dados em todo o seu ciclo de vida (Janssen et al 2020; Abraham 2018). Geralmente, uma unidade, uma pessoa ou comitê de governança de dados é criado para especificar direitos e responsabilidades sobre as decisões tomadas acerca dos dados pessoais da organização, e para formalizar políticas, padrões, processos e a monitoração da conformidade em relação aos dados (Vilminko-Heikkinen & Pekkola 2019; Abraham 2018; Carretero et al. 2017).

Em uma visão mais pragmática, pesquisas apontam a implantação de governança de dados como requisito básico ao estabelecimento de governança da IA (Kuziemski & Misuraca 2020). Como consequência, a avaliação de riscos e dos princípios éticos (Vetrò et al. 2021) passa a ser instrumento relevante à governança de dados (Kuziemski & Misuraca 2020; Alshahrani et al. 2021); algumas vezes, em interseção com a governança da IA (Medaglia et al. 2021; Andrews 2018; Özdemir & Hekim 2017; Dwivedi et al. 2021; Mäntymäki et al. 2022a).

O desafio da quebra de silos de dados (Abraham 2018) impõe grande esforço de colaboração de *stakeholders*, internos e externos, normas, e habilidades de negociação para obtenção da colaboração e consenso (Calzada & Almirall 2020; Micheli et al. 2020; Ruijer 2021; Benfeldt et al. 2020). Com a mesma fundamentação, Benfeldt et al. (2020) propõem que governança de dados seja compreendida como uma ação necessariamente coletiva. Seguindo as políticas estabelecidas

na governança de dados (Carretero et al. 2017), processos são implantados para atuar ao longo do ciclo de vida dos dados, abrangendo domínios de qualidade e proteção de dados (Abraham 2018).

A qualidade de dados corresponde à capacidade dos dados de serem úteis a um determinado contexto (Abraham 2018). O processo de gestão da qualidade de dados visa garantir que toda a linhagem de dados (jornada dos dados no seu ciclo de vida) (Rhahla et al. 2021) esteja adequada para cada sistema ou estudo analítico que deles necessite (Carretero et al. 2017). Na impossibilidade de contemplar todos os dados da organização, a gestão da qualidade de dados direciona esforços aos dados mestres (dados de maior importância e criticidade da organização), que sustentam informações relativas aos principais processos de negócio (Haneem et al. 2019; Vilminko-Heikkinen & Pekkola 2019, Benfeldt et al. 2020). Integrada à gestão de riscos, a gestão de qualidade de dados pode prover mecanismos de combate a vieses e seus efeitos (Vetrò et al. 2021); assim como a baixa qualidade de dados ocorre em situações de falhas de omissão, falhas de associação, descarte inadequado, e geração de decisões enviesadas e equivocadas pelos sistemas de IA (ISO 2021a; González et al. 2020; Liu et al. 2022b).

Geralmente, para atender legislação, existe processo específico para gestão da proteção de dados pessoais (Rhahla et al 2021; Vandercruysse et al. 2020; Kuziemski & Misuraca 2020). Sistemas que utilizem dados pessoais geram resultados que podem impactar a vida das pessoas, individualmente ou dos grupos aos quais pertencem; por isso, geralmente, a legislação para proteção de dados pessoais impõe a coleta e tratamento apenas dos dados cujos objetivos se justifiquem e sob consentimento, e o direito ao "esquecimento", por meio do qual, os dados podem ser retirados da base de dados, em situações onde a pessoa à qual o dado se refere, solicitar (Information Commissioner's Office 2020, EU GDPR 2016, Benfeldt et al. 2020; Rhahla et al 2021). Imersas nesse cenário, organizações têm sido encorajadas a reterem os dados pessoais de clientes e cidadãos pelo menor tempo possível. Conflito se estabelece porque, como sistemas de IA precisam de muitos dados, a solicitação de consentimento não é uma tarefa fácil. O problema se amplia quando forem necessárias exclusões de dados, decorrentes de demandas, à luz da lei, causando diferenças nos resultados (Information Commissioner's Office 2020).

Pelos motivos expostos, a produção de sistemas de IA que atendam a princípios éticos requer adequada governança de dados (Kuziemski & Misuraca 2020; Alshahrani 2021; Medaglia, 2021) amparada, pelo menos, por processos de gestão da qualidade de dados (Haneem et al. 2019;

Vilminko-Heikkinen & Pekkola 2019; Vetrò et al. 2021), e de gestão de proteção de dados pessoais (Rhahla et al. 2021; Kuziemski & Misuraca 2020).

A **proposição 2A** fundamenta-se na decomposição da dimensão "dados" nos processos de governança de dados (Kuziemski & Misuraca 2020), de gestão da qualidade de dados (Rhahla et al. 2021) e de gestão da proteção de dados pessoais (Rhahla et al 2021; Vandercruysse et al. 2020), cuja formulação é: Processo de governança de dados, processo de gestão da qualidade de dados e processo de gestão da proteção de dados pessoais devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

3.4.4 Diretrizes para processos e práticas no domínio "riscos"

A **proposição 2B** fundamenta-se na decomposição da dimensão "riscos" em ações identificadas para mitigar riscos próprios dos sistemas de IA (Wirtz et al. 2022; Vetrò et al. 2021; Manterelo 2018).

Sendo os riscos a gênesis das iniciativas de regulação e governança da IA (Vetrò et al. 2021; Medaglia 2021; Zuiderwijk et al. 2021), os princípios éticos são construídos a partir dos riscos que potencialmente os sistemas apresentam (Fjeld 2020), adicionados à diversidade de *stakeholders* e dimensões necessárias à Governança de IA, sugerem muitas adaptações na gestão de riscos tradicional das organizações (ISO 2022a; ISO 2022b; Zuiderwijk et al. 2021). A análise também requer o resgate de que a Teoria da Agência amplia a literatura de riscos com a inclusão do problema de agência no qual as partes cooperantes têm objetivos e visão diferentes sobre o trabalho (Eisenhardt 1989).

Os *stakeholders* impactados direta ou indiretamente pelo sistema de IA desde o projeto até o sistema de IA ser descontinuado, são o ponto central da proposta de gestão de riscos (NIST 2022; Wirtz et al. 2022; BSA 2021). A prática de identificação dos *stakeholders* (indivíduos, instituições, grupos), suas expectativas e condições de atuação frente aos sistemas de IA, constitui-se requisito a um processo de gestão de riscos que acompanhe todo o ciclo de vida dos sistemas de IA (Wright & Schultz 2018). Na mesma direção, as propostas de auditoria nos processos associados a sistemas de IA, também são orientadas a riscos (de Oliveira 2019; Erlina et al. 2020), permitindo-se a materialização de problemas com qualquer *stakeholder* (Zicari et al. 2021).

A amplitude da gestão de riscos para sistemas de IA impõe uma abordagem multidisciplinar às organizações, o que que impõe atenção na criação de cultura sobre tais riscos nos *stakeholders* (Wirtz et al. 2022).

Ao longo do tempo, mudanças de variáveis ambientais podem alterar o contexto para o qual o sistema de IA foi projetado, provocando comportamentos fora dos resultados desejados. Tal situação pode ser evitada com uma monitoração de mudanças no ambiente (regras de negócio, conformidades, composição do corpo funcional, tendências sociais) dentro e fora da organização, alimentando constantemente o processo de gestão de riscos (González et al. 2020).

Pelo exposto, a **proposição 2B** foi formulada como: Processo de gestão de riscos, processo de auditoria, prática para identificação de *stakeholders* em todo o ciclo de vida de sistemas de IA, e monitoração de mudanças no ambiente, devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

3.4.5 Diretrizes para processos e práticas no domínio "segurança"

Integrada à gestão de riscos (Breier et al. 2020) e à gestão de proteção de dados pessoais, a gestão de segurança de sistemas de IA é aplicada para evitar ataques cibernéticos (Eggers & Sample 2020) em todo o ciclo de vida dos sistemas de IA e dos dados (Rhahla et al. 2021; Jackson 2019b).

Além dos ataques gerais a serviços digitais, alguns ataques foram projetados especialmente para explorar vulnerabilidades de algoritmos de IA, preparadas a partir de cinco modelos básicos (Xue et al. 2020; Huawei 2018; European Union Agency for Cyber Security 2021; Mcgraw et al. 2020): *data poisoning* - dados falsos inseridos durante a fase de treinamento podem gerar decréscimo de acuraria ou erros; *backdoor* - códigos maliciosos implantados nos dados durante o treinamento, para ser acionados posteriormente (Gu et al. 2019); *adversarial examples* - perturbações cuidadosamente elaboradas na entrada do teste podem tornar o modelo errado (Chen et al. 2019); *model stealing attack* – permite que se roube os parâmetros do modelo ou que se recupere os dados de treinamento confidenciais; *recovery of sensitive training data* – tentativa de descobrir alguns dados usados durante o treinamento. Algumas situações favorecem tais vulnerabilidades, como por exemplo: *outsourcing* da fase de treinamento dos algoritmos ou modelos pré-treinados com grande volume de dados oriundos de usuários não confiáveis ou de terceiros, sem que tenha sido feita uma efetiva validação dos dados (Xue et al. 2020).

De difícil identificação, a depender do tipo do sistema de IA, os impactos de tais ataques podem ser apenas falhas no reconhecimento facial, até mortes em massa (Eggers & Sample 2020). Entre os mecanismos de controle de segurança devem ser implantados processo de gestão da segurança (European Union Agency for Cyber Security 2021) contemplando cada fase do ciclo de vida dos sistemas de IA, que, se não tratadas, geram brechas para tais ataques (Jing et al. 2021).

Para a dimensão "segurança" , portanto, **proposição 2C** foi formulada como: Processo de gestão de segurança de sistemas de IA deve seguir diretrizes estabelecidas no nível estratégico da governança de IA.

3.4.6 Diretrizes para processos e práticas no domínio "desenvolvimento"

O processo de desenvolvimento e sustentação de sistemas de IA apresenta passos específicos, distintos daqueles dos demais sistemas. De uma maneira geral, o ciclo de vida de construção em IA distribui-se em: concepção e design, extração de dados, preparação dos dados, construção de modelo e validação, compondo a fase de projeto; e a implantação, monitoração, reavaliação e retirada de operação como etapas de uma fase em que o sistema está em operação no ambiente real (De Silva & Alahakoon 2022; Fjeld et al. 2020; ISO 2022a; Laato et al. 2022).

Como primeira decomposição das práticas da dimensão "desenvolvimento", considerou-se aquelas dirigidas às questões éticas das fases do sistema de IA em projeto e em operação, permitindo-se formular a **proposição 2D** como: Práticas que visam princípios éticos no processo de desenvolvimento e sustentação de sistemas de IA, em projeto e em operação, devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

O início do projeto requer clareza na identificação do problema e de como a IA seria aplicada para resolvê-lo parcial ou totalmente; se for o caso, qual política pública ele se associa (González et al. 2020), permitindo precisão na identificação de *stakeholders* (Wright et al. 2018; De Silva & Alahakoon 2022).

No modelo de pesquisa (figura 2), o processo de desenvolvimento de sistemas de IA abrange a fase de projeto e a fase de operação. O aprofundamento na dimensão "desenvolvimento" com foco na associação às ações estratégicas de governança de IA, ocorre nas práticas do processo de desenvolvimento e sustentação de sistemas de IA que visam princípios éticos em cada uma dessas fases (Laato et al. 2022; González et al 2020). Visando investigar ações inerentes a cada uma dessas fases (González et al. 2020; Rajkomar et al. 2018; De Silva & Alahakoon 2022; Laato et al. 2022),

previram-se duas proposições derivadas da proposição 2D, quais sejam: 2D1 para analisar a fase de projeto e 2D2 para analisar a fase de operação.

Enquanto se faz a análise de impacto dos *stakeholders (*Wright et al. 2018), aprofunda-se na tradução dos princípios éticos para regras de transmitam clareza quanto ao comportamento do sistema na perspectiva ética (Dennis et al. 2016; Rubicondo & Rosato 2022; IEEE 2021a). Um dos focos é a busca de pessoas e organizações possivelmente afetados, o que permite a definição de grupos e/ou atributos protegidos (Rajkomar et al. 2018; González et al 2020; ISO 2021a). Dilemas éticos são identificados e analisados à luz dos princípios éticos estabelecidos no nível de governança (Bench-Capon & Modgil; 2017; Bonnemains et al. 2018; Anderson & Anderson 2018; Zicari et al. 2021), priorizando-se as ações para compatibilizar com tais princípios (Locher & Bolander 2019; Awad et. al 2022, Ma et al. 2018; Schrader & Ghosh 2018). A análise de impacto nos *stakeholders* evolui em um ciclo de *design* em que as ações citadas se repetem até que regras e dilemas éticos estejam completamente definidos e documentados.

A fase de extração e preparação dos dados requer atenção para entender suas características, formatos e qualidade, preparando-os para o pré-processamento (De Silva & Alahakoon 2022; Commission de Surveillance du Secteur Financier 2018), ocasião que requer aplicação de técnicas para identificação e prevenção de alguns vieses com o foco na amostra de dados (González et al 2020; Rubicondo & Rosato 2022) como vieses históricos, vieses de representação, vieses de população, vieses de amostra, paradoxo de Simpson, entre outros vieses de dados (Ashokan & Haas 2021; ISO 2021a; Baeza-Yates 2018; Oneto & Chiappa 2020). Considerando possíveis ajustes na amostra de dados (González et al 2020), podem ocorrer muitas iterações até que se tenha segurança de seguir para a construção do modelo.

A construção de modelos e validação envolve pesquisa por algoritmos, constituindo uma fase que requer importantes decisões quanto ao desenvolvimento. Se alguns erros ocorrem, novas brechas para vieses aparecem (González et al 2020; De Silva & Alahakoon 2022; Commission de Surveillance du Secteur Financier 2018). Trata-se de nova oportunidade de se encontrar problemas como: “contaminações” de dados no treinamento, dados selecionados que não serão utilizados, problemas de classes desbalanceadas (González et al 2020, Vetrò et al. 2021), e muitos vieses cognitivos (Mehrabi et al. 2019), vieses de algoritmos, vieses de variáveis omissas, vieses temporais, entre outros (González et al. 2020; Ashokan & Haas 2021; Rajkomar et al. 2018). Nesta etapa, busca-se *algorithmic fairness* como linha de pesquisa para definir, analisar e minimizar

potenciais vieses (Ashokan & Haas 2021; Abdollahi & Nasraoui 2018), os quais podem ter origem nos vieses, na formulação matemática (paridade preditiva e a igualdade de oportunidades), e legislação antidiscriminação, se existente (Makhlouf et al. 2021).

Adicionalmente, como prática contínua durante o desenvolvimento do sistema de IA, mecanismos para provimento de transparência são necessários com o propósito de construir uma documentação que permita uma explicação adequada dos resultados do sistema de IA, de maneira a reduzir a opacidade algorítmica (Arrieta et al. 2020; AI HLEG 2019b). Padrões de documentação, como por exemplo, o XAI de maneira a reduzir a opacidade algorítmica (Das 2020; Dazeley et al. 2021; Adadi & Berrada 2018; Phillips, et al. 2021) até que testes automatizados indiquem o momento da implantação do sistema de IA no ambiente real (González et al. 2020; Laato et al. 2022, Commission de Surveillance du Secteur Financier 2018; ISO 2022a).

Visando as práticas previstas na fase de projeto - representação de regras e dilemas éticos, práticas para minimizar vieses, e práticas para prover transparência na produção dos sistemas de IA, formulou-se a **proposição 2D1** como: As práticas de representação de regras e dilemas éticos, práticas para minimizar vieses, e práticas para prover transparência do processo de desenvolvimento dos sistemas de IA, devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

A sensibilidade em relação a variações no contexto no qual o sistema de IA foi inserido, aliada ao fato de que modelos de IA são menos complexos que realidades sociais (Strauß 2021), impõem monitoração contínua do sistema após ter sido disponibilizado ao uso (Laato et al. 2022; ISO 2022a; Rubicondo & Rosato 2022). Essa prática, compartilhada com o processo de gestão de riscos, inicia-se por meio da monitoração de desempenho automática (Fjeld et al. 2020; González et al 2020; De Silva & Alahakoon 2022; ISO 2022a); e robustecida com uma permanente investigação por vieses cujas origens possam ser profundas e externas ao contexto tecnológico, o que impõe a supervisão humana do comportamento do sistema de IA (Strauß 2021; Fjeld et al. 2020; González et al 2020; Zicari e*t al*. 2021; Dignum 2019; Hickman 2020).

A necessidade de obter *feedback* dos usuários impõe que esta prática passe a compor as fases de um sistema de IA em operação, em contínua observação de novas interpretações dos usuários (Rahwan et al. 2019; Wright & Schultz 2018, Dignum 2019; Hickman 2020, de Almeida et al. 2021; AI HLEG 2019b; AI4People 2018).

A combinação das ações de acompanhamento (monitoração automática, supervisão humana e coleta de *feedback*) permite encontrar explicações para o comportamento do sistema; validade das operações como foco na acurácia atingida; plausabilidade de maneira a não haver dúvidas quanto aos resultados, podendo esses serem reproduzíveis; e por fim, saber se são aceitáveis em uma perspectiva ética (Strauß 2021). Enquanto o sistema estiver operacional, as contribuições recebidas da monitoração automática, da supervisão humana, e do *feedback* dos usuários podem motivar evoluções, criando-se, assim, um contínuo *loop* em toda vida do sistema (De Silva & Alahakoon 2022; Laato et al. 2022), até o momento em que ele for retirado de operação (ISO 2022a).

Portanto, visando as práticas previstas na fase de operação relativas aos princípios éticos dos sistemas de IA, formulou-se a **proposição 2D2** como: As práticas de monitoração automática, supervisão humana, e coleta de *feedback* devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.

3.4.7 Treinamento de pessoas para Governança de IA

Na busca por mecanismos que garantam uma IA confiável, passa ser imperativa a criação de uma cultura com menor viés nas organizações (Awad et al. 2020; Ma et al. 2018). Pesquisas constatam a importância de uma percepção positiva e familiarizada da IA no engajamento de servidores públicos para projetos de serviços digitais baseados em IA, e a consequente necessidade de provimento do correto conhecimento a esse corpo funcional (Ahn & Chen 2022; Benfeldt et al. 2020), como ocorreu com a parceria do Governo da Finlândia com a iniciativa privada (Makarius et al. 2020). Na academia, estudos constatam a lacuna de conhecimento socio-ético na formação de profissionais de TI (Floridi 2018); e no campo organizacional, sugerem-se ações específicas à diversidade na formação de equipes, nos recrutamentos e nos processos decisórios (Dignum 2022; Vanhée & Borit 2022) em diversos níveis hierárquicos (Zuiderwijk et al. 2021; Bonnemains et al. 2018; Leitner & Stiefmueller 2019; Ma et al. 2018; Strauß 2021; Awad et al. 2020).

Considerado como habilitador para a implantação da governança de IA nas organizações, o treinamento de *stakeholders* em dados, em desenvolvimento de sistemas de IA, e em princípios éticos aplicados à IA (Calzada & Almirall 2020; Micheli et al. 2020; Ruijer 2021 e Benfeldt et al. 2020) fundamentou a **Proposição 3** formulada como: O treinamento de *stakeholders* em dados, em

desenvolvimento de sistemas de IA e em princípios éticos aplicados ao desenvolvimento de sistemas de IA é habilitador da implantação da governança de IA nas organizações públicas.

Com o foco nos *stakeholders* diretamente envolvidos nas deliberações e execuções relativas à gestão do desenvolvimento e sustentação de sistemas de IA, as pesquisas de Ahn e Chen (2022) e de Benfeldt et al. (2020) indicam a necessidade de capacitar *stakeholders* tanto da esfera gerencial, quanto técnica, para lidar com desafios da implantação da processo e práticas para lidar com as questões éticas dos sistemas de IA, como, por exemplo, a redução da opacidade algorítmica (Tutt 2017; Butterworth 2018; Buiten 2019). Em uma evolução da proposição 3, formulou-se nova proposição para compreender se a capacitação de gestores e desenvolvedores, ao longo do tempo, amplia a decisão por ter acesso aos códigos dos sistemas produzidos e contribui positivamente com a implantação da governança de IA nas organizações públicas. Pelo exposto, formulou-se a **proposição 4** como: Ao longo do tempo, o treinamento de gestores e desenvolvedores de sistemas de IA amplia a probabilidade das organizações públicas criarem condições de terem acesso aos códigos dos seus sistemas de IA e avançarem na implantação da governança de IA.

3.4.8 Integração dos processos auxiliares à Governança de IA

Em sua essência, a proposta de uma governança de IA fundamenta-se na abordagem de gestão de riscos das organizações e no ambiente onde se inserem (Wirtz et al. 2022; Vetrò et al. 2021; Zuiderwijk et al. 2021). Contudo, os processos tradicionais de gestão de riscos, amparados por métodos quantitativos (Chen & Deng 2022; Duijm 2015; ISO 2018), e por isso amplamente aceitos, têm sido criticados para aplicação em alguns domínios, requerendo abordagens complementares por meio de processos multidisciplinares (*multi-stakeholders*) que utilizem técnicas que integrem diferentes visões e assim permitindo uma camada qualitativa (Gerkensmeier & Ratter 2018; Fernandes et al. 2021; European Union Agency for Cyber Security 2022), de maneira a enriquecer análises necessárias aos propósitos da governança de IA (Vetrò et al. 2021; Vining et al. 2022; Breier et al. 2020; De Silva & Alahakoon 2022).

A integração do processo de gestão de riscos com os processos de gestão de qualidade de dados (Medaglia 2021), de gestão da segurança de dados (De Silva & Alahakoon 2022; Breier et al. 2020), de gestão da proteção de dados pessoais (Wirtz et al. 2022; De Silva & Alahakoon 2022), e de gestão do desenvolvimento e sustentação de sistemas de IA ocorre envolvendo *stakeholders* da gestão de riscos com aqueles próprios dos demais processos (De Silva & Alahakoon 2022). Pelo

exposto, formulou-se a **proposição 5** como: Para implantar a governança de IA nas organizações públicas, integra-se o processo de gestão de riscos aos processos de gestão da qualidade de dados, de gestão da proteção de dados pessoais, de gestão da segurança para sistemas de IA, de auditoria nos sistemas de IA e as práticas do processo de desenvolvimento e sustentação de sistemas de IA aplicadas aos princípios éticos.

3.5. Métodos e Técnicas de Pesquisa

Esta tese tem caráter exploratório e descritivo. A natureza exploratória se faz necessária dado o fato de que, apesar de muitos trabalhos sobre riscos associados ao uso e produção inadequada da IA, pesquisas sobre Governança da IA ainda compõem uma dimensão subdesenvolvida (Taeihagh 2021). A natureza descritiva decorre do objetivo de identificar como um determinado fenômeno está ocorrendo em organizações (Vergara 2005), conforme recomendam Mäntymäki et al. (2022) e Zuiderwijk et al. (2021) para que se investigue a lacuna existente na literatura de como organizações têm interpretado e incorporado, em suas práticas, processos e estrutura, os princípios éticos na produção de seus sistemas de IA.

A investigação foi realizada por meio de pesquisa empírica (Richie & Lewis 2003), de maneira a preencher a lacuna identificada por Zuiderwijk et al. (2021) para uso de métodos orientados a dados, com abordagens exploratórias além de análises qualitativas e quantitativas, para obter aprofundamento da Governança da IA no setor público sob a perspectiva das questões éticas.

3.5.1 Estratégia de seleção da amostra e coleta de dados

Sendo a Governança da IA uma necessidade global (Fjeld et al. 2020; OECD 2022a), procurou-se construir amostra de dados, a partir de uma população com alcance nos cinco continentes, onde existam organizações do setor público, em qualquer esfera de poder – Executivo, Legislativo ou Judiciário (Maluf 1995), com preferência de abrangência nacional e que atendesse à recomendação de Hair et al. (2009) para mínimo de cinco respondentes.

Considerando o interesse em investigar processos e práticas nos níveis de governança, de gestão e de desenvolvimento de sistemas de IA, entendeu-se como necessária a critério que a organização possua, pelo menos, um sistema de IA em operação no ambiente real, constituindo-se assim, parte de ações contínuas da organização em sua prestação e serviços públicos. Como consequência, não foram incluídas na população organizações que possuam somente sistemas de

IA em estado de protótipo, ou projetos em estágio muito embrionário, porque, provavelmente, não tenham ainda gerado o envolvimento necessário dos processos organizacionais.

O processo de identificação da população, seleção da amostra e coleta de dados ocorreu nos meses de agosto, setembro, outubro de 2022, constituindo-se de vários passos em linhas paralelas de pesquisa envolvendo diversos atores e fontes de informações, como apresentado na figura 3.

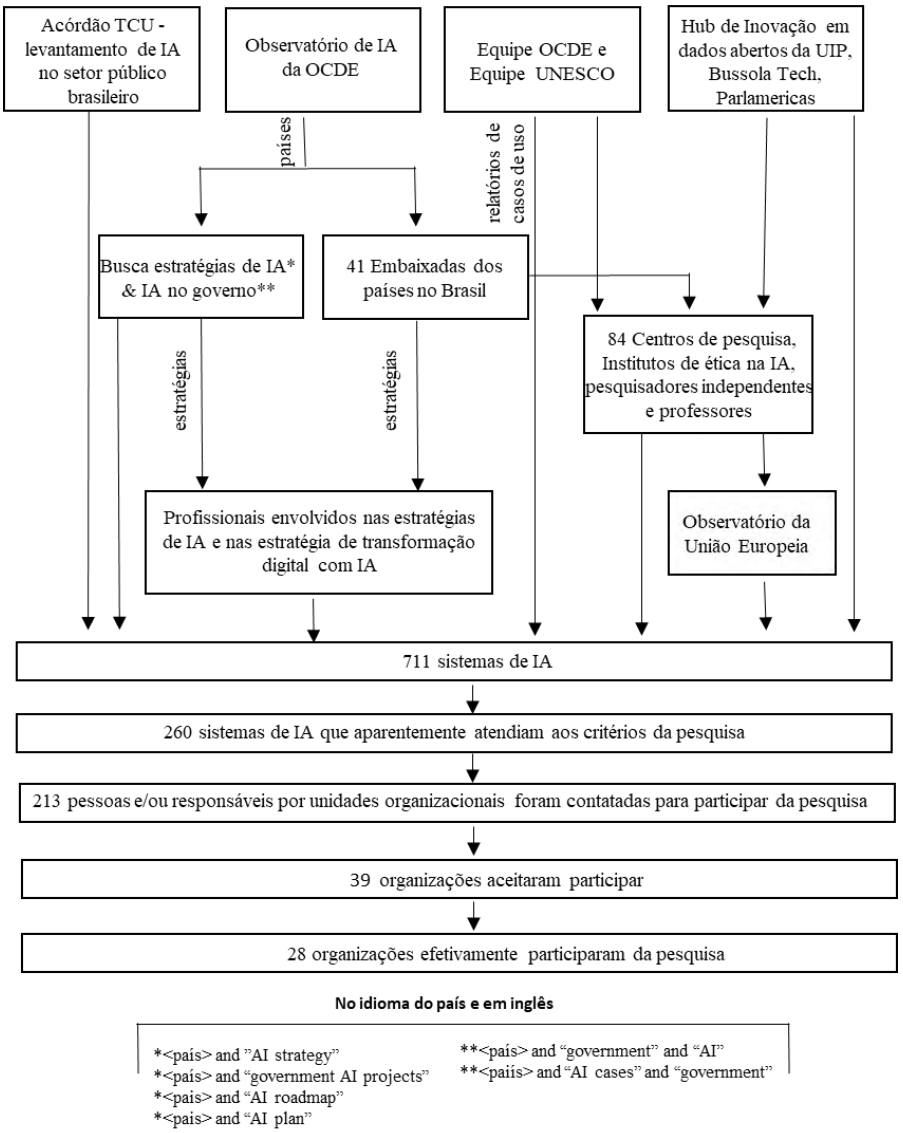

Figura 3: Processo de seleção da amostra
Fonte: Elaboração própria.

Iniciou-se pela pesquisa de países em que teriam, conforme fontes oficiais, organizações públicas utilizando sistemas de IA. O Observatório de Inteligência Artificial da OCDE (OECD 2022a) foi a primeira fonte de informações, o qual apresenta iniciativas de dezenas de países na estratégia de IA, na produção e na regulamentação da inteligência artificial. A partir da relação de

países extraída da OCDE, abriram-se novas frentes de busca com os responsáveis pelas estratégias de IA (tabela 3) , com envolvimento de embaixadas desses países no Brasil (MRE 2022), Observatório de Inteligência Artificial da Comissão Europeia (European Commission (2018b), União Interparlamentar (IPU 2022), Bussola Tech (2022), ParlAmericas (2022), Tribunal de Contas da União do Brasil (TCU 2021, 2022).

Tabela 3: Estratégias de IA identificadas durante a pesquisa.

País/Região	**Estratégia de IA/Estratégia Digital Contemplando IA**
Alemanha	German Federal Government (2020)
Argentina	Presidencia de la Nación (2019)
Australia	Australian Government (2021)
Austria	Government of Austria (2018)
Brasil	Ministério de Ciência, Tecnologia e Inovação do Brasil (2021)
Canadá	Government of Canada (2022b)
Chile	Gobierno de Chile (2020)
Colômbia	Republica de la Colombia (2019)
Dinamarca	Danish Government (2019)
Egito	The National Council for Artificial Intelligence (2020)
Emirados Árabes Unidos	United Arab Emirad (2018)
Espanha	Government of Spain (2020)
Estados Unidos	United States Government (2021)
Estonia	Government of the Republic of Estonia (2019)
Finlândia	Ministry of Economic Affairs and Employment of Finland (2017) Ministry of Economic Affairs and Employment of Finland (2019)
França	Gouvernment de France (2021)
Holanda	Amsterdam Data Science (2018)
Hungria	Hungarian Ministry for Innovation and Technology (2020)
India	NITI Aayog (2018) Government of India (2020b)
Indonesia	Sekretariat Nasional Kecerdasan Artifisial Indonesia (2020)
Irlanda	Government of Ireland (2021)
Itália	Ministero dello sviluppo economico (2019)
Japão	Japanese Strategic Council for AI Technology (2017)
Letônia	Ekonomikos ir Inovaciju Ministerija (2018)
Luxemburgo	Government of the Grand Duchy of Luxembourg (2018)
Noruega	Norwegian Ministry of Local Government and Modernisation (2020)
Polônia	Rzeczypospolitej Polskiej (2020)
Portugal	República Portuguesa (2021)
Reino Unido	Government of United Kingdom (2021e)
República da Coreia	Republic of Korea Government (2019)
Singapura	Government of Singapore (2019)
Suécia	Government of Sweden (2020)
União Africana	African Union (2018)
União Europeia	European Commission (2018a)

Fonte: Elaboração própria

Tais contatos evoluíram, por meio de reuniões remotas e mensagens de email e redes sociais, até obter contados de pesquisadores, de centros governamentais de produção e pesquisa sobre sistemas de IA, gestores responsáveis por implantação de estratégias de IA ou de transformação digital que apresentasse casos de uso da tecnologia.

As respostas, quando recebidas, entravam numa sequência de troca de mensagens e reuniões remotas até culminar com contatos de organizações e/ou pessoas de órgãos públicos conhecidos por terem iniciativas de IA em seu portfólio.

O percurso descrito culminou com 711 sistemas de IA em desenvolvimento e/ou uso por organizações públicas nos cinco continentes, apontados por diferentes meios de comunicação (Tangi et al. 2022; European Commission 2021b; IPS-X 2022, OECD/CAF 2022; Government of India 2020a; FCT 2021; WEF 2020a, 2020b, 2020c; Misuraca & Van Noordt 2020). Após remoção das redundâncias e dos sistemas que estavam divulgados como não operacionais (em estudo embrionário, em protótipo, ou já haviam sido retirados de uso), o refinamento prosseguiu até alcançar os contatos de organizações que atendiam aos critérios da pesquisa. Após algumas trocas de mensagens, 39 organizações aceitaram participar da pesquisa. Após disponibilizados os questionários, a pesquisadora ficou à disposição para eventuais dúvidas. Apenas 28 organizações efetivamente responderam às questões.

3.5.2 Instrumentos de coleta de dados

Para as análises quantitativa e qualitativa, foram utilizados dados primários, por meio de dois instrumentos de coleta - questionário online e entrevistas semiestruturadas, de maneira a se completarem na análise. As perguntas do questionário eram de múltipla escolha distribuídas nas seguintes seções: características da organização, características do respondente, características dos sistemas de IA da organização, modelos de governança, de gestão e de desenvolvimento de sistemas de IA, cujas questões se encontram no anexo 1. Com o objetivo de completar as informações do questionário, a entrevista (anexo 2) teve seu roteiro elaborado de maneira a se adaptar às respostas dele, e, assim, tentar compreender como algumas ações e decisões foram realizadas e ainda ocorrem nas organizações.

Tanto o questionário, quanto a entrevista requeriam conhecimentos básicos de como ocorre a governança e processos de gestão relacionados à produção de sistemas de IA, razão pela qual, considerou-se necessário um participante com perfil especializado, preferencialmente de unidade

de ciência de dados ou portfólio de projetos de sistemas de IA. Os questionários somente eram encaminhados após a concordância da organização em participar e a indicação da(s) pessoa(s) que a representaria na pesquisa. As entrevistas foram realizadas remotamente, por meio de plataforma de videoconferência da pesquisadora, exceto em duas organizações que preferiram usar sua própria plataforma.

Considerando a importância de um planejamento para aplicação dos instrumentos de coleta (Torlig et al. 2022), e que abordagens quantitativas podem ser utilizadas para avaliar ~~a~~ confiabilidade de pesquisas qualitativas (Morse et al. 2002), tanto o questionário, quanto o roteiro de entrevistas foram submetidos à avaliação por grupo de juízes, por meio de métodos indicados pela literatura para cada caso.

Para o questionário, foi utilizado o "Coeficiente de Validação de Conteúdo" (CVC) (Hernández-Nieto 2002; Silveira et al. 2017; Aburachid & Greco 2011), com avaliação de cada questão, em uma escala de pontuação de 1 a 5, quanto ao nível de clareza e de pertinência para a pesquisa (anexo 3). Utilizando as notas atribuídas pelos juízes, o CVC permite o cálculo: a) do CVC inicial (CVCi), erro provável (Pei), e CVC corrigido (CVCc) para cada pergunta, e em cada atributo (clareza e pertinência); b) CVC total (CVCt) de cada atributo analisado. Esta avaliação foi realizada por quatro avaliadores, o que atende à recomendação da literatura para que o número de juízes, especialistas no assunto, esteja entre três e cinco (Silveira et al. 2017).

O roteiro de entrevista foi avaliado por meio do método "Validação para Instrumentos de Pesquisa Qualitativa (VALI-QUALI) (Torlig et al. 2022), considerando as dimensões "Conteúdo" e "Semântica" (anexo 4). A avaliação de conteúdo previa pontuação para cada questão em relação aos atributos "alinhamento de cada questão ao objetivo da pesquisa", e para a "aderência da questão ao construto investigado". A análise semântica considerou os atributos "clareza" e a "expectativa qualitativa de resposta para cada pergunta". O VALI-QUALI, uma evolução do MRPQ (Torlig et al. 2019), recomenda a adoção de três perfis de juízes: o especialista prático, o especialista teórico, e o especialista metodológico em pesquisa qualitativa. Para atender a este último perfil, uma juíza foi incluída aos quatro juízes selecionados. Desta forma, foram quatro avaliadores para o questionário e cinco avaliadores para o roteiro de entrevista. Para cada questão e cada atributo, os juízes dão pontuação de 1 a 5, de maneira a se calcular Qi (Indicador VALI-QUALI para a questão i). Os resultados foram comparados aos critérios sugeridos por Torlig et al. (2022) e reproduzidos no anexo 4.

Como proposto por Torlig et al. (2022), foram realizados pré-testes utilizando pessoas de perfil semelhante ao público-alvo para confirmar o alinhamento com os objetivos da pesquisa (Manzini 2004). Tendo em vista o fato de que a entrevista foi planejada para ser complementar ao questionário, o pré-teste foi aplicado para o conjunto completo: questionário e entrevista. Na figura 4, pode-se observar a evolução das validações dos instrumentos de coleta de dados, e nos anexos 1 e 2, as versões finais dos instrumentos de coleta de dados.

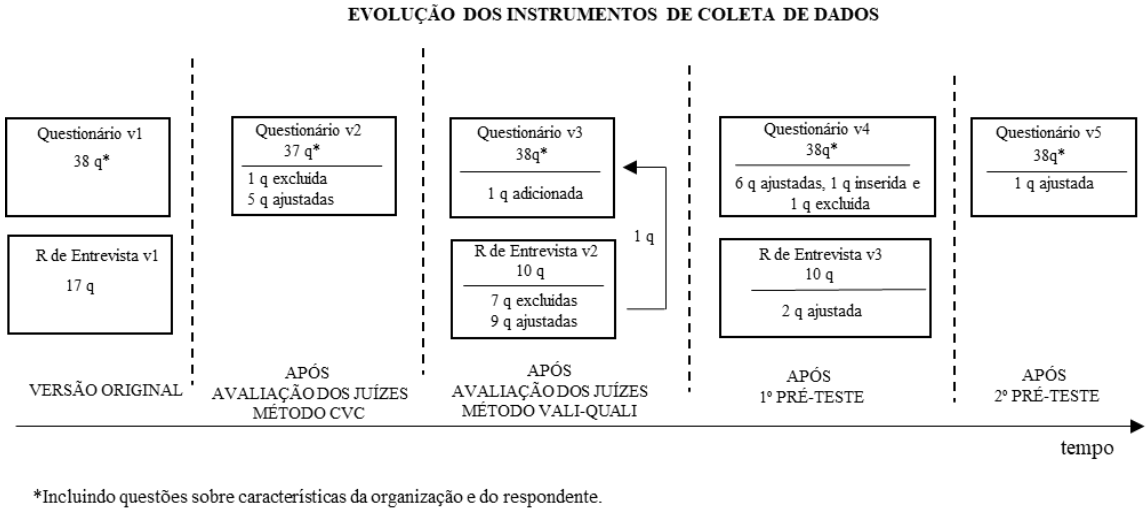

Figura 4: Evolução dos instrumentos de coleta de dados.
Fonte: Elaboração própria

3.5.3 Estratégia de análise

Realizou-se a análise dos dados primários dos casos da amostra, obtidos pelos instrumentos de coleta, utilizando-se uma combinação da *Qualitative Quantitative Analysis* – QCA (Rihoux & Ragin, 2008; Ragin 2008), e de análise de conteúdo das entrevistas e dos documentos compartilhados (krippendorff 2013; Saldaña 2013).

3.5.3.1 QCA

A QCA é uma técnica de pesquisa qualitativa que também considera aspectos quantitativos, utilizada nas ciências sociais (Dias 2011), e recentemente utilizada nas pesquisas sobre gestão nas organizações (Rihoux & Ragin 2008), baseada na Teoria dos Conjuntos e em operações booleanas para estabelecer as relações lógicas entre os conjuntos (Freitas & Neto 2016; Betarelli-Júnior & Ferreira 2018), para resolver problemas provocados pela necessidade de se fazer inferências causais em estudos de casos. Interpretam-se os dados qualitativamente, enquanto também se

procuram relações de causalidade entre as variáveis, admitindo-se situações em que algumas variáveis não apareçam na coleta (Dias 2011).

A adoção da QCA em pesquisas sociais nasceu da identificação de limitação da maioria das aplicações de métodos quantitativos convencionais ao assumirem que os efeitos das variáveis independentes são lineares e aditivos, onde o impacto de uma determinada variável independente na variável dependente é considerado o mesmo, a despeito dos valores das outras variáveis (Ragin 2008). O método busca mostrar quais condições ou combinações de condições ocorreram em um cenário de um resultado esperado (Rihoux & Ragin 2008). Duas estratégias podem ser utilizadas para a investigação: casos que compartilham um dado resultado, ou casos que compartilham as mesmas combinações de condições causais para identificar se possuem o mesmo resultado. Tecnicamente, as análises consistem em examinar se as instâncias de um resultado específico são subconjuntos de instâncias de uma condição causal ou se instâncias de combinações causais são subconjuntos da instância de um resultado.

Denominado de método configuracional, a QCA propõe-se a analisar os casos, preservando suas configurações complexas a partir de suas características qualificadas e quantificadas, de maneira a realizar análises comparativas, por meio de associações entre determinadas condições e o resultado (Freitas & Neto 2016), em lugar de correlações (Ragin 2008; Korjani & Mendel 2012). Considera-se uma condição como necessária para um determinado resultado, se estiver sempre presente quando este ocorrer, ou seja, o resultado não ocorrerá na ausência da condição. E uma condição é suficiente para um determinado resultado, se este sempre ocorrer quando a condição for presente (Rihoux & Ragin, 2008). Portanto, com a necessidade, o resultado é um subconjunto da condição causal (CONDIÇÃO $\supset$RESULTADO); e, com suficiência, a condição causal é um subconjunto do resultado (CONDIÇÃO $\subset$ RESULTADO) (Betarelli-Júnior & Ferreira 2018).

Freitas e Neto (2016) e Bertarelli & Ferreira (2018) fizeram uma comparação entre os métodos quantitativos convencionais, geralmente de base estatística, e a QCA, reproduzidas na tabela 4, onde se destacam os pressupostos de equifinalidade, causalidade conjuntural e causalidade assimétrica (Betarelli-Júnior & Ferreira 2018; Rihoux & Ragin, 2008). A equifinalidade é a propriedade de que várias combinações de condições conduzem a um mesmo resultado. A causalidade conjuntural representa as condições não necessariamente conduzindo ao resultado de modo isolado uma da outra, mas podendo ser combinadas entre si para revelar padrões causais de um resultado. E a causalidade assimétrica representa o fato de que não somente a ocorrência do

fenômeno requer análise separada, mas também sua ausência, porque a presença ou ausência das condições podem produzir diferenças no resultado (Rihoux & Ragin 2008).

Tabela 4: Comparações entre QCA e os métodos quantitativos convencionais

Principais diferenças entre QCA e técnicas quantitativas

Técnicas Quantitativas Tradicionais	Análise Qualitativa Comparativa	
Variáveis	Conjuntos	Fonte: Betarelli & Ferreira (2018)
Variável dependente	Resultado	
Variáveis Independentes	Condições	
Correlações	Relações entre conjuntos	
Matriz de correlação	Tabela Verdade	
Efeitos líquidos das variáveis	Caminhos causais	
Relações de aditividade e lineares	Relações não aditivas	
Causalidade múltipla ou singular	Causalidade conjuntural múltipla	
Universalidade ou equifinalidade	Equifinalidade	
Unifinalidade	Multifinalidade	
Causalidade simétrica	Causalidade assimétrica	
Análise dos efeitos das variáveis	Análise dos efeitos das configurações	
Amostra aleatória	Seleção intencional dos casos para incluir casos típicos	Fonte: Freitas & Neto (2016)
Generalização estatística	Generalização modesta, limitada no tempo e no espaço	
Causalidade única ou múltipla	Causalidade múltipla conjuntural	
Desmembra os casos em um conjunto de variáveis independentes	Desmembra casos em um conjunto de atributos inter-relacionados.	
Foco nas variáveis e nas relações entre variáveis causais e dependentes	Foco em configurações de variáveis que gerem diferentes resultados	

Buscando uma generalização limitada no tempo e no espaço, em lugar da generalização estatística, a QCA pode ser utilizada para verificar se os dados são coerentes com as alegadas relações entre os conjuntos; testar hipóteses e teorias; dar uma visão global sobre as suposições básicas da análise; desenvolver novos argumentos teóricos; e criar tipologias empíricas (Betarelli-Júnior & Ferreira 2018). Estendendo-se em situação de pequenas e médias amostras, Dias (2011) apresenta o uso da QCA para amostras de 3 a 250 casos.

A utilização das relações de conjuntos nas pesquisas sociais para envolver conexões causais ou outras conexões integrais que ligam fenômenos sociais, possuem forte dependência da teoria que suporta a pesquisa, requerendo uma explicação específica (Ragin 2008). Conforme a Teoria dos Conjuntos, na QCA operações booleanas entre os conjuntos podem ser realizadas para situações de conjunção [E(*)], disjunções[OU(+)], e negação [NÃO(~)]. O núcleo das técnicas

com utilização de QCA requer a construção de tabelas-verdade nas quais cada linha representa uma combinação de condições logicamente possíveis (conjunções) (Betarelli-Júnior & Ferreira 2018).

A presente pesquisa utilizou o *crisp-set* QCA (admite os valores 0 e 1, respectivamente para ausência ou presença da relação entre os conjuntos) e a variação *fuzzy* QCA por permitir mais precisão em função de sua maior flexibilidade ao permitir a utilização de um conjunto contínuo de valores no intervalo 0 (ausência completa de membresia) e 1 (membresia completa) (Ragin 2008; Rihoux e Ragin 2008).

De maneira análoga aos conjuntos dicotômicos, os conjuntos *fuzzy* podem ser analisados sob a mesma Teoria dos Conjuntos, ou seja, A é um subconjunto *fuzzy* de um conjunto *fuzzy* B, se os escores de membresia dos casos em A são menores ou iguais aos escores de membresia dos casos correspondentes em B. Em representação matemática, sejam dois conjuntos *fuzzy*, A={s1,s2,.., sn} e B={g1,g2,.., gn}, e si, gi, os escores de i-ésimo caso em cada conjunto, e si ϵ [0;1] $\subset$ R, gi ϵ [0;1] $\subset$ R, $\forall$i; então, A $\subset$ B se si $\leq$ gi, $\forall$i.

Tanto Ragin (2008) quanto Freias e Neto (2016) sugerem uma calibragem dos valores originais dos conjuntos, passando-os para os conjuntos *fuzzy*, cujos valores são distribuídos no intervalo entre 0 e 1, a partir do nível de presença das condições no conjunto-resultado, representando em um extremo com a completa exclusão , e do outro, a completa inclusão do conjunto. A calibragem requer a definição do valor do ponto de cruzamento (*crossover point)* de cada variável-conjunto, ou seja, o valor, na dimensão da matriz de dados, que representa a membresia 0,5 numa escala *fuzzy* (ponto de indiferença, valor que não caracteriza um conjunto estar dentro nem fora de outro conjunto). A partir dos valores fuzzy, gera-se a tabela verdade com a indicação de presença ou ausência da combinação de soluções e resultados (Ragin 2008; Schneider & Wagemann 2012; Betarelli-Júnior & Ferreira 2018; Meijerink & Bondarouk 2018).

O *fuzzy* QCA permite o cálculo de três possíveis soluções para cada análise: complexa, parcimoniosa e intermediária, relacionadas entre si. A solução complexa considera as combinações de condições consistentes com o resultado, ou seja, ignora os remanescentes lógicos do processo de minimização da tabela verdade. As demais condições consideram, além das combinações contempladas na solução complexa, remanescentes lógicos. Na solução parcimoniosa, esses remanescentes conduzem a soluções mais simples possível, e a solução intermediária, somente os remanescentes que são contrafactuais. Assim, as soluções parcimoniosa e intermediária são conjuntos maiores que os conjuntos da solução complexa (Betarelli-Júnior & Ferreira 2018).

Similarmente, os conceitos de suficiência e necessidade também se aplicam aos conjuntos *fuzzy*, pois se os conjuntos A e B forem teoricamente relacionados de tal modo que A é condição para o resultado B, então A é condição necessária, mas, não suficiente para o resultado B, quando as instâncias do resultado B, constituem um subconjunto das instâncias da causa (Schneider & Wagemann 2012).

O principal critério de validação do *fuzzy* QCA é a medida de consistência, cujo propósito é mensurar a proximidade da relação entre conjuntos, indicando o grau em que os casos que compartilham uma condição, ou combinação de condições, concordam com o resultado, sendo um valor compreendido entre 0 e 1 (Betarelli-Júnior & Ferreira 2018). Consistências próximas a 1 indicam que (quase) todos os casos que compartilham uma condição causal também compartilham o resultado. Ragin (2008) considera aceitável consistência $\geq 0{,}75$ para análises de condições necessárias. Complementando à interpretação dos resultados, a medida de cobertura oferece a quantificação da relevância empírica de uma condição ou combinação causal no conjunto das combinações causais (Thiem 2010); ou seja, a quantidade de casos com o resultado é representada por uma condição causal em particular (Rihoux & Ragin 2008), cujo valor reside no intervalo entre 0 e 1. A existência de múltiplos caminhos para um mesmo resultado implica coberturas baixas para cada combinação (Ragin 2008). Em conjuntos dicotômicos, a consistência e a cobertura podem ser visualizadas por meio de diagramas de Venn como na figura 4.

Figura 4: Diagramas Venn de condições suficientes consistentes e inconsistentes.
Fonte: Adaptado de Schneider e Wagemann (2012).

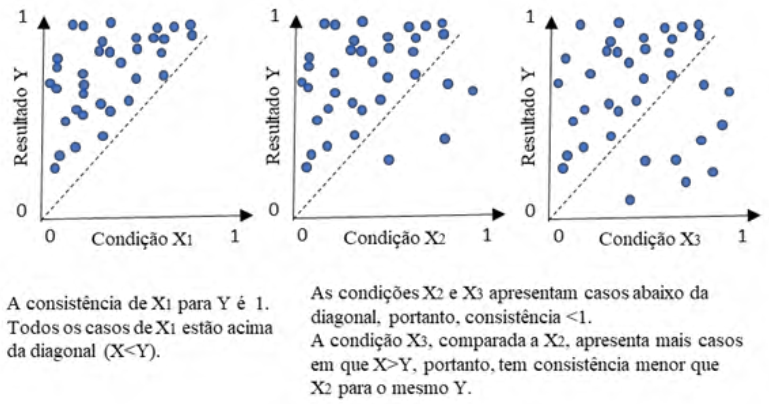

Figura 5: Diagramas XY de condições suficientes consistentes e inconsistentes
Fonte: Adaptado de Schneider e Wagemann (2012)

Em conjuntos *fuzzy*, os diagramas XY são mais adequados para o cálculo da consistência (figura 5), a qual é definida matematicamente como a equação 1: Consistência(Xi≤Yi) = $\sum \frac{MIN(Xi,Yi)}{\sum Xi}$,∀i; onde X é o escore de membresia na combinação causal e Y é o escore de membresia no resultado. O cálculo da cobertura em conjuntos *fuzzy* é obtido como na equação 2: Cobertura(Xi≤Yi) = $\sum \frac{MIN(Xi,Yi)}{\sum Yi}$,∀i; onde X é o escore de membresia na combinação causal e Y é o escore de membresia no resultado.

3.5.3.2 Plano de análise

Conforme roteiro de análise exposto na figura 5, aplicou-se o *crisp-set* QCA para análises com variáveis-conjunto dicotômicas e *fuzzy* QCA para aqueles de valores contínuos, utilizando informações recebidas do questionário e das entrevistas associadas aos construtos definidos no modelo de pesquisa. A parcela de informação recebida das entrevistas foi extraída a partir de análise de conteúdo, considerando os construtos mapeados no modelo de pesquisa e no questionário. Por meio da QCA, analisaram-se proposições fundamentadas na literatura, a partir de condições e resultados observados nos dados coletados por meio do fluxo Q1E1. Para as análises *crisp-set QCA*, utilizou-se o software TOSMANA versão 1.6.1 *(*https://www.tosmana.net/*)*, e para o *fuzzy* QCA, usou-se o fsQCA versão 3.1 (http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml) .

A análise do fluxo E1, recebido das entrevistas semiestruturadas, permitiu a atribuição de valores a algumas variáveis-conjunto utilizadas na QCA, assim como a análise de conteúdo e identificação de novas informações associadas aos construtos definidos no modelo de pesquisa,

utilizando-se o software MAXQDA 2022 (https://www.maxqda.com/). Também foi possível a identificação de novas categorias no conteúdo analisado (fluxo E2), e; por fim, a união das duas análises (fluxo Q1E1E2) permitiu a consolidação das discussões para se chegar às conclusões.

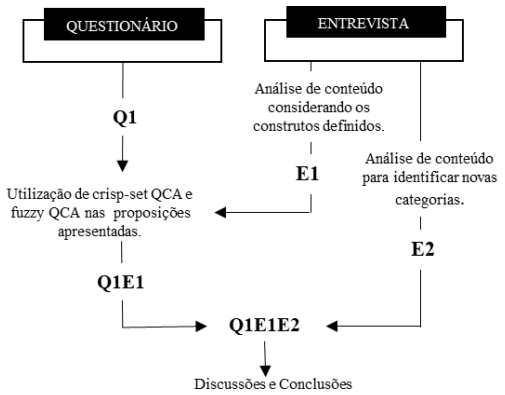

Figura 5: Plano de análise.
Fonte: Elaboração própria

Para o fluxo Q1E1, a atribuição de valores às respostas não dicotômicas relacionados às práticas de governança, de gestão e de desenvolvimento de sistemas de IA, tanto do questionário (anexo 1), quanto da entrevista (anexo 2), ocorreram considerando-se o contexto em que os temas "Governança de IA" e "Ética na IA" são pouco difundidos entre as organizações (Taeihagh 2021), gerando abordagem não profunda quanto às práticas necessárias para implantação da governança de IA (Zuiderwijk et al. 2021; Mäntymäki et al. 2022a) e com raros casos de legislação estabelecida, apesar de muitas discussões nas casas legislativas. Sob o cenário descrito, optou-se por simplificar a manifestação do participante da pesquisa (tabela 5), sem exigir graus de resposta refinados que sugerissem níveis de maturidade de tais práticas.

Tabela 5: Critério de pontuação às respostas do questionário.

OPÇÃO DE RESPOSTA	SIGNIFICADO	PONTUAÇÃO
Sim, completamente.	A organização possui ou implantou o item com todos, ou a maior parte dos requisitos necessários.	100
Sim, parcialmente.	A organização possui ou implantou o item contemplando menor parte dos requisitos necessários. Alguns exemplos: limitação de recursos, projeto ainda em andamento, projeto implantado contemplando requisitos básicos, programas distribuídos em muitos projetos.	67
Não, mas, há decisão formal para implantar.	A organização decidiu formalmente implantar, porém, ainda não iniciou projeto para esse fim.	33
Não, nem há decisão formal para implantação.	A organização não implantou, nem decidiu formalmente por implantar o item.	0

Fonte: Elaboração própria.

3.5.3.3 Variáveis-conjunto

Fundamentadas no modelo de pesquisa (figura 2), as variáveis-conjunto utilizadas nas QCA (anexo 5) endereçam perguntas do questionário e algumas perguntas das entrevistas.

No construto "Fatores Geradores de Expectativas", previram-se variáveis-conjunto dicotômicas para representar a existência de legislação (*hard law)* para regular a IA (Fjeld et al. 2020) e para proteção de dados pessoais (Information Commissioner's Office 2020, EU GDPR 2016; Benfeldt et al. 2020; Rhahla et al. 2021), dada a estreita relação entre elas (Bogucki et al. 2022; Georgieva et al. 2022), além da existência de política governamental de boas práticas para uso e provimento de sistemas de IA - *soft law* (Marchant 2019; Gutierrez & Marchant 2021) (anexo 5).

Compondo o construto "Governança de IA", um primeiro grupo de variáveis-conjunto definidas distribuem-se no nível estratégico de uma organização pública, e atuam na formalização de uma governança de IA: estratégia de IA, política para sistemas de IA, declaração de princípios éticos dos sistemas de IA, processo de governança de IA, estrutura/comitê para governança de IA. Um segundo grupo associado ao mesmo construto envolve habilitadores para a governança de IA, treinamento para tomadores de decisão, treinamento para desenvolvedores de sistemas de IA, treinamento para usuários, treinamento para auditores internos (anexo 5).

Compondo o construto "Processos e práticas auxiliares", as variáveis-conjunto representativas dos processos que auxiliam a governança de IA: processo para governança de dados, processo para

gestão da qualidade de dados, processo para gestão de dados pessoais, processo para gestão de riscos dos sistemas de IA, processo para gestão da segurança de sistemas de IA, processo para auditoria interna nos sistemas de IA, monitoração de mudanças de ambiente e tendências sociais (anexo 5).

O construto "Gestão do Desenvolvimento e Sustentação de sistemas de IA" contemplou apenas as práticas dirigidas aos princípios éticos do ciclo de vida de um sistema de IA: identificação de *stakeholders*, representação formal dos princípios e dilemas éticos nas regras de negócio, práticas para minimizar vieses, transparência do processo de desenvolvimento de sistemas de IA, monitoração automática, supervisão humana na busca por vieses, e coleta de *feedback* dos usuários (anexo 5). Não foi exigida a normatização desses processos por meio de instrumentos formais, deixando livre o entendimento de que a existência de uma sequência de ações para determinado fim pudesse ser considerada como processo.

Tabela 6: Critérios de interpretação das variáveis-conjunto extraídas das entrevistas.

Variável-conjunto	**Critério**	**Pontuação**
Processo de governança de IA	Há instrumento formal estabelecendo as atribuições e os responsáveis sobre as decisões estratégicas quanto o uso da IA na organização **OU** Há comitê formalmente constituído que, entre suas atribuições, encontram-se as deliberações sobre a IA.	100
	Há clareza quanto às atribuições e os responsáveis sobre as decisões estratégicas quanto o uso da IA na organização, porém sem formalidade.	67
	Não há clareza quanto às atribuições e os responsáveis sobre as decisões estratégicas quanto o uso da IA na organização; mas, há decisão formal por estabelecê-las.	33
	Não há clareza quanto às atribuições e os responsáveis sobre as decisões estratégicas quanto o uso da IA na organização; nem decisão formal por estabelecê-las.	0
Práticas para aumentar a transparência do processo de desenvolvimento de sistemas de IA	(A organização tem acesso aos códigos dos sistemas de IA) **E** ((Organização publica código dos algoritmos em formato aberto à sociedade) **OU** **(**Organização utiliza práticas conhecidas da literatura para prover explicabilidade dos sistemas de IA. Ex:XAI.))	100
	Excluindo-se as ações do item acima, a organização tem acesso completo aos códigos dos sistemas de IA.	67
	A organização não tem acesso aos códigos, mas, há decisão formal por ter tal acesso.	33
	A organização não tem acesso aos códigos, nem há decisão formal por tê-lo.	0

Fonte: Elaboração própria

As variáveis-conjunto "processo para governança de IA" e "transparência do processo de desenvolvimento de sistemas de IA" foram coletadas das entrevistas, seguindo os critérios estabelecidos na tabela 6. E, como sistemas de IA podem possuir diferentes níveis de complexidade, propósitos e riscos, podem requerer diferentes ajustes nos processos da organização. Posto isso, visando permitir análises que envolvam as características dos sistemas declarados no

questionário, criou-se um bloco de variáveis dicotômicas que versam sobre existência de acesso do código dos sistemas pelas organizações pesquisadas, contratação externa ou desenvolvimento interno dos sistemas de IA, localização dos dados utilizados pelos sistemas, e tempo de experiência da organização na produção de sistemas de IA (anexo 5 tabela 5).

3.6. Resultados e Discussão

3.6.1 Composição da amostra

A amostra, resultante do caminho percorrido na figura 3, é composta por vinte e oito organizações públicas distribuídas nos três poderes de Estado, e em nove áreas de conhecimento (tabelas 7 e 8). A maior parte dos respondentes do questionário era formada de gestores (71,43%) e profissionais lotados em unidade de TI ou de estatística e ciência de dados, quando estes últimos não estavam dentro da própria TI (tabela 7).

Tabela 7: Características das organizações públicas da amostra utilizada no estudo.

AMOSTRA - 28 ORGANIZAÇÕES PÚBLICAS

País	N	%
Angola	1	3.57
Alemanha	2	7.14
Argentina	3	10.71
Austrália	1	3.57
Brasil	4	14.29
Canadá	2	7.14
Dinamarca	1	3.57
Estônia	2	7.14
Finlândia	2	7.14
Islândia	1	3.57
Itália	1	3.57
Japão	1	3.57
Luxemburgo	1	3.57
Noruega	3	10.71
Reino Unido	1	3.57
Suécia	1	3.57
Suíça	1	3.57

Nº de funcionários	N	%
Até 100	2	7.14
101 a 500	3	10.71
501 a 1000	8	28.57
1001 a 10000	12	42.86
Mais de 10000	3	10.71

Abrangência	N	%
Nacional	24	85.71
Grupo de países	1	3.57
Estadual/Municipal	3	10.71

Esfera de Governo	N	%
Executivo	17	60.714
Legislativo	9	32.143
Judiciário	2	7.14

Área de Atuação	N	%
Assistência Social	3	10.71
Centro de Pesquisa e Dados	3	10.71
Econômica, Financeira e Tributária	7	25.00
Fiscalização/Auditoria	1	3.57
Gestão de Políticas Públicas	1	3.57
Justiça	2	7.14
Parlamento	8	28.57
Saúde	2	7.14
Transporte	1	3.57

Tempo com IA	N	%
Menos de 1 ano	1	3.57
1-3 anos	8	28.57
3-5 anos	11	39.29
Mais de 5 anos	8	28.57

Fonte: Elaboração própria.

Tabela 8: Características dos participantes que responderam aos questionários

AMOSTRA - PARTICIPANTES DA PESQUISA

Unidade onde trabalha	N	%
Tecnologia da Informação	17	60.71
Inovação	2	7.14
Ciência de Dados e Estatística*	3	10.71
Governança Corporativa	4	14.29
Outros	2	7.14

Formação	N	%
Doutorado/Pós-doutorado	9	32.14
Mestrado	12	42.86
Especialização/MBA	4	14.29
Graduação	3	10.71

Posição/Cargo	N	%
Diretor/Gerente/Coordenador de portfolio	20	71.43
Cientista de Dados/Analista de TI	6	21.43
Consultor	2	7.14

* Quando a unidade de ciência de dados não se localiza na área de TI.

Fonte: elaboração própria.

3.6.2 Dados primários utilizados na QCA

O anexo 6 reúne os valores das variáveis-conjunto coletadas e utilizadas na QCA, assim como um resumo estatístico descritivo desses conjuntos.

Os valores coletados para o construto “Fatores Geradores de Expectativas”, “Governança de IA”, “Processos e práticas auxiliares”, “Desenvolvimento e Sustentação de Sistemas de IA” estão disponíveis no anexo 6.

No construto “Fatores Geradores de Expectativas”, foram consideradas apenas leis aprovadas por casas legislativas nacionais (Parlamentos). Não foram encontradas na amostra leis estaduais. Resoluções, decretos e outros normativos estritos a uma determinada esfera de governo foram considerados na variável “Políticas e normativos governamentais sobre IA”, acrescidos de quaisquer outros documentos que estabeleçam regras sobre a IA. Não foram consideradas como leis, os projetos de lei ou atos que ainda estejam tramitando no processo legislativo, e, portanto, ainda em discussão nos parlamentos.

3.6.3 Aplicação da QCA

As análises que envolveram simultaneamente variáveis dicotômicas e variáveis de valores contínuos foram realizadas em tabelas-verdade de valores binários, por meio da prévia conversão das variáveis contínuas em binárias (Betarelli-Júnior & Ferreira 2018) e aplicado o *crisp-se*t QCA.

Para as análises em que todas as variáveis possuíam valores contínuos, foi utilizada *fuzzy* QCA, transformando-se os valores da tabela original em valores *fuzzy*, por meio da calibragem de cada

variável-conjunto, de maneira a ser distribuída no intervalo [0;1], cujos limites representam, respectivamente, ausência completa de membresia, e membresia completa na relação entre os conjuntos (Schneider e Wagemann 2012; Ragin 2008). Os valores extremos considerados para a calibragem, correspondentes ao valor *fuzzy* de 0,05 e de 0,95, foram estabelecidos, respectivamente, como sendo o menor e o maior valor de cada conjunto, como também estabeleceram Navarro et al. 2015 e Codá et al. (2022) em seus estudos.

3.6.3.1 Análise da proposição 1

Para investigação da Proposição 1 aplicou-se a *crisp-set* QCA, com variáveis binárias definidas no anexo 6, considerando como condições os três fatores geradores de expectativas: lei dirigida a regular as questões que oferecem risco ao uso e desenvolvimento inadequado de sistemas de IA (EXP 1), lei dirigida à proteção de dados pessoais (EXP2), e políticas ou normativos governamentais dirigidos às boas práticas no uso e desenvolvimento de sistemas de IA (EXP3).

Para o cálculo do conjunto-resultado da QCA, definiu-se, primeiramente, PGERALGOVIA como a pontuação geral das ações para a governança de IA, considerando as variáveis-conjunto constantes no anexo 6 - tabelas 2, 4 e 5. Desta forma, PGERALGOVIA foi definida como na equação 3: PGERALGOVIA=MÉDIA(FGOVIA;PROCESSOS,PDESENV), onde FGOVIA representa as ações de nível estratégico geralmente utilizadas para formalizar a governança de IA em uma organização; PROCESSOS são os processos auxiliares à implantação da governança da IA; PDESENV representa práticas a serem aplicadas durante o processo de desenvolvimento e sustentação de sistemas de IA, com foco no atendimento de princípios.

A partir do conhecimento do conjunto PGERALGOVIA, com o objetivo na análise com valores dicotômicos, definiu-se o conjunto binário, MAIORGOVIA, contendo valor “1” para os casos em que PGERALGOVIA for maior ou igual a 60,0, e “0” para os casos em que PGERALGOVIA for menor que 60,0. A escolha do valor 60,0 ocorreu para manter coerência com os critérios utilizados nas demais proposições em que se objetiva considerar práticas e processos implantados em qualquer estágio (critérios da tabela 5). Assim, para a análise da proposição 1, MAIORGOVIA presente (igual a 1) significa maiores valores pontuados nas ações para governança de IA. E, MAIORGOVIA ausente (igual a 0) significa não fazer parte dos maiores valores pontuados nas ações para governança de IA.

A distribuição de valores de PGERALGOVIA com seus correspondentes MAIORGOVIA, encontram-se no anexo 6. Destaca-se o fato de que 100% da amostra é composta de organizações sujeitas à alguma lei de proteção de dados pessoais. A partir da interpretação da tabela verdade gerada com MAIORGOVIA (tabela 9), três combinações foram encontradas.

Tabela 9: Tabela verdade utilizada no QCA para a proposição 1

Tabela Verdade - Proposição 1

EXP1	EXP2	EXP3	MAIORGOVIA	ORGANIZAÇÕES
0	1	0	C	1(0), 2(0), 5(1), 10(0), 11(0), 12(0), 15(0), 18(0), 19(1), 20(1), 21(0), 22(1), 27(1)
0	1	1	C	3(1), 4(0), 6(1), 7(0), 8(1), 9(1), 14(0), 16(0), 17(0), 23(0), 24(1), 25(1), 26(0), 28(1)
1	1	1	1	13

Fonte: Valores gerados pelo software Tosmana 1.6.1

A combinação 1 (tabela 10), caracterizada por 27 organizações que não estão sujeitas a qualquer lei para IA, e estão sujeitas a uma lei para proteção de dados pessoais. Desse conjunto, apenas 44,44% apresentaram valores altos de ações dirigidas à governança de IA (5, 19,20,22,27,3,6,8,9,24,25,28). Observa-se, portanto, a existência de contradições, casos que, apesar de estarem na mesma combinação 1, não apresentaram maiores pontuações para ações de implantação da governança de IA). Tal proporção entre casos MARGOVIA=1 e MAIORGOVIA=0 não permite que se estabeleça associação da combinação 1 à alta pontuação nas ações para governança de IA.

Tabela 10: Combinações 1 e 2 da solução para a proposição 1

Combinação 1		Resultado
condições	casos com o resultado investigado	
EXP1{0} * EXP2{1}	5,19,20,22,27,3,6,8,9,24,25,28	MAIORGOVIA(1)
	contradição	
	1,2,10,11,12,15,18,21,4,7,14,16,17,23,26	
Combinação 2		**Resultado**
condições	casos com o resultado investigado	
EXP2{1} * EXP3{1}	3,6,8,9,24,25,28,13	MAIORGOVIA(1)
	contradição	
	4,7,14,16,17,23,26	

Fonte: Adaptação a partir das informações geradas pelo software Tosmana 1.6.1.

Compondo a análise, a combinação 2 (tabela 10), caracteriza-se por 15 organizações que estão sujeitas à lei para proteção de dados pessoais e estão sujeitas a políticas ou recomendações de governo versando sobre desenvolvimento e uso de sistemas de IA. Apenas 53% dos casos desse conjunto (3,6,8,9,24,25,28,13) apresentaram alta pontuação nas ações para implantação da

governança de IA. A análise da tabela verdade também nos revela uma terceira combinação em que apenas uma organização com todos os fatores geradores de expectativas presentes (caso 13): lei para IA (EXP1), lei para proteção de dados pessoais (EXP2) e políticas ou normas de governo que versem sobre uso e desenvolvimento de sistemas de IA (EXP3). Trata-se do maior valor de pontuação de toda a amostra quanto a ações voltadas à governança de IA (94,87).

Apesar da existência de vários projetos de lei tramitando em casas legislativas (OECD 2022b), nos dezessete países contemplados pelos vinte e oito casos da amostra, somente na Dinamarca (Danish Government 2020) identificou-se lei aprovada versando sobre ética em dados, dirigida tanto a uso e desenvolvimento de sistemas de IA, quanto a qualquer análise de dados em organizações de grande porte. Convém a reflexão de que, versões de leis para proteção de dados pessoais foram criadas em vários países logo após a EU GDPR (2016) ter sido sancionada na Comissão Europeia. E, portanto, situam-se em momento diferente das leis sobre o uso e a desenvolvimento de sistemas de IA (OECD 2022b; European Commission 2021a), apesar dos avanços na tramitação das propostas de emendas ao ato europeu para IA (European Parliament 2022a; European Parliament 2022b), referencial tanto para a Europa, quanto para outros países.

Quanto à análise da combinação 2, 53,57% da amostra declarou existir algum documento do governo com direcionamento ou recomendações sobre boas práticas no uso e desenvolvimento da IA (EXP2). Observa-se que a contribuição dada pelas *soft laws* na forma de políticas e normativos de governo voltados ao uso e desenvolvimento de sistemas de IA (EXP3), associada à existência de lei para proteção de dados pessoais, apresentaram um grau maior de efetividade para alta pontuação das ações voltadas à governança de IA em relação à combinação 1. Contudo, a proporção ainda não é suficiente para se afirmar alguma associação entre a combinação 2 e alta pontuação para ações voltadas à governança de IA. Uma reflexão sobre provável razão para esse cenário é de que ainda não tenha havido tempo suficiente para que as recomendações sejam colocadas em prática, dado que a maior parte dos normativos governamentais é muito recente.

3.6.3.2 Análise da proposição 2

Para análise da proposição 2, considerou-se agrupamentos de processo e práticas nos temas “dados”, “riscos”, “segurança” e desenvolvimento”, como condições que podem ser combinadas. Para o conjunto resultado, consideram-se os casos que estejam em um estágio mais avançado de implantação das ações estratégicas para a governança de IA. Na aplicação da *fuzzy* QCA, para

representar as práticas da dimensão "dados" (PDADOS), foram considerados os processos de governança de dados, de gestão da qualidade de dados e gestão da proteção de dados pessoais. Para as práticas na dimensão "riscos" (PRISCOS), foram considerados o processo de gestão de riscos de sistemas de IA, o processo de auditoria em sistemas de IA, a prática de identificação dos *stakeholders* em todo o ciclo de vida dos sistemas de IA, e a monitoração de mudanças no ambiente. Para as práticas na dimensão "segurança" (PSEG) foi utilizado somente o processo de gestão da segurança de sistemas de IA. E na dimensão "desenvolvimento" (PDESENV), foram consideradas práticas das fases de projeto e de sustentação dos sistemas de IA dirigidas ao atendimento de princípios éticos. Para representar as ações estratégicas que formalizam a existência de uma governança de IA (FGOVIA), foram considerados: estratégia para a IA (GOV1); política ou normativo dirigido a sistemas de IA(GOV2); código ou guia de princípios éticos que sejam aplicados aos sistemas de IA(GOV3); processo para governança de IA (GOV4); e a existência de uma estrutura, pessoa responsável ou comitê para tratar a governança de IA (GOV5). A composição de cada dimensão é apresentada nas equações 4, 5, 6 e 7, a partir das variáveis-conjunto coletadas.

PDADOS = MÉDIA(PROGOVD,PROQUAD,PRODPES) Equação 4;
PRISCOS = MÉDIA(PRORISC, AUDIT, STAKEH, MAMBIENTE) Equação 5;
PDESENV = MÉDIA(SISPROJETO,SISOPERACAO) Equação 6;
FGOVIA = MÉDIA(MÁXIMO(GOV1,GOV2),GOV3,MÁXIMO(GOV4,GOV5)) Equação 7.

A calibragem para a passagem dos valores originais para os valores *fuzzy* correspondentes foi realizada considerando-se que interessam à presente análise as condições cujas práticas tenham sido ou estejam sendo implantadas em qualquer proporção, ou seja, valores ≥ 67. Para tais propósitos, em todas as análises da pesquisa que utilizam *fuzzy* QCA, definiu-se o ponto de cruzamento=60,0 tanto para as variáveis-conjunto condição, quanto para a variável-conjunto resultado (tabela 11). No caso da proposição 2 e das proposições derivadas dela, convencionou-se como altos valores de FGOVIA aqueles maiores que 60,0.

A partir da tabela verdade da proposição 2, as soluções, complexa, parcimoniosa e intermediária foram calculadas, tendo-se optado pela solução complexa para todas as análises *fuzzy* QCA, por apresentar menor grau de simplificação de suas equações, portanto mais útil à descrição dos casos da pesquisa. Distribuída em três combinações, a solução complexa é apresentada na tabela 11.

As ações da dimensão “segurança” e “riscos” tiveram menor adesão e implantação em menor proporção; enquanto as ações da dimensão “dados” e “desenvolvimento”, um maior avanço na implantação da governança de IA.

A combinação 1 (tabela 11) é caracterizada por baixa pontuação nas práticas na dimensão “riscos”, por alta pontuação nas práticas nas dimensões “segurança” e “desenvolvimento”. Das cinco organizações que optaram pela combinação 1, quatro (25,5,6,26), 80%, apresentaram alta pontuação em ações estratégicas que formalizam a governança de IA; e, em situação contraditória, um caso (20) não obteve alta pontuação em suas ações estratégicas para governança de IA. Tal cenário sugere que organizações situadas na combinação 1 já identificaram as peculiaridades dos riscos em segurança de sistemas de IA, o que as alertam para prevenção durante o próprio desenvolvimento de tais sistemas, permitindo que se utilize o mesmo corpo funcional atuando nos dois processos.

Tabela 11: Parâmetros e valores gerados da QCA para análise da proposição 2.

Proposição 2 - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Vmínimo	Vmáximo	Limiares para Calibragem fsQCA		
					Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	PDADOS	71.12	22	100	100	60	22
	PRISCOS	47.38	0	100	100	60	0
	PSEG	46.5	0	100	100	60	0
	PDESENV	60.23	0	100	100	60	0
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2 - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.797232	0.825606	0.811073
Consistência da solução	0.890263	0.881093	0.88253
Corte de frequência	1	1	1
Corte de consistência	0.914783	0.914783	0.914783

Proposição 2 - Tavela Verdade

PDADOS1	PRISCOS1	PSEG1	PDESENV1	FGOVIA1	FREQ
1	1	1	0	1	1
0	0	1	1	1	1
1	1	1	1	1	6
1	0	1	1	1	4
1	1	0	1	1	2
0	0	0	1	0	1
1	0	0	1	0	1
1	0	1	0	0	3
0	1	0	1	0	1
1	0	0	0	0	3
0	0	0	0	0	5

Combinação 1

condições	casos com o resultado investigado	Resultado
~PRISCOS *PSEG *PDESENV	25,5,6,26	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.408997	0.12872	0.929245	20	

Combinação 2

condições	casos com o resultado investigado	Resultado
PDADOS *PRISCOS *PSEG1	22,9,13,8,3,19	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.548789	0.0228372	0.950839	28	

Combinação 3

condições	casos com o resultado investigado	Resultado
PDADOS *PRISCOS *PDESENV	3,9,22,13,27,8	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.645675	0.119723	0.914706	28	

Fonte: Adaptação a partir das informações geradas pelo software FSQCA versão 3.1.

A combinação 2 (tabela 11) é caracterizada por alta pontuação nas ações da dimensão “dados”, nas ações da dimensão “riscos” e nas ações na dimensão “segurança”. Das organizações que

optaram pela combinação 2, 85,71% (22,9,13,8,3,19) apresentaram alta pontuação em suas ações estratégicas para governança de IA Nessa combinação, maior ênfase ou priorização foi dada à implantação de processos de dados, nas práticas e processos que mitigam riscos nos sistemas de IA e no processo de gestão de segurança dos sistemas de IA. Tal fenômeno pode decorrer de vários fatores, como escolhas das organizações por contratar ou estabelecer parcerias nas quais elas não sejam protagonistas do desenvolvimento, mas, fornecedoras de dados de qualidade e de ambiente seguro. Um segundo possível fator é o simples desconhecimento das práticas próprias do processo de desenvolvimento dirigidas a questões éticas. E uma terceiriza razão pode ser o fato de terem considerado seus sistemas de baixo risco, não exigindo esforços adicionais dirigidos a questões éticas durante o desenvolvimento dos sistemas de IA.

A combinação 3 (tabela 11) é caracterizada por alta pontuação nas ações das dimensões "dados", "riscos" e "desenvolvimento". Observa-se, assim, a priorização nos processos de gestão de dados, nas práticas dirigidas à mitigação de riscos de sistemas de IA, e nas práticas distribuídas no processo de desenvolvimento e sustentação de sistemas de IA com foco em atender a princípios éticos. Das organizações que optaram pela combinação 2, 85,71% (3,9,22,13,27,8) apresentaram alta pontuação em suas ações estratégicas que formalizam a governança de IA. Na combinação 3, percebe-se provável envolvimento do nível estratégico, tático e operacional, pelo fato da dimensão "dados" envolver a governança de dados, geralmente protagonizada por um comitê de gestores de dados em nível estratégico das unidades administrativas (Vilminko-Heikkinen & Pekkola 2019, Abraham 2018), e o processo de desenvolvimento de sistemas de IA atingir o nível operacional.

Um quarto grupo de casos pode ser observado, como a interseção entre as combinações 2 e 3. As organizações 3,8,9,13,22,28 implantaram, em qualquer estágio, práticas nas quatro dimensões: "dados", "riscos", "segurança" e "desenvolvimento" (PDADOS1*PRISCOS*PSEG1*PDESENV1). Faz-se importante o registro de que todas apresentaram pontuação alta para ações estratégicas que formalizam a governança de IA, exceto o caso 28, que se apresenta como uma contradição em todas as combinações.

Portanto, para a proposição 2, as combinações 2 e 3 apresentaram maior valor da proporção entre alta e baixa pontuação das ações estratégicas para a governança de IA (85,71%). Destaca-se a combinação 3, por contemplar as ações da dimensão "dados" maior média do grupo; sendo também a combinação, potencialmente, com a maior diversidade de *stakeholders* – níveis estratégico, tático e operacional.

3.6.3.2.1 Análise da Proposição 2A

A partir da decomposição das práticas previstas na dimensão “dados” (PDADOS) em processos de governança de dados (PROGOVD), de gestão da qualidade de dados (PROQUAD) e gestão da proteção de dados pessoais (PRODPES), procedeu-se a análise da proposição 2A, repetindo a comparação com os casos que estavam em um estágio mais avançado de implantação das ações estratégicas para a governança de IA.

Tabela 12: Parâmetros e valores gerados da QCA para análise da proposição 2A.

Proposição 2A - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Mínimo	Máximo	Limiares para Calibragem: Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	PROGOVD	64.39	0	100	100	60	0
	PROQUAD	58.46	0	100	100	60	0
	PRODPES	90.5	33	100	100	60	33
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2A - Tavela Verdade

PROGOVD1	PROQUAD1	PRODESP1	FGOVIA1	FREQ
0	1	1	1	1
1	0	1	1	2
1	1	1	1	19
0	0	0	1	2
0	0	1	0	4

Proposição 2A - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.893426	0.893426	0.893426
Consistência da solução	0.657332	0.646794	0.646794
Corte de frequência	1	1	1
Corte de consistência	0.75	0.75	0.75

Combinação 1

condições			casos com o resultado investigado	Resultado
PROGOVD1 * PRODPES1			3,5,9,19,22,27,6,8,13,25	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradições	
0.825606	0.103114	0.71781	4,20,24,2,7,11,15,16,23	

Combinação 2

condições			casos com o resultado investigado	Resultado
PROQUAD1 * PRODPES1			8,19,22,25,5,6,9,13,27	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.733564	0.0110726	0.688312	20,23,2,4,7,21,15,16,18,28	

Combinação 3

condições			casos com o resultado investigado	Resultado
~PROGOVD1 * ~PROQUAD1 * ~PRODPES1			21	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.215917	0.0567474	0.75	1	

Teste de Necessidade

condições	consistência	cobertura	Resultado
PROGOVD1 * PROQUAD1 * PRODPES1	0.959862	0.581795	FGOVIA1

Fonte: Adaptação das informações geradas pelo software fsQCA 3.1.

O processo de gestão de proteção de dados pessoais apresentou a maior média (90,5), o que reforça a interpretação de estágio mais avançado na implantação das ações internas realizadas pelas organizações da amostra para atender à lei de proteção de dados pessoais a que estão sujeitas, como discutido na proposição 1. Adicionalmente, entende-se que o reflexo dessa média se estende à média da dimensão “dados”, na análise da proposição 2 . Para o cálculo da tabela verdade, o máximo corte de consistência possível foi 0,75, o que é aceito por Ragin (2008) e por Schneider & Wagemann (2012). Fundamentada na tabela verdade, a solução complexa apresentou três combinações (tabela 12).

A baixa membresia das combinações 1 e 2, abaixo do valor aceito por Ragin (2008) e Schneider & Wagemann (2012) não permite que se reconheça associação entre cada uma dessas combinações e conjunto-resultado (FGOVIA1). Similarmente, a combinação 3 não agrega valor à pesquisa visto se caracterizar pela ausência de implantação dos três processos analisados, e mesmo número de casos que apresentaram alta e baixa pontuação das suas ações estratégicas para governança de IA.

Contudo, convém a observação de que na tabela verdade, há dezenove casos em que houve implantação dos três processos – governança de ados, gestão da qualidade de dados e gestão da proteção de dados pessoais (PROGOVD1*PROQUAD1*PRODPES1); e que das doze maiores pontuações de ações estratégicas para a governança de IA, nove casos (75%) haviam implantado, em qualquer estágio, os três processos dirigidos a dados. Tais casos materializam a interseção das combinações 1 e 2, fato compreensível, haja vista nenhuma delas fazer menção à ausência de outro processo. O teste de avaliação de condição necessária, disponível no FSQCA 3.1 permitiu que se conhecesse mais sobre a combinação PROGOVD1*PROQUAD1*PRODPES1 (tabela 12), haja vista os resultados indicarem ser pouco provável de se encontrar altos valores nas ações estratégicas para a governança de IA (FGOVIA1) sem a existência simultânea dos três processos testados.

O impacto positivo quando os três processos sobre dados são implantados permite a reflexão de que a implantação da governança de dados requer definição de uma estratégia de dados, definição dos gestores e curadores desses dados e, de forma sistemática, dar transparência das decisões sobre o que deve e pode ser feito com os dados, forma de acesso, parâmetros para descarte (Sivarajah 2017). São ações com necessidade de envolvimento de gestores de todos os níveis hierárquicos com autoridade para tais deliberações (Benfeldt, 2020; Vilminko-Heikkinen, 2020).

Em um nível tático, a busca por qualidade dos dados de uma organização requer o estabelecimento de regras claras para a entrada dos dados (Rhahla et al. 2021) e seu tratamento nos sistemas de cada processo de negócio (Haneem et al. 2019; Vilminko-Heikkinen & Pekkola 2019, Benfeldt et al. 2020). Qualquer ação em torno de padronizações, ou regras para os dados, requer autorização dos gestores desses dados das áreas de negócio, ação pertencente à governança de dados. E tais regras precisam estar registradas em um catálogo corporativo de dados (Labadie et al. 2020).

Também de nível tático, o trabalho de gestão de proteção de dados pessoais envolve ações que dependem de uma boa gestão de dados (Rhahla 2021), o que também requer um catálogo atualizado das bases de dados pessoais (Labadie et al. 2020). Assim, a plena implantação de gestão de dados pessoais, também envolve ações de governança de dados, e de qualidade de dados estrita aos dados pessoais. A combinação dessas condições revela implantação abrangente das recomendações por gestão da proteção de dados pessoais constantes em legislação discutida sobre o tema, fato que corrobora com a análise feita na proposição 1 (100% da amostra sujeitas à alguma lei sobre proteção de dados pessoais).

Pelo exposto, a proposição 2A confirma uma associação entre a combinação da implantação dos processos de governança de dados, gestão da qualidade de dados, gestão da proteção de dados pessoais com organizações em estágio avançado nas ações estratégicas para governança de IA.

3.6.3.2.2 Análise da Proposição 2B

A partir da decomposição das práticas previstas na dimensão "riscos" (PRISCOS) em processo de gestão de riscos (PRORISC), processo de auditoria em sistemas de IA (AUDIT), prática para identificação de *stakeholders* (STAKEH), monitoração de mudanças do ambiente e tendências sociais (MAMBIENTE) como representado na equação 5, procedeu-se a análise da proposição 2B, repetindo a comparação com os casos que estavam em um estágio mais avançado de implantação das ações estratégicas para a governança de IA. Após calibragem (tabela 13), fundamentadas na tabela verdade, as três soluções (complexa, parcimoniosa e intermediária) apresentaram alta consistência. Optou-se pela solução complexa que apresentou três combinações.

A combinação 1 (tabela 13) contemplou quatro casos em que se implantou, em qualquer estágio, práticas para identificação dos *stakeholders* dos sistemas de IA, e práticas para a monitoração de mudanças no ambiente com identificação de tendências sociais; mas, não se implantou qualquer prática de auditoria em sistemas de IA. O número de casos que obteve alta pontuação nas ações estratégias para governança de IA foi igual aos casos com baixa pontuação. Apesar da alta consistência da combinação, a proporção entre casos com alta e baixa pontuação nas ações estratégicas para governança de IA impede que se considere a combinação 1 como uma associação válida para o presente estudo.

Foi identificada alta pontuação nas ações estratégias para a governança de IA em 85,71% dos casos apresentados pela combinação 2 (3,9,13,22,27,8), os quais implantaram, em qualquer estágio, processo de gestão de riscos, práticas para identificar *stakeholders* em todo o ciclo de vida de sistemas de IA, e práticas de monitoração de mudanças de ambiente e tendências sociais . Uma vez definidos claramente todos os *stakeholders* de um sistema de IA, um processo de gestão de riscos se alimenta de tal informação para nutrir análises de impacto que vão acompanhar a vida de tais sistemas (Wirtz et al 2022; NIST 2022). De maneira similar, o processo de gestão de riscos pode consumir informações sobre mudanças nas regras de negócio, em normativos ou em percepções da sociedade sobre algum tema diretamente abordado pelo sistema de IA (Zicari e*t al*. 2021). Sendo assim, este último deve gerar informações suficientes e precisas para viabilizar um confiável

processo de gestão de riscos (Breier et al 2020; Vetró et al 2021). Logo, as ações para mitigação de riscos em sistemas de IA são enriquecidas, quando a correta e ampla identificação de *stakeholders* se une a um processo de gestão de riscos que se alimenta de uma monitoração constante de mudanças em variáveis diversas de ambiente e aspectos sociais, de maneira a permitir identificar se o contexto em que o sistema se encontra é diferente daquele para o qual foi projetado.

Tabela 13: Parâmetros e valores gerados da QCA para análise da proposição 2B.

Proposição 2B - Calibragem para aplicação da fuzzy QCA

	Variáveis-conjunto	Média	Mínimo	Máximo	Limiares - Calibragem: Adesão plena	Limiares - Calibragem: Ponto de Cruzamento	Limiares - Calibragem: Não-adesão plena
Condição	PRORISC	34.5	0	100	100	60	0
	AUDIT	45.32	0	100	100	60	0
	STAKEH	56.04	0	100	100	60	0
	MAMBIENTE	53.68	0	100	100	60	0
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2B - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.690657	0.731488	0.690657
Consistência da solução	0.900722	0.889731	0.8879
Corte de frequência	1	1	1
Corte de consistência	0.913043	0.913043	0.913043

Proposição 2B - Tavela Verdade

PRORISC1	AUDIT1	STAKEH1	MAMBIENTE	FGOVIA1	FREQ
1	1	1	1	1	5
1	0	1	1	1	2
0	1	1	0	1	3
0	0	1	1	1	2
0	1	0	1	0	1
0	0	1	0	0	2
0	0	0	1	0	2
1	1	0	1	0	2
0	1	1	1	0	4
0	0	0	0	0	5

Combinação 1				Resultado
condições			casos com o resultado investigado	
~AUDIT1 * STAKEH1 * MAMBIENTE1			27,25	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.319031	0.0581315	0.893411	2,16	

Combinação 2				Resultado
condições			casos com o resultado investigado	
PRORISC1 * STAKEH1 * MAMBIENTE1			3,9,13,22,27,8	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.49135	0.23045	0.941645	2	

Combinação 3				Resultado
condições			casos com o resultado investigado	
PRORISC1 * AUDIT1 * STAKEH1 * ~MAMBIENTE			5,21	FGOVIA1
cobertura total	cobertura única	consistência	casos em contradição	
0.293426	0.107958	0.925764	20	

Fonte: Adaptação a partir dos valores calculados pelo software fsQCA 3.1

A combinação 3 (tabela 13) apresenta dois casos (5,21) em que se implantou, em qualquer proporção, práticas para identificação dos *stakeholders* dos sistemas de IA, e se implantou, em qualquer proporção, práticas para auditoria em sistemas de IA; mas, não se implantou processo de gestão de riscos, nem práticas que façam a monitoração de mudanças no ambiente e se identifiquem tendências sociais. Houve uma contradição (20) em que a mesma combinação fez parte de um cenário que não atingiu alta pontuação das ações estratégicas para governança de IA. Assim, apesar da alta consistência, a proporção entre casos com alta e baixa pontuação nas ações estratégicas para governança de IA impede que se considere como uma associação válida para o presente estudo.

Portanto, por meio da análise da proposição 2B identificou-se que a combinação composta por um processo para gestão de riscos, prática para definição de *stakeholders*, e monitoração de mudanças no ambiente apresentou associação com estágios mais avançados da implantação de ações estratégicas para a governança de IA. Tal constatação revela que a percepção dos gestores estratégicos das organizações contempladas com essa combinação coaduna com o pensamento de Wirtz et al. (2022), NIST (2022), González et al. (2020) e Raji et al. (2020), quanto à contribuição positiva à governança de IA quando existe processo de gestão de riscos alimentado por uma precisa e correta definição de *stakeholders* em todo o ciclo de vida dos sistemas de IA, e também consome as informações fornecidas por uma monitoração de mudanças no ambiente.

3.6.3.2.3 Análise da Proposição 2C

Composta apenas pelo processo de gestão de segurança (PROSEG), a dimensão "segurança" teve sua análise simplificada. A calibragem da QCA foi realizada com os mesmos parâmetros das demais *fuzzy* QCA, e a tabela verdade fundamentou os índices de análise, tendo-se escolhido a solução complexa (tabela 14).

A análise da combinação única (tabela 14) revela que, dos casos que haviam implantado, em qualquer estágio, processo para gestão de segurança em sistemas de IA, 66,67% obtiveram alta pontuação para ações estratégicas para a governança de IA. Apesar da consistência aceitável, a proporção de casos com 3. presente sugere uma baixa associação entre o processo de gestão de segurança para sistemas de IA e estágios avançados da implantação de ações estratégicas para governança de IA.

Tabela 14: Parâmetros e valores gerados da QCA para análise da proposição 2C.

Proposição 2C - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Mínimo	Máximo	Limiares para Calibragem: Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	PROSEG1	46.5	0	100	100	60	0
Resultado	FGOVIA1	56.02	0	100	100	60	0

Combinação 1

condições	casos com o resultado investigado	Resultado
PROSEG1	5,8,9,13,22,3,6,19,25,26	FGOVIA1

cobertura total	cobertura única	consistência	casos em contradição
0.719031	0.719031	0.8312	11,2,16,20,28

Proposição 2C - Tavela Verdade

PROSEG1	FGOVIA1	FREQ
1	1	15
0	0	13

Proposição 2C - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.719031	0.719031	0.719031
Consistência da solução	0.8312	0.8312	0.8312
Corte de frequência	13	13	13
Corte de consistência	0.8312	0.8312	0.8312

Fonte: Adaptação a partir dos valores resultantes dos cálculos realizados pelo software fsQCA 3.1

3.6.3.2.4 Análise da Proposição 2D

A análise das práticas da dimensão "desenvolvimento" requereu dois níveis de decomposição: primeiramente com as fases do projeto de gestão do desenvolvimento e sustentação de sistemas de IA – projeto (SISPROJETO) e operação (SISOPERACAO); e, posteriormente, decompondo cada uma dessas fases. Após calibragem e geração da tabela verdade, as soluções complexa, parcimoniosa e intermediária apresentaram mesmas consistências tendo-se escolhido a solução complexa que apresentou única combinação (tabela 15)

Tabela 15: Parâmetros e valores gerados da QCA para análise da proposição 2D.

Proposição 2D - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Mínimo	Máximo	Limiares para Calibragem: Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	SISPROJETO	54.07	0	100	100	60	0
	SISOPERACAO	66.38	0	100	100	60	0
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2D - Tavela Verdade

SISPROJETO	SISOPERACAO	FGOVIA1	FREQ
1	0	1	4
1	1	1	8
0	1	0	7
0	0	0	9

Proposição 2D - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.818685	0.818685	0.818685
Consistência da solução	0.864766	0.864766	0.864766
Corte de frequência	4	4	4
Corte de consistência	0.878943	0.878943	0.878943

Combinação 1				Resultado
condições			casos com o resultado investigado	
SISPROJETO1			9,13,27,14,3,8,26,5,21,22	FGOVIA1
cobertura tota	cobertura única	consistência	casos em contradição	
0.818685	0.818685	0.864766	24,23	

Fonte: Adaptação a partir dos valores resultantes dos cálculos realizados pelo software fsQCA 3.1

Apesar da média de pontuação das práticas aplicadas na fase de operação (66,38) ter sido superior à média de pontuação das práticas aplicadas na fase de projeto (54,07), a *fuzzy* QCA apresentou única combinação com somente a condição SISPROJETO1 para as soluções complexa, parcimoniosa e intermediária, sugerindo que práticas da fase de projeto dirigidas às questões éticas dos sistemas de IA são necessárias e suficientes para que se atribua sua associação à alta pontuação das ações estratégicas para governança de IA.

Portanto, identificou-se associação entre estar mais avançado na implantação das ações estratégicas para a governança de IA e a implantação das práticas da fase de projeto do processo de desenvolvimento e sustentação de sistemas de IA dirigidas aos princípios éticos. Entre possíveis razões para este cenário, destaca-se o fato de que projetos são aprovados por gestores de nível estratégico que se empenham no patrocínio do projeto até o seu lançamento como um serviço digital; a partir de então, tais sistemas, já em operação, passam a ser foco apenas das equipes técnicas. Considerando as médias de cada fase, as práticas da fase de operação aparentemente estão

em estágios mais avançados, em relação às ações da fase de projeto. Entre as possíveis razões, podem ser consideradas: atividades de monitoração e coleta de *feedback* são menos complexas e por isso mais facilmente implantadas e compartilhadas por profissionais das unidades de negócio.

3.6.3.2.4.1 Análise da Proposição 2D1

A *fuzzy* QCA das práticas da fase de projeto considerou os mesmos parâmetros das demais análises; porém, com um desmembramento da prática de representação de regras e dilemas éticos, visto que a identificação de todos os *stakeholders* durante o ciclo de vida do sistema de IA ser necessária a tal atividade. Ademais, a identificação dos *stakeholders* é uma atividade que, apesar de importante, tem complexidade menor que a atividade de representação formal dos princípios e dilemas éticos; sendo esta última, se efetuada com profundidade, requerente de conhecimentos filosóficos sobre ética, sobre comportamentos sociais e equações que representem não somente quais são, mas, como o sistema vai lidar com cada cenário (Bonnemains et al. 2018; Anderson & Anderson 2018; Bench-Capon & Modgil 2017). Assim, para efeito da pontuação geral ações para implantação da governança de IA (proposições 1, 3 e 4), agrupou-se essas duas práticas como RDILEMA, seguindo a equação 8: RDILEMA=0,2*STAKEH+0,8*DILEMA .

Pelo exposto, a QCA foi aplicada às variáveis-conjunto das seguintes condições: representação de regras e dilemas éticos (RDILEMA), identificação de *stakeholders* (STAKEH), práticas para minimizar vieses (PVIESES), e práticas para prover transparência a todo o processo produtivo de sistemas de IA (PTRANSP).

A combinação 1 da solução complexa caracteriza-se por organizações que implantaram em qualquer estágio, práticas para representar formalmente regras de negócio com os princípios éticos e/ou dilemas éticos, e práticas para prover transparência na produção dos sistemas de IA. 75% dos casos que fizeram a opção desta combinação (9,13,27,3,21,26), apresentaram alta pontuação para ações estratégicas para a governança da IA; havendo contradição em 25% dos casos.

Destaque-se que a combinação 1 (tabela 16) não exclui a possibilidade de inclusão de casos com as práticas para minimização de vieses (PVIESES1); pois, não houve a expressa negação dessa condição. Tal situação justifica o fato de 100% dos casos da combinação 1 com presença de FGOVIA1 (alta pontuação nas ações de nível estratégico para a governança de IA), terem sido casos que também implantaram as três práticas (RDILEMA1*PTRANSP1* PVIESES1). A aplicação do teste de condições necessárias do fsQCA para a combinação RDILEMA1*PTRANSP1 e RDILEMA1*PTRANSP1*PVIESES1 (tabela 16) resultou, como

esperado, em consistência maior para a combinação das três condições, visto que {RDILEMA1*PTRANSP1*PVIESES1}⊂ {RDILEMA1*PTRANSP}.

Tabela 16: Parâmetros e valores gerados da QCA para análise da proposição 2D1.

Proposição 2D1 - Calibragem para aplicação da fuzzy QCA

	Variáveis-conjunto	Média	Mínimo	Máximo	Limiares para Calibragem: Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	RDILEMA	51.29	0	100	100	60	0
	PVIESES	60.82	0	100	100	60	0
	PTRANSP	50.1	0	100	100	60	0
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2D1 - Tavela Verdade

RDILEMA1	PVIESES1	PTRANSP1	FGOVIA1	FREQ
1	1	1	1	7
1	0	1	1	1
0	1	0	0	2
1	1	0	0	5
0	1	1	0	6
0	0	1	0	3
0	0	0	0	4

Proposição 2D1 - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.538408	0.538408	0.538408
Consistência da solução	0.913146	0.913146	0.913146
Corte de frequência	1	1	1
Corte de consistência	0.935294	0.935294	0.935294

Combinação 1 — Resultado: FGOVIA1

condições	casos com o resultado investigado
RDILEMA1*PTRANSP1	9,13,27,3,21,26

cobertura total	cobertura única	consistência	casos em contradição
0.538408	0.538408	0.913146	15,24

Teste de Necessidade — Resultado: FGOVIA1

condições	consistência	cobertura
RDILEMA1*PTRANSP1	0.929412	0.727125
RDILEMA1*PTRANSP1* PVIESES1	0.96263	0.679531

Fonte: Adaptação a partir dos valores gerados pelo software fsQCA 3.1.

A baixa média das práticas voltadas à transparência confirma desafio de se obter explicação dos resultados dos algoritmos previsto por Tutt (2017), Butterworth (2018), Buiten (2019) e Zuiderwijk et al. (2021). Entre outros fatores, tal cenário pode ter sido amplificado pelo fato de 73,33% da amostra ter contratado o desenvolvimento de, pelo menos, uma parte dos seus sistemas de IA, e apenas 46,43% da amostra têm acesso a 100% dos códigos dos seus sistemas de IA declarados na pesquisa.

Quanto aos baixos valores para a representação princípios e dilemas éticos, podem ser decorrentes da falta de definição clara dos princípios éticos, da falta de percepção de que possam existir dilemas éticos quando as regras de negócio são transferidas para os sistemas de IA, ou a inexistência de profissionais com conhecimento necessário para implantar a prática, como alertam Ahn e Chen (2022).

O esforço para implantação de prática para minimização de vieses alinha-se ao pensamento de Strauß (2021) quando destaca que, sem o foco no alcance dos efeitos que um sistema de IA possa gerar, não existe confiabilidade numa gestão de riscos. O fato da média das práticas para minimizar vieses ter sido a mais elevada deste grupo revela que, onde foi aplicado, estavam em estágio mais

avançado de implantação. Uma razão para as médias mais altas pode residir no fato de que durante treinamentos para desenvolvimento de sistemas de IA, geralmente, práticas para minimizar vieses de dados por meios de técnicas estatísticas são apresentadas. No entanto, apenas tais técnicas podem não ser suficientes para cobrir a variedade de tipos de vieses, a depender do tipo de dados, do tipo de sistema e do contexto em que este se insere, como destacam Lin et al. (2021), Strauß (2021) e Leavy et al. (2020).

A QCA aplicada para a proposição 2D1 revelou que as organizações que implantaram práticas para representação formal de regras e dilemas éticos, práticas para prover transparência e práticas para minimizar vieses nos sistemas de IA durante o projeto desses sistemas, também apresentaram estágio mais avançado na implantação de ações de nível estratégico dirigidas à governança de IA, confirmando as argumentações de Anderson e Anderson (2018), Bonnemains et al. (2018) e Oppy e Dowe (2011) na defesa de uso de representação formal dos princípios e dilemas éticos; de Dazeley et al. (2021), Adadi e Berrada (2018) e Phillips et al. (2021) na defesa de práticas para a transparência do processo de desenvolvimento; assim como Ashokan e Haas (2021), Makhlouf et al. (2021), González et al. (2020) na defesa da implantação de práticas para minimização de vieses.

3.6.3.2.4.2 Análise da Proposição 2D2

A análise das práticas da fase de operação do processo de desenvolvimento e sustentação de sistemas de IA, variáveis-conjunto de condições monitoração automática (MAUTO), supervisão humana (SHUMANA) e coleta de *feedback* dos usuários (FEEDBACK). Identificou-se a maior média entre as práticas desta fase na monitoração automática (71,57) revelando a maior facilidade de se monitorar quando não se depende de recursos humanos.

Após calibragem e geração da tabela verdade, as soluções complexa, parcimoniosa e intermediária apresentaram valores de consistência idênticos (tabela 17), tendo a solução complexa apresentado única combinação, cuja consistência é superior a 0,75; mínimo aceito por Ragin (2008) e, Schneider e Wagemann (2012).

A combinação 1, composta por prática de supervisão humana e prática de coleta de *feedback*, foi opção de dezesseis organizações da amostra, das quais onze (3,6,9,13,22,25,5,14,19,26,27), 68,75%, obtiveram alta pontuação nas ações de nível estratégico para a governança de IA.

A análise mais precisa da proposição 2D2 impõe uma leitura da tabela verdade (tabela 17) onde se constata que dos dezesseis casos da combinação 1, quinze (93,75%), confirmando as percepções de Rahwan et al. (2019), Wright e Schultz (2018), Dignum (2019), e De Silva e Alahakoon (2022); implantaram em alguma proporção, as três práticas, monitoração automática, supervisão humana e coleta de *feedback*, correspondendo a um subconjunto dos casos da combinação 1. E de fato, {MAUTO*SHUMANA*FEEDBACK}⊂ {SHUMANA*FEEDBACK}, além de que a combinação 1 não exclui a possibilidade de inclusão de outro conjunto.

Tabela 17: Parâmetros e valores gerados da QCA para análise da proposição 2D2.

Proposição 2D2 - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Mínimo	Máximo	Limiares para Calibragem: Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	MAUTO	71.57	0	100	100	60	0
	SHUMANA	60.79	0	100	100	60	0
	FEEDBACK	66.79	0	100	100	60	0
Resultado	FGOVIA	56.02	0	100	100	60	0

Proposição 2D2 - Tavela Verdade

MAUTO	SHUMANA	FEEDBACK	FGOVIA1	FREQ
0	1	1	1	1
1	1	1	1	15
1	1	0	0	3
1	0	0	0	1
0	0	0	0	2
1	0	1	0	6

Proposição 2D2 - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.741869	0.741869	0.741869
Consistência da solução	0.790561	0.790561	0.790561
Corte de frequência	1	1	1
Corte de consistência	0.794207	0.794207	0.794207

Combinação 1

condições	casos com o resultado investigado	Resultado
SHUMANA1 * FEEDBACK1	3,6,9,13,22,25,5,14,19,26,27	FGOVIA1

cobertura total	cobertura única	consistência	casos em contradição
0.741869	0.741869	0.790561	10,24,16,20,28

Teste de Necessidade

condições	consistência	cobertura	Resultado
SHUMANA1 * FEEDBACK1	0.891349	0.627069	FGOVIA1
MAUTO1 * SHUMANA1 * FEEDBACK1	0.972318	0.576529	

Fonte: Adaptação a partir dos valores gerados pelo software fsQCA 3.1.

A aplicação do teste de necessidade das condições para a combinação 1 e para a seu subconjunto com as três práticas, indica que seria pouco provável obter a alta pontuação nas ações de nível estratégico para a governança de IA, sem que as três práticas tivessem sido implantadas, independente do estágio em que estejam.

Pelo exposto, a análise da proposição 2D2 indicou a existência de associação entre as organizações em estágio mais avançado na implantação das ações de nível estratégico para a governança de IA e casos de implantação, em qualquer estágio, de práticas para monitoração automática, supervisão humana e coleta de *feedback*, sugerindo que o pensamento dos gestores dessas organizações alinha-se às argumentações de De Silva e Alahakoon (2022), Laato et al. (2022), Straub (2021), González et al. (2020), Zicari et al. (2021), Dignum (2019) e Hickman

(2020), quando destacam a necessidade de tais práticas. Resumem-se na figura 6 as associações encontradas nas *fuzzy* QCA aplicadas às proposições 2, 2A, 2B, 2C, 2D, 2D1, e 2D2.

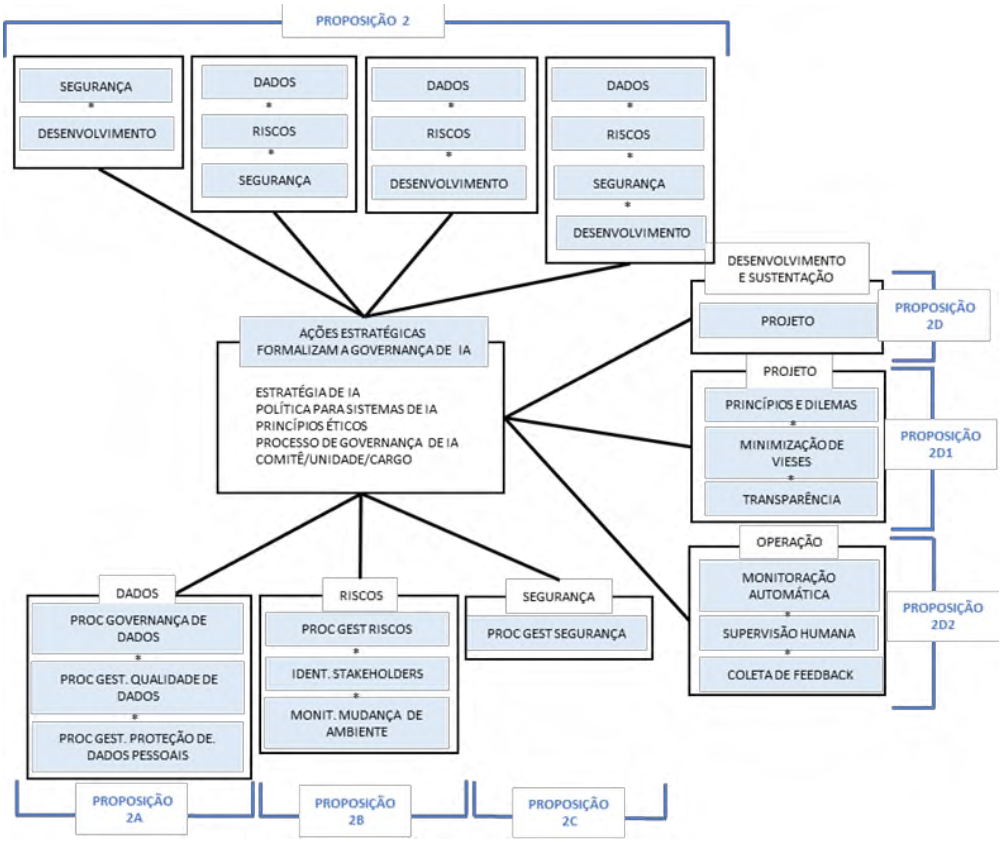

Figura 6: Associações encontradas nas QCA para as proposições 2, 2A, 2B, 2C, 2D, 2D1, 2D2.
Fonte: elaboração própria

Ressalta-se que todas as práticas e processos considerados nos grupos analisados apresentaram associação com estágios mais avançados de ações estratégicas que formalizam a existência de uma governança de IA, exceto o processo de auditoria de sistemas de IA. Tais associações podem indicar que a existência de ações estratégicas para governança de IA produz algum direcionamento na implantação de tais práticas e processos.

3.6.3.3 Análise da Proposição 3

A análise da proposição 3 contemplou como variáveis conjunto da *fuzzy* QCA, a realização de treinamento sobre dados, riscos e princípios éticos em IA para tomadores de decisão (TDECISOR); treinamento sobre dados, desenvolvimento de sistemas de IA, riscos e princípios éticos em IA para desenvolvedores (TDESENV); a treinamento e programas de comunicação sobre dados para usuários (TUSUARIO); e treinamento sobre dados, riscos e princípios éticos em IA para auditores internos (TAUDITOR). Como variável-conjunto resultado, considerou-se a pontuação geral das ações voltadas à implantação da governança de IA (PGERALGOVIA), definida na equação 3.

PGERALGOVIA=MÉDIA(FGOVIA;PROCESSOS,PDESENV) Equação 3;

PDESENV=MEDIA(SISPROJETO;SISOPERACAO) Equação 6;

FGOVIA=MEDIA(MÁXIMO(GOV1;GOV2);GOV3;MÁXIMO(GOV4;GOV5) Equação 7;

PROCESSOS= MEDIA(PROGOVD; PROQUAD; PRODPES; PRORISC; PROSEG; AUDIT) Equação 8.

Comparados às demais variáveis-conjunto analisadas nas proposições, os treinamentos apresentaram baixas médias. Entre os treinamentos, aqueles dirigidos aos usuários apresentaram maior média (53,68), e os treinamentos para auditor (34,61), menor média. Registre-se, ainda, que 14,29% das organizações declararam não ter oferecido qualquer um dos treinamentos citados. Após calibragem, geração de tabela verdade, e escolhida a solução complexa (tabela 18), duas combinações foram geradas.

A combinação 1 (tabela 18) composta pela presença de treinamento para desenvolvedores de sistemas de IA e ausência de treinamento para auditores. Entre os casos que optaram pela cominação 1, 57,14% (27,19,20,28) obtiveram alta pontuação geral para ações de implantação da governança de IA, enquanto os demais casos (14,25,26) não obtiveram a citada pontuação. A pequena diferença entre os casos de presença e ausência de PGERALGOVIA1 não permite que se reconheça associação da combinação 1 com alta pontuação para ações de implantação da governança de IA.

Tabela 18: Parâmetros e valores gerados da QCA para análise da proposição 3.

Proposição 3 - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Mínimo	Máximo	Limiares para Calibragem		
					Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	TDECISOR	40.57	0	100	100	60	0
	TDESENV	41.71	0	100	100	60	0
	TUSUARIO	53.68	0	100	100	60	0
	TAUDITOR	34.61	0	100	100	60	0
Resultado	PGERALGOVIA	57.62	0	100	96.33	60	5.56

Proposição 3 - Tavela Verdade

TDECISOR1	TDESENV1	TUSUARIO1	TAUDITOR1	PGERALGOVIA1	FREQ
0	1	0	0	1	1
1	1	1	0	1	2
1	1	0	0	1	3
0	1	1	0	1	1
1	1	1	1	1	6
1	0	0	0	0	1
1	0	1	1	0	2
0	0	0	1	0	1
0	0	1	1	0	3
0	0	0	0	0	4
0	0	1	0	0	4

Proposição 3 - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.669199	0.68232	0.68232
Consistência da solução	0.919355	0.883721	0.883721
Corte de frequência	1	1	1
Corte de consistência	0.938333	0.938333	0.938333

Combinação 1			Resultado	
condições			resultado investigado	
TDESENV1 * ~TAUDITOR1			27,19,20,28	PGERALGOVIA1
cobertura tota	cobertura única	consistência	contradição	
0.452348	0.174724	0.927762	14,25,26	

Combinação 2			Resultado	
condições			resultado investigado	
TDECISOR1 * TDESENV1 * TUSUARIO1			13,22,27,8,9,24,28	PGERALGOVIA1
cobertura tota	cobertura única	consistência	contradição	
0.494475	0.216851	0.937173	11	

Fonte: Adaptação a partir dos valores resultantes dos cálculos realizados pelo software fsQCA 3.1

A combinação 2 (tabela 18) é caracterizada pela presença de treinamento para os tomadores de decisão, treinamento para os desenvolvedores, além de treinamento e campanhas de comunicação para os usuários. Entre os casos que se situam na combinação 2, 87,5% (13,22,27,8,9,24,28) obtiveram alta pontuação para ações voltadas à governança de IA. Não obstante a existência de um caso contraditório, a grande diferença da proporção dos casos de implantação mais avançada em ações para governança de IA e aqueles de menor avanço, permite a confirmação das previsões de Calzada e Almirall (2020); Micheli et al. (2020); Ruijer (2021) e Benfeldt et al. (2020), de que ampla capacitação na organização voltada a dados, a riscos e a princípios éticos no uso e desenvolvimento de sistemas de IA seria necessária para habilitar as organizações na implantação da governança de IA. Kuziemski e Misuraca (2020) e Benfeldt et al. (2020) destacam a necessidade de sensibilização e capacitação de gestores, além dos próprios usuários, na proteção de dados pessoais, e na mitigação de problemas com entrada inadequada dos dados.

A variável-conjunto de treinamento dos usuários apresentou a maior média (53,68) entre as capacitações estudadas, o que pode ter ocorrido, entre outras razões, pelo grande número de leis para proteção de dados pessoais que geralmente estabelecem sanções quando inconformidades são

detectadas, e pelo fato de 100% dos casos da amostra estarem sujeitos a alguma lei de proteção de dados pessoais (conforme análise da proposição 1).

A baixa média do treinamento para auditores e ausência do treinamento para auditores nas duas combinações revelam que auditoria interna em boas práticas para desenvolvimento de sistemas de IA não ter sido ainda uma prioridade a essas organizações, apesar de um pequeno grupo ter feito a opção pelos quatro treinamentos. Tais resultados confirmam as argumentações de Mökander e Foridi (2021) de que as organizações ainda carecem de métricas padronizadas para avaliação que considerem os princípios éticos no desenvolvimento de sistemas de IA. Entre as razões para esse cenário, podem ser citadas: a inexistência de lei específica dirigida a aspectos éticos de tratamento de dados ou de sistemas de IA na maioria dos casos da amostra (discutido na análise da proposição 1); ou seja, ainda não há conformidade a ser atendida. Adicionalmente, convém a reflexão de que a decisão por implantar auditoria interna passa por gestores que ainda estão no desafio de verem os protótipos que usam IA se concretizarem como sistemas do portfólio de serviços digitais da organização; como demonstrou o processo de seleção da amostra (figura 3) em que muitos sistemas de IA ainda eram protótipos. Destaca-se a presença do treinamento de desenvolvedores de sistemas de IA nas combinações 1 e 2, sugerindo ser uma condição necessária, porém, não suficiente, à alta pontuação nas ações para governança de IA. Convém ainda o registro de que, a decisão por capacitar desenvolvedores não foi detectada apenas em organizações que possuem sistemas de IA desenvolvidos pelo seu próprio corpo funcional; tendo havido treinamentos também por aquelas que contrataram ou estabeleceram parcerias para o desenvolvimento de sistemas de IA, destacando o reconhecimento de sua responsabilidade pelos resultados dos sistemas de IA que suportam os serviços públicos, como argumentam Hickman (2020) e Zuiderwijk et al. (2021).

Organizações em estágio mais avançado de implantação das ações para governança de IA haviam realizado treinamento para tomadores de decisão, para desenvolvedores de sistemas de IA e para usuários; sugerindo assim, associação entre maiores pontuações nas ações para implantação da governança de IA e a combinação de oferta de treinamento para esses *stakeholders* (figura 7), em pensamento alinhado a Calzada e Almirall (2020), Micheli et al. (2020), Ruijer (2021), Benfeldt et al. (2020), Kuziemski e Misuraca (2020) e Benfeldt et al. (2020).

3.6.3.4 Análise da Proposição 4

A análise da proposição 4 requereu, para as condições da *crisp-set* QCA, a criação de variáveis-conjunto dicotômicas para indicar a realização de treinamento de tomadores de decisão (TREINATD), a realização de treinamento para desenvolvedores (TREINAD), o acesso da organização a pelo menos, 80% do código dos seus sistemas de IA (ACOD80), e mais de três anos de experiência no desenvolvimento de sistemas de IA (TSIA) (tabela 19). Para variável-conjunto resultado, considerou-se a pontuação geral das ações para implantação da governança de IA (PGERALGOVIA) como utilizada nas proposições 1 e 3.

Tabela 19: Parâmetros e valores gerados da QCA para análise da proposição 4.

Proposição 4 - Tabela Verdade

ACOD80	TSIA	TREINATD	TREINAD	MAIORGGOVIA	ORGANIZAÇÕES
0	0	0	0	0	17
0	0	0	1	1	19
0	0	1	0	1	5
0	0	1	1	C	20(1), 22(0), 28(1)
0	1	0	0	C	1(0), 4(0), 6(1)
0	1	1	1	1	25
1	0	0	0	0	12
1	0	1	0	0	7
1	0	1	1	0	11
1	1	0	0	C	2(0), 3(1), 10(0), 15(0), 18(0), 21(0), 23(0)
1	1	0	1	0	14
1	1	1	0	0	16
1	1	1	1	C	8(1), 9(1), 13(1), 24(1), 26(0), 27(1)

Combinação 1		Resultado
condições	casos com o resultado investigado	
TSIA{1} * TREINATD{0} * TREINAD{0}	6,3	MAIORGGOVIA
	contradições	
	1,4,2,10,15,18,21,23	

Combinação 2		Resultado
condições	casos com o resultado investigado	
ACOD80{0} * TSIA{0} * TREINATD{1}	5,20,28	MAIORGGOVIA
	contradições	
	22	

Combinação 3		Resultado
condições	casos com o resultado investigado	
ACOD80{0} * TSIA{0} * TREINAD{1}	19,20,28	MAIORGGOVIA
	contradições	
	22	

Combinação 4		Resultado
condições	casos com o resultado investigado	
TSIA{1} * TREINATD{1} * TREINAD{1}	8,9,13,27,24,25	MAIORGGOVIA
	contradições	
	26	

Fonte: Geração feita pelo software Tosmana 1.6.1

A tabela verdade (tabela 19) gerada permitiu que se configurassem quatro combinações. A combinação 1 é caracterizada por casos com mais de 3 anos na produção de sistemas de IA sem treinamento aos tomadores de decisão e sem treinamentos aos desenvolvedores de sistemas de IA. Destaca-se o fato de que 80% dos casos dessa combinação não obtiveram alta pontuação geral para ações de implantação da governança de IA. As combinações 2 e 3 apresentam muitas semelhanças. Ambas, caracterizam-se pela inexistência de acesso da organização pública analisada a, pelo menos, 80% do código dos sistemas de IA declarados no questionário, e por terem até três anos de experiência na produção de sistemas de IA. A inexistência de acesso aos códigos dos seus próprios sistemas retrata a situação de organizações que contrataram externamente o desenvolvimento de tais sistemas.

Importa o registro de que, enquanto a combinação 2 treinou tomadores de decisão, a combinação 3 treinou desenvolvedores de sistemas de IA. Em cada uma das combinações, 75% dos casos apresentaram alta pontuação geral de ações para implantação da governança de IA. Adicionalmente. identificou-se uma interseção de 75% dos casos entre os conjuntos das combinações 2 e 3, o que significa serem organizações que treinaram tanto tomadores de decisão quanto desenvolvedores.

A combinação 4 caracteriza-se por organizações com mais de 3 anos atuando com produção de sistemas de IA, tendo provido treinamento tanto para tomadores de decisão, quanto para desenvolvedores de sistemas de IA. Observou-se que 85,71% das organizações que fizeram tal opção (8,9,13,27,24,25), obtiveram alta pontuação nas ações voltadas à governança de IA. Percebe-se que a combinação 4 apresentou maior proporção entre aqueles que obtiveram alta pontuação das ações para governança de IA, em relação aos que não a obtiveram. Este cenário ilustra a união das percepções de Ahn e Chen (2022) e de Benfeldt et al. (2020), quanto à necessidade de envolver e capacitar *stakeholders* tanto da esfera gerencial, quanto técnica. Cumpre a reflexão de que treinar tomadores de decisão pode significar capacitar gestores de níveis estratégicos, geralmente, responsáveis por aprovações da estratégica, da política que pode proibir ou autorizar a terceirização, do orçamento e dos editais. Geralmente são os responsáveis pela assinatura do contrato em nome da organização pública. Ademais, responsáveis por unidades de ciência de dados, geralmente, são gestores que possuem conhecimento técnico, estando, portanto, aptos aos dois tipos de treinamentos em evidência nesta análise.

Observando a tabela verdade, verifica-se que a combinação ACOD80{1} * TSIA{1} * TREINATD{1} * TREINAD{1} é um subconjunto da combinação 4, correspondente a todos os casos daquela combinação que apresentaram alta pontuação geral para ações de implantação da governança de IA. Logo, organizações com mais de três anos de experiência na produção de sistemas de IA, que tinham acesso a, pelo menos, 80% do código de seus sistemas de IA, e que treinaram tomadores de decisão e desenvolvedores nesses sistemas, apresentaram associação a um estágio mais avançado na implantação da governança de IA.

Resgatando-se a análise da combinação 1, pode-se considerar a existência de associação entre não treinar gestores e desenvolvedores com baixa pontuação nas ações para governança de IA, mesmo em casos com mais de três anos de experiência na produção de sistemas de IA.

Confirma-se, para a amostra estudada, a proposição 4, pois, após três anos de experiência na produção de sistemas de IA, a realização de treinamento de gestores e desenvolvedores de sistemas de IA associou-se positivamente à probabilidade das organizações estudadas criarem condições de terem acesso aos códigos dos seus sistemas de IA e se posicionarem em estágios mais avançados de implantação da governança de IA (figura 7).

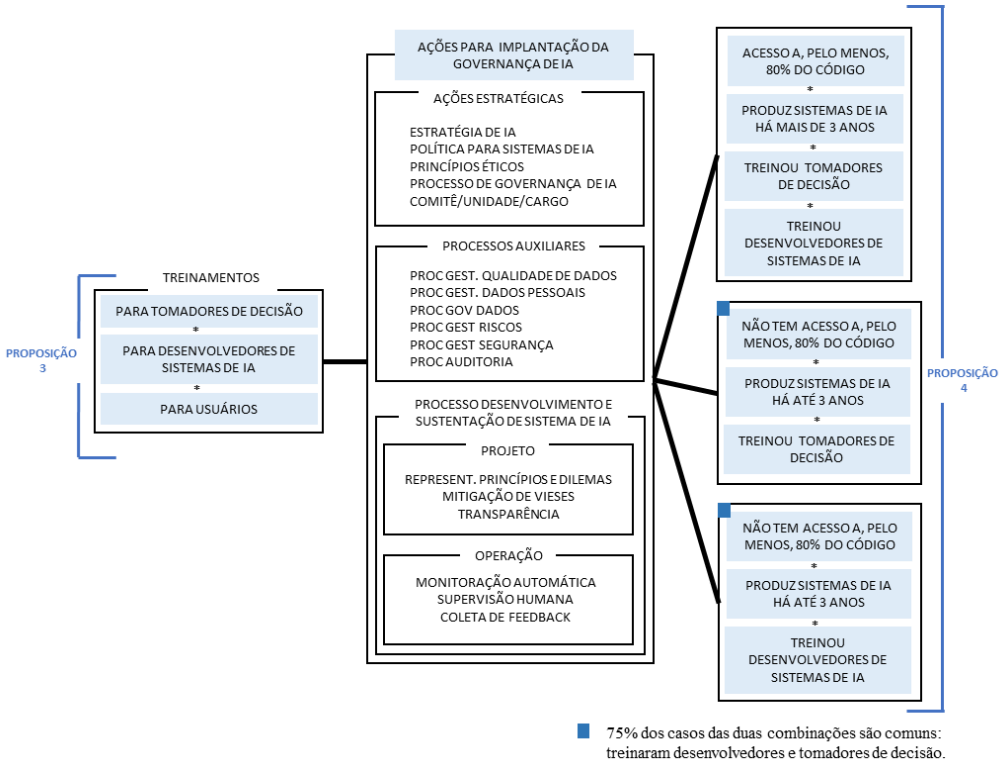

Figura 7: Resumo das associações encontradas das fuzzy QCA realizadas para as proposições 3 e 4.
Fonte: Elaboração própria

3.6.3.5 Análise da Proposição 5

Para analisar a integração entre o processo de gestão de riscos e os demais processos estudados, aplicou-se a *fuzzy* QCA às variáveis-conjunto: processo de gestão da qualidade de dados (PROQUAD), processo de gestão da proteção de dados pessoais (PRODPES), processo de gestão da segurança de sistemas de IA (PROSEG), processo de auditoria de sistemas de IA (AUDIT), práticas do processo de desenvolvimento e sustentação de sistemas de IA dirigidas aos princípios éticos (PDESENV); assim como à variável-conjunto resultado: processo de gestão de riscos em sistemas de IA (PRORISC).

Após calibragem, conversão para valores *fuzzy*, a tabela verdade fundamentou a geração das soluções, complexa, parcimoniosa e intermediária (tabela 20), as quais apresentaram consistência no limite inferior aceito por Ragin (2008), 0,75. A solução complexa apresentou única combinação caracterizada pela presença de processo de gestão da qualidade de dados, processo de gestão da proteção de dados pessoais, processo de gestão da segurança e processo de auditoria dos sistemas de IA (PROQUAD1* PRODPES1* PROSEG1 *AUDIT1). Apenas 60% dos casos que haviam optado pela combinação 1 (8,22,3,6,9,13), também haviam implantado, em qualquer estágio, o processo de gestão de riscos aplicável a sistemas de IA.

Conforme a tabela verdade, observa-se, no entanto, dos dez casos que compõem a combinação 1, nove casos apresentam, além das quatro condições integrantes da combinação 1, a implantação das ações associadas ao processo de desenvolvimento de sistemas de IA dirigidas ao atendimento de princípios éticos. E todos os casos que optaram por esta combinação e apresentaram alta pontuação nas ações para a governança de IA, haviam também implantado as práticas do processo de desenvolvimento e sustentação de sistemas de IA voltadas às questões éticas (8,22,3,6,9,13). Importa ainda resgatar que tal conjunto é parte do conjunto de casos da combinação 1, haja vista {PROQUAD1*PRODPES1*AUDIT1*PROSEG1*PDESENV1}⊂ {PROQUAD1*PRODPES1*AUDIT1*PROSEG1.

Importa registrar que o processo de gestão de riscos apresentou a menor média das variáveis-conjunto deste bloco (34,5), o que gera uma percepção de que as organizações da amostra estão menos avançadas na implantação do processo de gestão de riscos, em comparação à implantação dos processos para gestão da qualidade de dados, para gestão de proteção de dados pessoais, para gestão de segurança dos sistemas de IA, de processo para auditoria, e das práticas do processo de desenvolvimento de sistemas de IA dirigidas ao atendimento de princípios éticos. O fenômeno

descrito pode ser resultado de uma priorização em tratar riscos associados às questões éticas dos sistemas de IA diretamente nos processos sobre dados, segurança, desenvolvimento e auditoria, sem criar ou envolver nova equipe especializada para articular a gestão de riscos.

Tabela 20: Parâmetros e valores gerados da QCA para análise da proposição 5.

Proposição 5 - Calibragem para aplicação da fuzzy QCA

Variáveis-conjunto		Média	Vmínimo	Vmáximo	Limiares para Calibragem		
					Adesão plena	Ponto de Cruzamento	Não-adesão plena
Condição	PROQUAD	58.46	0	100	100	60	0
	PRODPES	90.5	33	100	100	60	33
	PROSEG	46.5	0	100	100	60	0
	AUDIT	45.32	0	100	100	60	0
	PDESENV	60.23	0	100	100	60	0
Resultado	PRORISC	34.5	0	100	100	60	0

Proposição 5 - Tavela Verdade

PROQUAD	PRODPES	AUDIT	PDESENV	PROSEG	PRORISC	FREQ
1	1	1	0	1	1	1
1	1	1	1	1	1	9
0	1	1	1	0	0	2
0	0	1	0	0	0	1
0	1	0	1	1	0	1
1	1	1	1	0	0	1
1	1	1	0	0	0	1
1	1	0	1	0	0	1
1	1	0	1	1	0	1
1	1	0	0	1	0	3
0	1	0	1	0	0	1
0	0	0	0	0	0	1
1	1	0	0	0	0	3
0	1	0	0	0	0	2

Proposição 5 - Soluções

Índices de Análise	Complexa	Parcimoniosa	Intermediária
Cobertura da solução	0.62527	0.69379	0.625268
Consistência da solução	0.7449	0.764151	0.744898
Corte de frequência	1	1	1
Corte de consistência	0.79063	0.790633	0.790633

Combinação 1			Resultado
condições			casos com o resultado investigado	
PROQUAD1 * PRODPES1 * AUDIT1 * PROSEG1			8,22,3,6,9,13	PRORISC1
cobertura total	cobertura única	consistência	contradição	
0.625268	0.625268	0.744898	5,19,20,28	

Fonte: Adaptação a partir dos valores gerados pelo software fsQCA 3.1.

Pelo exposto, a *fuzzy* QCA aplicada à amostra permitiu observar uma baixa associação do processo de gestão de riscos, nos poucos casos onde ele foi implantado, a um pequeno grupo que também implantou todos os processos auxiliares e as práticas aplicadas às questões éticas no desenvolvimento e sustentação de sistemas de IA.

3.6.4 Análise de conteúdo das entrevistas e documentos disponibilizados

A análise de conteudo das entrevistas e material disponibilizado pelas organizações foi realizada usando o software MAXQDA versão 2022, primeiramente, tentando entender como foram aplicados práticas e processos categorizados no modelo de pesquisa e utilizados nas análises das proposições nas QCA apresentadas. A seguir são apresentadas as categorias em que houve relato ou documento disponibilizado.

3.6.4.1 Padrões internacionais

No contexto de padronizações, apenas duas organizações declaram utilizar o padrão XAI (Das 2020; Dazeley et al 2021; Adadi & Berrada 2018) no propósito de prover transparência e explicabilidade durante o processo de desenvolvimento de sistemas de IA. Em sentido amplo, entre os governos da amostra, foram identificadas três iniciativas em defesa do uso de padrões internacionais. O governo do Reino Unido, por meio do Instituto Alan Turing (Leslie 2019) apresentou o XAI entre as alternativas de prover transparência ao código dos sistemas. O governo do Japão, por meio do *National Institute of Advanced Industrial Science and Technology* (2022) recomendou adesão a padrões ISO de qualidade de software e de segurança. E o governo da Alemanha apresentou um conjunto de padrões ISO voltados diretamente ao desenvolvimento de sistemas de IA (nos temas riscos, qualidade de dados, governança de dados e segurança) auxiliares no desafio de prover a governança de IA (German Federal Ministry for Economic Affairs and Energy 2020).

3.6.4.2 Ações estratégicas de governança de IA

No contexto das ações estratégicas para governança de IA, quando questionadas sobre como ocorriam a decisões mais estratégicas relativas à IA, citando-se os seus componentes, as organizações manifestaram-se como segue.

3.6.4.2.1 Relatos sobre comitês

Além dos comitês executivos da organização (anexo 7 - figura 1.a), foram identificados comitês de inovação, comitês de dados, comitês de ética, comitês diretivos de TI e comitês de especialistas de TI, neste último caso, para discutir a viabilidade técnica do futuro sistema de IA e para discutir aspectos multidisciplinares referentes ao sistema de IA (anexo 7 – figura1.b). Os

comitês de dados, geralmente criados para atender às questões de conformidade com a proteção de dados pessoais, apresentaram atuação técnico jurídica, o que coaduna com o fato de que todas as organizações da amostra são sujeitas a algum tipo de legislação para proteção de dados pessoais (anexo 7 – figura1.c). A atuação do comitê de ética foi encontrada tanto com uma conotação voltada para a legalidade dentro dos aspectos sociais, sendo independente do comitê de dados; quanto em uma abordagem dirigida às questões que envolvem pesquisas com pacientes, sendo nesse caso, com atuação híbrida, para atender aspectos ligados à proteção de dados pessoais (anexo 7 – figura1.d).

3.6.4.2.2 Relatos sobre processo para a governança de IA

Algumas organizações expressaram mais claramente o processo decisório, ou parte dele, em que as questões de nível tático e estratégico referentes à IA são deliberadas (anexo 7 – figura2.a), corroborando com a percepção do diversificado conjunto de *stakeholders* necessários à governança de IA em uma organização pública (Dignum 2022; Vanhée & Borit 2022). Em alguns casos, as decisões são tomadas pela unidade de ciência de dados localizada dentro da área de TI, e todas as demais decisões, em comitê de nível hierárquico mais alto da organização, acompanhado por unidade de proteção de dados (anexo 7 – figura2.b).

A percepção da necessidade de se trabalhar com projetos formalmente definidos, e a existência de processo para aprovação do desenvolvimento de sistema de IA foram relatados (anexo 7 – figura2.d). Faz-se igualmente importante o registro de que, quando os gestores de nível estratégico não estão deliberando sobre o uso da IA, outras instâncias sem o necessário empoderamento, estão sendo obrigadas a deliberar, o que pode ser um elemento de fragilidade quanto à sustentabilidade das decisões a longo prazo (anexo 7 – figura2.c).

6.4.2.3 Relatos sobre políticas para o uso da IA

A adoção de algum normativo dirigido ao uso e desenvolvimento de sistemas de IA tem ocorrido, basicamente, em dois formatos: elaboração de normativo com política para IA própria da organização, ou incorporação de política para IA desenvolvido por alguma agência central do Governo com o propósito de ser aplicado aos demais departamentos.

Nas entrevistas, houve relatos nas duas situações. Na primeira opção, além dos relatos sobre normativos criados pelas organizações da amostra (anexo 7 – figura3 a), houve relatos sobre uso de normativos governamentais sobre dados, sobre responsabilizações das áreas de negócio em

relação aos sistemas de IA, enquanto os normativos mais específicos da organização pesquisada estavam sendo construídos, permitindo a implantação dos requisitos mais básicos dos processos e práticas.

Em organização que contratou externamente o desenvolvimento dos seus sistemas de IA, centralizou-se nos editais as regras para o desenvolvimento de manutenção dos sistemas (anexo 7 – figura3.b). O tempo e o perfil dos profissionais envolvidos na elaboração da política de IA também foram itens contemplados (anexo 7 – figura3.c). Na opção de se incorporar políticas gerais para todo o governo, os relatos ocorreram no sentido do reconhecimento de sua importância e abrangência a todas as agências do governo (anexo 7 – figura3.d)

Um caso de abordagem híbrida ocorreu com uma das organizações da amostra, que apesar de estar sujeita à uma política em sua esfera de governo, produziu sua própria política (LIAA-3R 2022) incorporando as diretrizes da primeria política (CNJ 2020), mas, permitindo o acréscimo de novos elementos associados às particularidades do órgão.

3.6.4.2.4. Relatos sobre princípios éticos

Referindo-se especificamente aos princípios éticos dirigidos a sistemas de IA da amostra, observou-se que, entre aqueles que haviam declarado no questionário possuir um guia de princípios éticos, uma parte das organizações havia criado seus próprios princípios (anexo 7 – figura4.a)(LIAA-3R 2022; Vero 2019), e outra parte havia adotado os princípios estabelecidos por algum normativo de governo construído para todos os ministérios, agências, departamentos ou tribunais em sua esfera de competência (Government of Canada 2020a; Government of United Kingdom 2020a.) (anexo 7 – figura4.c).

3.6.4.3. Relatos sobre governança de dados, gestão da qualidade de dados, gestão da proteção de dados pessoais

Alinhando-se às variáveis-conjunto estudadas nas QCA, foram extraídas menções sobre governança de dados, qualidade de dados e proteção de dados pessoais.

A percepção dos entrevistados em relação a Governança de Dados foi presente tanto em organização de avançado estágio na implantação da governança de dados, quanto naquelas que ainda estão em processo inicial de sua implantação. No primeiro caso, o relato descreve com clareza ações deliberativas acerca dos dados em todo o seu ciclo de vida – nomeação de proprietário dos

dados, aprovações de tratamento nos dados, definição do tempo de retenção dos dados, etc. (anexo 7 – figura5.a). E em casos de implantação da governança de dados em estágio inicial, alguns relatos demonstram familiaridade com os termos próprios desse processo (anexo 7 – figura5.b). Sendo a governança de dados um processo dirigido a deliberações, em geral, feitas por atores das unidades de negócio sobre os dados dos serviços digitais sob sua responsabilidade; observaram-se dificuldades quando a agência ou departamento é um centro de produção de sistemas de IA para todo o governo. Nessas situações, o conhecimento do negócio ocorre na organização demandante dos sistemas de IA, deixando a organização que produz e mantém tais sistemas, na dependência da primeira, a cada ação que requerer autorização de tratamento nos dados (anexo 7 – figura5.c).

A dependência que a qualidade de dados em relação aos atores que dão entrada e alteram os dados é clara para alguns entrevistados, assim como o desafio de melhorar o nível dessa qualidade. Como ilustração, uma organização relatou ter criado uma premiação para a unidade que apresentasse menor ocorrência de erros na entrada dos dados (anexo 7 – figura6.b), e outra implantou práticas de padronização de modelos de entrada e armazenamento de dados com o propósito de melhorar a qualidade de dados (anexo 7 – figura6.b). Destacam-se as percepções de que a gestão da proteção de dados pessoais seja mecanismo de viabilizar parte da governança de dados, identificada na combinação 1 da QCA realizada para a proposição 2A (anexo 7 – figura6.a).

Como reflexo de estarem sujeitas à legislação sobre proteção de dados pessoais, os relatos sobre ações com esse propósito foram frequentes. Há casos em que formulários próprios para proteção de dados pessoais fazem parte da documentação do processo de desenvolvimento de sistemas de IA (LIAA-3R 2022), assim como orientações técnicas gerais do governo, para todas as agências sobre sua responsabilidade, são incorporados ao processo de desenvolvimento (Government of United Kingdom 2021d; Information Commissioner's Office 2021, Government of United Kingdom 2017) . Na mesma direção, identificou-se atuação combinada entre gestão de proteção de dados pessoais e a gestão da segurança de dados em sistemas de IA, apontadas na QCA para estudo da proposição 5. Nesses casos, também apontadas as dificuldades em se construir sistemas de IA, quando estes envolvem dados pessoais e segurança de dados, algumas vezes, inviabilizando todo o projeto. O problema se amplia quando a organização que vai desenvolver o sistema de IA não é a proprietária dos dados, mesmo que ambas sejam organizações do mesmo governo (anexo 7 – figura6.b).

3.6.4.4. Gestão da Segurança de Sistemas de IA

As percepções acerca do tema segurança foram observadas em relatos com ênfase no controle de acesso aos dados durante o processo de desenvolvimento, especialmente, quando este ocorre por meio de contratos com outras empresas do setor privado ou instituições públicas (anexo 7 – figura7.a), tendo também ocorrido relatos da segurança na perspectiva de proteção dos dados pessoais (anexo 7 – figura7.b), como apontada a combinação gerada pela QCA da proposição 5.

3.6.4.5 Relatos sobre Riscos

As percepções acerca do tema riscos foram observadas em três conotações. Primeiramente, por meio de guias estabelecidos de forma centralizada de Governo para as agências e departamentos a ele subordinados, iniciando-se por uma avaliação de riscos (Moeini & Rivard 2019 (Government of Canada 2020a; Government of Canada 2021) (Government of United Kingdom 2022a) (anexo 7 – figura8.a).

Nesses modelos de avaliação, os resultados da aplicação do questionário em organizações públicas dos governos citados estão disponívies na internet (Government of Canada 2022; Government of United Kingdom 2022b). Ainda na amostra em estudo, uma organização da Dinamarca criou um processo para avaliação de riscos de sistemas de IA, AIRA (Nagbøl et al. 2021). Baseada no modelo de avaliação canadense, o AIRA, por meio de processo, amplia a abordagem por enfatizar e promover a comunicação entre *stakeholders* da área de negócio com aqueles especialistas de dados e sistemas, o que permite obter análises qualitativas e quantitativas na avaliação dos riscos.

Observam-se, portanto, a existência de elementos informacionais sobre proteção de dados pessoais, segurança, auditoria e desenvolvimento de sistemas de IA nos três modelos de avaliação de riscos analisados. E em apenas um deles, sobre qualidade de dados.

3.6.4.6 Relatos sobre auditoria em sistemas de IA

As percepções sobre auditoria somente ocorreram em um caso, cuja organização definiu um documento próprio a ser preenchido por unidade para esse fim (LIAAR-3R 2022).

3.6.4.7 Relatos sobre contratações

Em 73,33% das organizações da amostra, houve contratação de terceiros ou estabelecimento de parceria para desenvolvimento de, pelo menos, parte dos seus sistemas de IA. Identificou-se, desde a ausência completa de acesso aos códigos contratados, até a exigência contratual de que os códigos fossem disponibilizados no GitHub para a sociedade. Durante a pesquisa, alguns participantes relataram práticas para que requisitos fossem inseridos nos contratos com o intuito de minimização de riscos de segurança (anexo 7 – figura9.a), questões de propriedade intelectual (anexo 7 – figura9.b), controle do código (anexo 7 – figura9.c) e, razões e limitações dessas contratações (anexo 7 – figura9.d).

A ciência de que as organizações públicas não são autossuficientes para produção de sistemas de IA na proporção e especialidade de que necessitam (Bendszus 2022), organizações e pesquisadores têm buscado modelos de especificações que garantam o atendimento de princípios éticos; o que tem motivado acordos entre o *World Economic Forum* e vários governos (Government of United Kingdom 2020b; Government of United Kingdom 2020c; World Economic Forum 2020a; World Economic Forum 2020b; World Economic Forum 2020c; C4IR Brasil 2022). O modelo do WEF endereça alternativas de soluções para mitigar muitos riscos apontados por Hickok (2022), e fundamenta-se em uma avaliação de riscos e impactos dos sistemas a serem contratados, antes de se chegar aos requisitos para cada caso.

3.6.4.8 Relatos sobre minimização de vieses no desenvolvimento de sistemas de IA

No escopo de um processo de desenvolvimento de sistemas de IA, os relatos e documentos disponibilizados referiam-se a práticas para minimizar vieses e prover transparência na fase de projeto, e formas de acompanhamento dos sistemas de IA, quando em operação. Algumas organizações expressaram que seu fluxo de trabalho para minimização de vieses envolvia equipe de ciência de dados, e/ou comitê de ética e/ou profissionais da área de negócio (anexo 7 – figura10.a). Observou-se a preocupação com diferentes percepções do alcance da palavra viés, requerendo clareza quando se trabalha com times multidisciplinares (anexo 7 – figura10.b).

Uma das organizações inseriu no seu *framework* para construção de sistemas de IA, um *sub-framework* que contempla práticas para minimização de vieses de desenvolvimento (Nagbøl & Müller 2020) (anexo 7 – figura10.c). Diversas experiências em ações para minimização de vieses foram relatadas contemplando evitar discriminação de etnia, de classe social, e de religião

(islamismo e judaísmo) (anexo 7 – figura10.d). Tanto as técnicas apontadas no anexo 7 - – figura10.c, quanto anexo 7 – figura10.d coadunam com as práticas recomendadas por Ashokan e Haas (2021), Makhlouf et al. (2021) e González et al. (2020).

Considerando apenas a amostra da presente pesquisa, o governo do Reino Unido, em vários guias sobre viés algorítmico (Government of United Kingdom 2019b; Leslie 2019; Government of United Kingdom 2021d) apresenta aos servidores públicos diversos tipos de aplicações em que a existência de viés tem alta probabilidade de ocorrer, instrui como identificá-los e aponta atributos comuns entre legislação sobre equidade e sobre proteção de dados pessoais, assim como atributos que só dizem respeito a uma abordagem e não à outra (Government of United Kingdom 2021d). Em uma abordagem dirigida a desenvolvedores, o governo do Japão (National Institute of Advanced Industrial Science and Technology 2022) apresenta técnicas a serem utilizadas para prover *algorithm fairness* em várias etapas do desenvolvimento de sistemas de IA com o propósito de minimizar vieses.

A contratação de profissional com perfil adequado para fazer análises sobre potenciais vieses foi destacada, como habilitadora das práticas para minimizar vieses.

> *"para garantir que não incorreríamos em nenhum viés, montamos um grupo paralelo de pessoas lideradas por um especialista em ética de dados que se concentrava em monitorar se estávamos usando os dados de maneira correta ou se havia algum viés"*

O tipo de especialização profissional para lidar com questões éticas ainda não tem sido muito explorado pelos governos. O governo britânico, em Government of United Kingdom (2020c) recomenda a formação de equipe multidisciplinar, assim como a consideração de perfis de arquiteto de dados, cientista de dados, engenheiro de dados, arquiteto de tecnologia, gestor de implantação, arquiteto de segurança e gerente de negócio. O governo australiano (Australian Government 2019a) recomenda a multidisciplinaridade no projeto com especialistas em ética, cientistas sociais, cientistas de dados, especialistas em privacidade e advogados. Além da multidisciplinaridade, o Instituto Alan Turing recomenda uma equipe inclusiva (Leslie 2019).

3.6.4.9. Relatos sobre transparência no processo de desenvolvimento de sistemas de IA

O grande número de relatos sobre transparência na produção de sistemas de IA revela a variedade de percepções sobre sua aplicabilidade em alguns casos e sobre ações e decisões vistas como viabilizadoras de algum nível de transparência (anexo 7 – figura11.c). Contudo, apenas

6,67% da amostra publica parte dos códigos de seus sistemas de IA para a sociedade. Ademais, o desafio de prócer sistemas de IA transparente não se restringe apenas ao código; envolve transparência do processo e transparência dos resultados, podendo este último ser dividido como clareza do resultado e justificativa dos resultados (Leslie 2019).

As organizações da amostra que apresentaram projetos ou práticas em curso para prover transparência relataram diferentes percepções (Dazeley et al 2021; Adadi & Berrada 2018; Phillips, et al. 2021), como pode ser observado no anexo 7 – figura11. Em estágio mais avançado, uma organização registra ações tomadas, resultados alcançados, e razões para as definições, formando uma sequência de documentos padronizados assinados digitalmente pelas pessoas envolvidas na governança de IA (LIAA-3R 2022).

Por meio de parceria, organizações em dois países seguem processo de transparência criado por uma delas para contemplar todo o ciclo de vida dos sistemas de IA, desde a concepção, com envolvimento direto dos *stakeholders* em reuniões, passando por todas as iterações que o desenvolvimento exige até a fase de monitoração. Publicado para a sociedade, o X-RAI provê transparência do processo de desenvolvimento de sistemas de IA em três subníveis: "*simulabilidade, decomposição, transparência algorítmica. Além disso, ele se baseia em tipos de interpretabilidade post-hoc com as seguintes abordagens: explicações de texto, visualização, explicações locais e explicação por meio de exemplo*" (Nagbøl & Müller 2020 p.2).

Identificaram-se organizações que dirigem suas ações de transparência a publicar artigos científicos explicando a construção dos sistemas de IA. Outras, abrem a documentação apenas para profissionais da própria organização; e há aquelas que abrem apenas para organizações regidas pelo mesmo governo (anexo 7 – figura11.a).

A proximidade e parceria da TI com a área de negócio fez parte dos relatos como mecanismo para viabilizar a transparência, haja vista o fato de que somente o profundo conhecimento do contexto do sistema permite compreensão necessária ao modelo mais adequado e à explicação de seu funcionamento (anexo 7 – figura11.b). Destaca-se organização defensora da maior simplificação possível dos modelos como estratégia de viabilização da transparência, não excluindo outras.

> "*Acho que o problema com a transparência é quando os algoritmos são muito complicados"..." às vezes você obtém modelos muito complicados porque passou por uma evolução e adicionou solicitações ao longo do tempo.. mas, acho que porque tentamos*

acomodar o maior número possível de requisitos e isso muda com o tempo, então os modelos se tornam complicados demais."

De maneira similar ao enfrentamento de vieses, os governos da Noruega, e do Reino Unido têm produzido orientações ao provimento de transparência em todo o processo de desenvolvimento de um sistema de IA (Norwegian Data Protection Authority 2018; Government of United Kingdom 2021a; Government of United Kingdom 2021b; Government of United Kingdom 2021b).

3.6.4.10. Relatos sobre monitoração, supervisão e *feedback* dos sistemas de IA

Os relatos reconhecem tanto da monitoração automática (De Silva & Alahakoon 2022), quanto na supervisão humana (Strauß 2021; González et al. 2020; Zicari et al. 2021; Dignum 2019; Hickman 2020) e na coleta de *feedback* (AI4People 2018; Rahwan 2017; Wright & Schultz 2018). Nessas categorias, houve convergência de relatos de que elas ocorrem na responsabilidade do gestor de negócio do sistema de IA, mesmo que, em algumas situações a execução tenha, operacionalmente, ajuda da unidade de TI (anexo 7 – figura12.a). Também identificada a percepção de que a monitoração é essencial para se manter o ciclo contínuo que sustenta um serviço digital baseado em IA, como apresentado por Laato et al. (2022). A categoria "coleta de *feedback*" recebeu dois relatos de situações bem distantes. Enquanto uma organização deixava claro que o *feedback* era baseado na intuição do usuário, a outra apresentou um método interativo de buscar uma contribuição mais efetiva para a evolução de chatbot (anexo 7 – figura12.b).

3.6.4.11. Interação entre a área de negócio e a área de TI

Além das categorias previamente identificadas no modelo de pesquisa e já trabalhadas nas QCA, uma nova categoria foi identificada ao longo dos relatos e dos documentos analisados: a integração entre a área de negócio e a área de TI (anexo 7– figuras 13, 14 e 15).

A percepção da atuação conjunta "negócio+TI" como estratégia para minimizar vieses e prover transparência (anexo 7 – figura 13), amparam-se na argumentação de que para identificar alguma anormalidade nos resultados e até a identificação de grupos e atributos a serem protegidos depende de longa vivência com o negócio automatizado. Analogamente, ter prontidão para responder questionamentos sobre as regras, assim como explicar profundamente os modelos, ou até auxiliar

a avaliação de modelos mais simples que possam trazer resultados confiáveis e mais facilmente explicáveis, requerem profundo conhecimento do negócio.

O acompanhamento dos sistemas de IA em operação também foi considerado pelos entrevistados como ação que requer estreita parceria "negócio+TI" (anexo 7 – figura 13) com a argumentação de que a área de negócio conhece ou interage com os usuários mais que os profissionais da TI o fazem. Um terceiro grupo de relatos ocorreu por profissionais que implantaram processos de gestão de riscos e processos de desenvolvimento de sistemas de IA, construídos com papéis a serem desempenhados e artefatos a serem preenchidos por atores das unidades de negócio (Nagbøl et al. 2021) (anexo 7 – figura 14).

O reconhecimento dos benefícios que a parceria "negocio+TI" traz para a governança de dados nas atribuições dos proprietários dos dados, gestores da área de negócio (anexo 7– figura 14), e os benefícios dessa parceria para a construção de projetos de sistemas de IA, foram amplamente apresentados durante as entrevistas (anexo 7 – figura 15).

Em suma, as entrevistas e documentos disponibilizados confirmaram a distribuição das práticas e processos contemplados no modelo de pesquisa como integrantes de ações para implantação da governança de IA da amostra. Acrescentaram a existência de relações com agências de governo que estabelecem diretrizes e definem padrões a serem implantados nas organizações públicas, de maneira a facilitar e, muitas vezes, viabilizar a governança de IA no governo de um país. Dentro das organizações, destacaram a necessidade da parceria entre unidades de negócio, responsáveis pelos sistemas de IA, e as unidades de TI, para viabilizar a implantação de algumas práticas contempladas na presente pesquisa.

3.7. Conclusões, agenda e contribuições

O estudo investigou como as organizações públicas têm incorporado as diretrizes apresentadas pela academia, pelos padrões internacionais e pela legislação, ao seu modelo de desenvolvimento de sistemas de IA considerando princípios éticos.

3.7.1. Síntese da análise das proposições

Utilizou-se uma amostra de vinte e oito organizações públicas de dezessete países, com coleta de dados por meio de questionário on-line e, de maneira complementar, entrevista. A primeira parte da análise contemplou a realização de análises de proposições por meio de *fuzzy* QCA e *crisp-set* QCA, cujos resultados são sintetizados na tabela 21.

Tabela 21: Resumo dos resultados das análises das proposições.

Proposição		Resultado
1	A combinação entre legislação e normativos governamentais para regulação da IA (*hard law* e *soft law*) são fatores geradores de expectativas de que organizações públicas implantem sua própria governança de IA.	Não foi possível encontrar associação entre a existência de lei para IA, lei para proteção de dados pessoais, e normativos governamentais com a existência de um estágio mais avançado na implantação da governança de IA. Possíveis justificativas: somente uma organização estava sujeita à lei aplicada a sistemas de IA, e provavelmente, ainda não houve tempo suficiente para que normativos e políticas governamentais tenham provocado reação diferenciada na implantação de governança de IA nas organizações públicas da amostra.
2	Processos e práticas na dimensão "dados", "riscos", "segurança", e "desenvolvimento" devem seguir diretrizes estabelecidas no nível estratégico da governança de IA.	Foi encontrada associação entre as ações estratégicas da governança de IA e processos na dimensão "dados" (processo de governança de dados, processo de gestão da qualidade de dados, processo de gestão da proteção de dados pessoais), na dimensão "riscos" (processo de gestão de riscos, definição de *stakeholders* monitoração de mudanças de ambiente), na dimensão "segurança" (processo de gestão da segurança de sistemas de IA), na dimensão "desenvolvimento" (fase de projeto do sistema de IA, representação formal de princípos e dilemas éticos, práticas para minimização de vieses, práticas para provimento de transparência, monitoração automática, supervisão humana, coleta de *feedback*). Não foi encontrada associação entre o processo de auditoria e as ações estratégicas da governança de IA. Proposição 2 confirmada para todas as práticas e processos testados, exceto para o processo de auditoria de sistemas de IA.
3	O treinamento de *stakeholders* em dados, em desenvolvimento de sistemas de IA e em princípios éticos aplicados ao desenvolvimento de sistemas de IA é habilitador da implantação da governança de IA nas organizações públicas.	Foi encontrada associação entre estágio avançado de implantação das ações para governança de IA e a realização de treinamento para tomadores de decisão, para desenvolvedores de sistemas de IA e para usuários. Não foi encontrada associação entre estágio avançado de implantação das ações para governança de IA e a realização de treinamento para auditores internos. Proposição 3 confirmada para treinamentos dirigidos a tomadores de decisão, a desenvolvedores de sistemas de IA e a usuários. Proposição 3 não confirmada para terinamentos dirigidos a auditores de sistemas de IA.
4	Ao longo do tempo, o treinamento de gestores e desenvolvedores de sistemas de IA amplia a probabilidade das organizações públicas criarem condições de terem acesso aos códigos dos seus sistemas de IA e avançarem na implantação da governança de IA.	Após três anos de experiência na produção de sistemas de IA, a realização de treinamento de tomadores de decisão e desenvolvedores de sistemas de IA associou-se positivamente à probabilidade das organizações estudadas criarem condições de terem acesso aos códigos dos seus sistemas de IA e se posicionarem em estágios mais avançados de implantação da governança de IA. E organizações que, após três anos de experiência na produção de sistemas de IA, não treinaram tomadores de decisão nem desenvolvedores de sistema de IA, apresentaram associação a estágios menos avançados de implantação da governança de IA. Proposição 4 confirmada.
5	Para implantar a governança de IA nas organizações públicas, integra-se o processo de gestão de riscos aos processos de gestão da qualidade de dados, de gestão da proteção de dados pessoais, de gestão da segurança para sistemas de IA, de auditoria nos sistemas de IA e as práticas do processo de desenvolvimento e sustentação de sistemas de IA aplicadas às questões éticas.	Foi encontrada associação entre a implantação do processo de gestão de riscos aplicado a sistemas de IA a apenas organizações que estavam em maior avanço na implantação da governança de IA, as quais haviam implantado, em qualquer proporção o processos de gestão da qualidade de dados, e processo de gestão da proteção de dados pessoais, e processo de auditoria em sistemas de IA, e processo de gestão da segurança e práticas aplicadas às questões éticas no desenvolvimento e sustentação de sistemas de IA. Proposição 5 confirmada apenas para o restrito grupo descrito.

Fonte: elaboração própria

3.7.2 Modelos de relacionamento entre *stakeholders*

As entrevistas e o material disponibilizado pelos governos de alguns países da amostra permitiram que se evidenciassem os benefícios da existência de um órgão do governo que, centralizadamente, produza guias claros e acessíveis aos demais órgãos públicos com recomendações de boas práticas para desenvolvimento de sistemas de IA. Alguns guias compuseram o portfólio de padrões dos governos da amostra para promoção da governança de IA nas agências e departamentos sob sua responsabilidade (Tabela 22).

Além de suprir conhecimento inexistente em muitas organizações públicas, esses órgãos especializados do governo geram celeridade às implantações, e promovem um padrão de conceitos e conhecimento que facilita a comunicação entre departamentos do mesmo governo que precisem compartilhar projetos, dados e experiência. Transcendendo o domínio público, tais padrões auxiliam na preparação de empresas do setor privado para contratos com o governo visando o provimento de serviço de desenvolvimento e sustentação de sistemas de IA que atendam a princípios éticos.

Ainda no contexto de ações governamentais, os depoimentos e documentos compartilhados confirmaram os estudos de Gutierrez e Marchant (2021) e de Marchant (2019) sobre a contribuição das *soft laws* na implantação da governança de IA; enquanto as discussões em torno da legislação ocorrem nas instâncias legislativas.

Adicionalmente, os guias e padrões governamentais permitem ações regulatórias em processos que requerem um nível de adequação ao contexto cultural da organização, não alcançável pela legislação, como, por exemplo, os guias do governo britânico que podem ser combinados, a depender do tipo do sistema de IA (Government of United Kingdom 2017, 2019b, 2020a, 2020b, 2020c, 2020d, 2020e, 2020f, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2022a, 2022b, 2022c, 2022d). Identificou-se também a utilização de parceria entre o setor público e o privado (PPP), como fez o governo da Finlandia (AURORA AI 2019), unindo governo e iniciativa privada na construção de padrões para boas práticas de uso da IA e transferência de conhecimento aos servidores públicos. E, com alcance internacional, foi identificada parceria entre países nórdicos para implantar boas práticas do desenvolvimento de sistemas de IA com foco nas questões éticas (Nordic Council of Ministers 2018), cuja materialização foi encontrada envolvendo organizações da Dinamarca, Finlândia e Islândia na amostra estudada.

Adicionalmente aos guias e padrões, alguns governos atribuem a responsabilidade pelas diretrizes estratégias da IA a uma Agência que elabora e acompanha uma estratégia de IA nacional, estabelece uma política para os sistemas de IA e define os princípios

éticos para uso e desenvolvimento dos sistemas de IA, como observado na análise da proposição 1 e durante as entrevistas. Tal situação permite aos órgãos governamentais adotarem a política de sistemas de IA e os guias éticos estabelecidos ou, construírem versões próprias alinhadas às definições da Agência do governo para esse propósito, muitas vezes, também responsável pela estratégia de transformação digital do governo.

Como observado em algumas situações, a Agência responsável pelas diretrizes estratégicas pode ser distinta da responsável pele elaboração dos guias e padrões, modelo representado na figura 8.

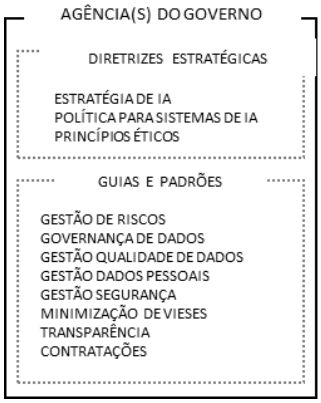

Figura 8: Produtos disponibilizados por Agências de governo para a governança de IA.
Fonte: Elaboração própria

Ao longo das entrevistas, confirmaram-se as percepções de Benfeldt (2020) e Vilminko-Heikkinen (2019) quanto à parceria entre unidades de negócio e unidade de TI para uma efetiva governança de dados; às argumentações de De Silva e Alahakoom (2022), Morley et al. (2020a) e Ashaye et al. (2019) quanto à necessidade de trabalho conjunto entre desenvolvedores e especialistas das unidades de negócio no desenvolvimento e sustentação de sistemas de IA. Contudo, além dessa parceria, complexas relações entre *stakeholders* foram identificadas, confirmando as argumentações de Mäntymäki et al. (2022), e gerando uma especial categoria de análise: a necessidade de forte e contínua integração entre unidades de negócio e a unidade de TI.

A percepção de imprescindibilidade de uma atuação conjunta "negócio+TI" ocorreu em diversas abordagens. No processo de desenvolvimento, quando se abordou a transparência e práticas para minimizar vieses (anexo 7 - figura 13), a unidade de negócio foi considerada imprescindível para a identificação da existência ou possibilidade de ocorrência de vieses e de como o modelo deveria ser explicado. E, na mesma perspectiva, a área de negócio é consultada para confirmar se as medidas tomadas para a minimização de vieses foram suficientes, assim como se a documentação de todo o funcionamento do

sistema é suficiente e corretamente compreensível a uma pessoa externa à área de TI . Ainda no processo de desenvolvimento e sustentação de sistemas de IA, houve relatos da utilização de métodos ágeis de desenvolvimento, confirmando os *loops* internos da fase de projeto do processo de desenvolvimento de sistemas e sustentação de IA, cuja participação conjunta da área de negócio e da área de TI mostrou-se como condição básica (Mohammad 2017); mas, também da união das equipes simplesmente pela alta frequência em que profissionais de negócio precisavam ser acionados. Em um caso, houve relato detalhado de envolvimento de muitas pessoas da unidade de negócio na tarefa de anotação durante a construção supervisionada de um modelo de *machine learning* (anexo 7 - figura 14), e ainda de documento específico a ser preenchido pela unidade de negócio durante essa tarefa (anexo 7 - figura 14), reforçando assim, uma participação na construção do modelo do sistema de IA. Quando se abordou o acompanhamento dos sistemas de IA em operação, a interpretação do *feedback* foi considerada uma ação com a participação do bloco "negócio+TI", e a responsabilidade por monitorar, declarada como sendo da unidade de negócio, mesmo que a área de TI tenha uma participação na operacionalização (anexo 7 - figura 13). Na abordagem de gestão de riscos, a ferramenta AIRA (Nagbøl et al. 2021), criada por uma organização da amostra para gestão de riscos de sistemas de IA, requer participação ativa e integrada da área de TI e das áreas de negócio (anexo 7 - figura 14). Durante as declarações sobre governança de dados, as unidades de negócio foram reconhecidas como as responsáveis e maiores conhecedoras dos dados; e, portanto, mais confiáveis explicadoras desses ativos e de seus impactos (figura anexo x - 14).

Importa observar duas situações encontradas na amostra que potencializam o desafio de se ter um bloco "negócios+TI" constantemente atuante ao longo de todo o ciclo de vida dos sistemas de IA: quando há a terceirização do desenvolvimento dos sistemas de IA, e quando os governos possuem uma unidade centralizadora de desenvolvimento de sistemas de IA para as demais agências e departamentos do governo.

3.7.3 *Framework* AIGov4Gov

Unindo as QCA e análise do conteúdo das entrevistas e dos documentos disponibilizados, foi construída uma visão unificada das ações e relações encontradas na presente pesquisa, quanto a modelos de governança e de gestão de organizações que produzem sistemas de IA considerando princípios éticos: o *framework* AIGov4Gov (figura 9).

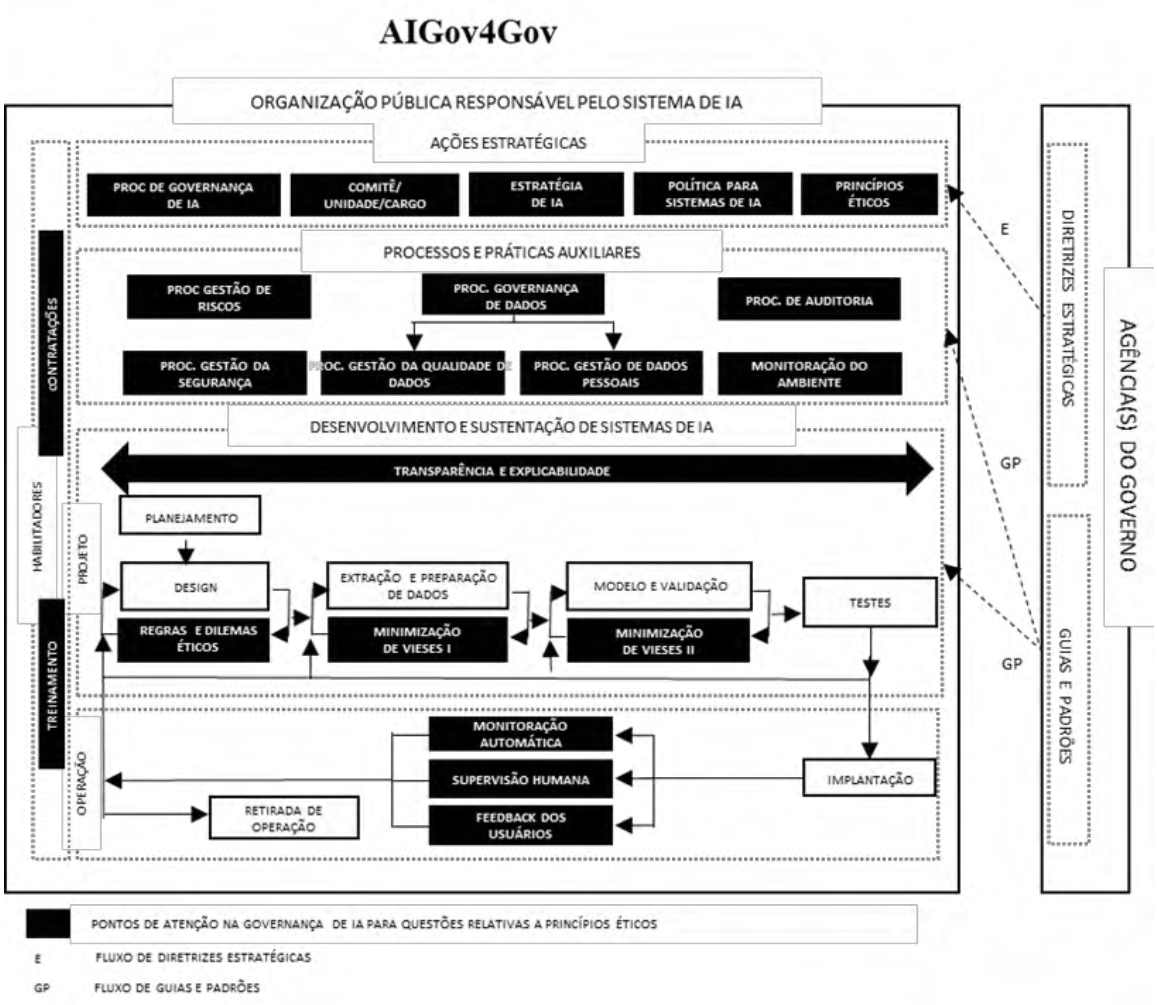

Figura 9: *Framework* AIGov4Gov – governança de IA para organizações públicas.
Fonte: Elaboração própria

O AIGov4Gov apresenta um modelo de governança de IA para uma organização pública, aplicável às três esferas de governo, Executivo, Legislativo e Judiciário; contemplando processos e práticas distribuídas nos níveis estratégico, tático e operacional. No nível estratégico, encontram-se ações dirigidas à construção e implantação de uma estratégia de IA, uma política para sistemas de IA, definição de princípios éticos aplicáveis a sistemas de IA, implantação de processo que viabilize a governança de IA, e a existência de comitê, unidade ou cargo com a atribuição da responder pela governança de IA da organização.

Processos e práticas auxiliares às diretrizes estratégicas, se interagem em diversos níveis da organização. No nível estratégico, o processo de governança de dados. No nível tático, processos para gestão de riscos, para gestão da qualidade de dados, para gestão de dados pessoais, para gestão da segurança de sistemas de IA, e para auditoria nos sistemas de IA. Adicionalmente, o AIGov4Gov prevê práticas distribuídas nos níveis estratégico

e tático, visando monitoração do ambiente interno e externo, por meio da coleta de informações sobre eventuais mudanças e tendências sociais que caracterizem alterações no contexto para o qual o sistema de IA foi projetado.

Atuando no nível tático e o operacional, o AIGov4Gov prevê práticas do processo de desenvolvimento e sustentação de sistemas de IA, com foco nos princípios éticos, distribuídas na fase de projeto e na fase em que o sistema é colocado em operação, de maneira a contemplar todo o ciclo de vida dos sistemas de IA.

Na fase de projeto, logo após o planejamento, *loops* internos caracterizam o desenvolvimento seguindo a metodologia Ágil, ocorrendo durante a especificação e desenho do problema a ser resolvido com a representação formal dos princípios éticos nas regras de negócio considerando todos os *stakeholders* do sistema de IA; na identificação de dilemas éticos, se existirem, e na definição das decisões que a organização dará para cada dilema. Práticas para minimizar vieses ocorrem em sucessivas iterações durante a fase de extração e preparação dos dados, durante a construção do modelo e validação, seguidos de testes finais, antes de serem implantados em ambiente real. E, apesar da delegação do desenvolvimento a outra organização, a *accountability* da organização pública se mantém, com a responsabilidade pela transparência de todo o ciclo de vida do sistema de IA para seus resultados sejam explicáveis.

Na fase em que o sistema está operacional, logo após a implantação, práticas conjuntamente aplicadas permitem acompanhar o sistema de IA na complexidade do ambiente real. Por meio de monitoração automática, a maioria das organizações da amostra que implantaram tal prática, tinha como objetivo, monitorar somente acurácia, mas, houve relatos de também se automatizar indicadores utilizados para identificar eventuais vieses de dados. Como observado na literatura e na amostra, a supervisão humana foi prevista para identificação de vieses cognitivos que não foram encontradas durante a fase de projeto; mas, também útil para identificar mudanças no ambiente real. E, por fim, a coleta de *feedback* de usuários.

Como habilitadores da implantação da governança de IA, o AIGov4Gov prevê a realização de treinamentos (formais ou na forma de workshops e palestras) para tomadores de decisão, desenvolvedores, usuários e auditores internos à organização, assim como a contratação do serviço de profissionais qualificados na análise de questões éticas envolvendo dados.

Também estão contemplados no AIGov4Gov, a interação entre a organização pública responsável pelo sistema de IA e, se houver, agência de sua esfera de governo

responsável por estratégia de IA, política para sistemas de IA e princípios éticos para sistemas de IA aplicáveis aos departamentos ou organizações sob sua responsabilidade. Na existência de tal situação, a organização responsável pelo sistema de IA pode adotar as diretrizes estratégicas centralizadas do seu governo, ou adaptá-las respeitando o alinhamento entre elas (fluxo E da figura 9). De maneira similar, foram previstos guias com recomendações e padrões para gestão de riscos, para governança de dados, para gestão da qualidade de dados, para gestão de dados pessoais, para gestão da segurança dos sistemas de IA, para minimização de vieses e para prover transparência em todo o processo de desenvolvimento de sistemas de IA (fluxo GP da figura 9).

A parceria entre unidade de negócio e unidade de TI novamente aparece presente nas relações com a agência centralizadora do governo para diretrizes estratégicas ou para os guias e padrões; com o propósito de dirimir dúvidas ou até mesmo fazer o registro de que esteja seguindo, parcial ou totalmente, as recomendações.

No contexto da governança de IA, no caso de se contratar o desenvolvimento externamente à organização pública, o processo de desenvolvimento e sustentação de sistemas de IA ainda requer o envolvimento da parceria “negocio+TI” (figura 10), visto que ainda se faz necessário viabilizar os processos e práticas não apenas diretamente associados ao desenvolvimento do sistema de IA, mas, também aqueles dos níveis superiores do modelo de governança de IA. No caso de se contratar tanto o desenvolvimento, quanto a sustentação, cobrindo todo o ciclo de vida do sistema de IA, a existência de gerência de projeto e gerência do produto após entrar em produção, se fazem necessárias para garantia de que as práticas e processos da governança de IA possam ser implantados.

Considerando as argumentações de Dignum (2022) e de Aasi et al. (2014) de que para desenvolver sistemas de IA, é necessário uma profundo conhecimento do contexto da organização e onde ela se insere; e tendo-se uma expectativa de que os princípios tenham algo da cultura e da estratégia da organização, um nível básico de representação das regras de negócio à luz dos princípios éticos, assim como os possíveis dilemas éticos, requerem ser realizados pela organização contratante, mesmo que uma versão mais técnica seja realizada pela organização contratada. E, considerando o fato de que a supervisão humana na busca por vieses requer uma visão profunda do negócio, como relatado na amostra, propõe-se sua realização ser pela organização contratante.

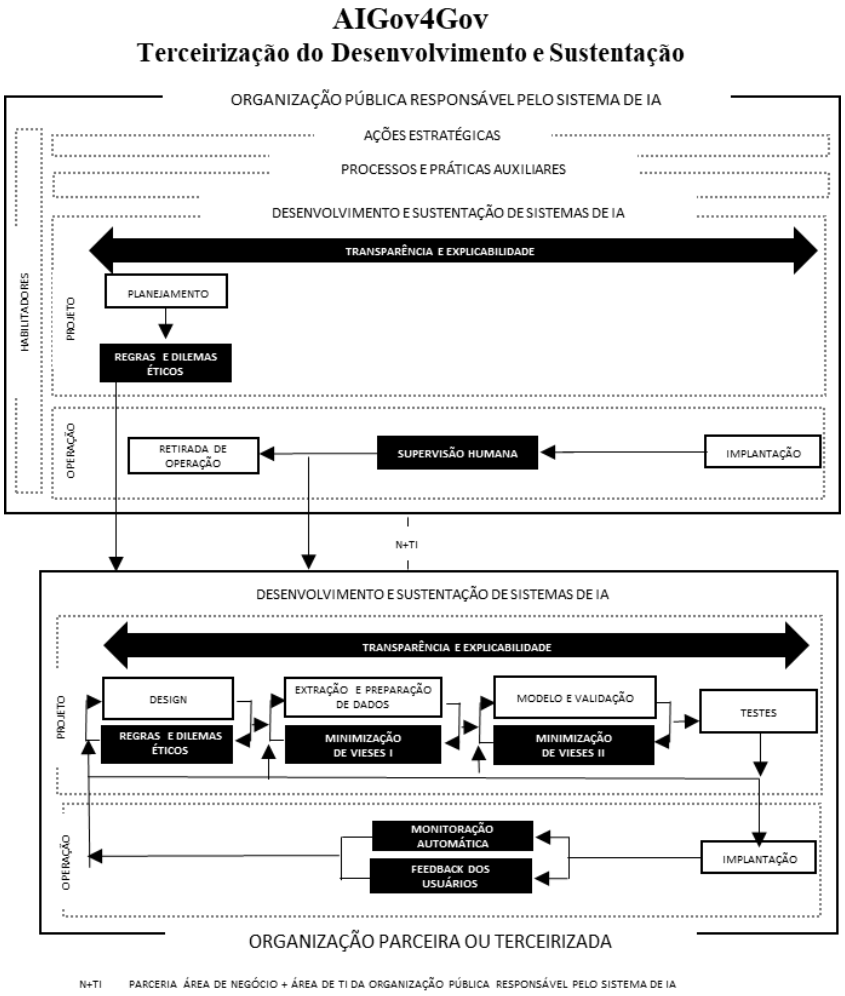

Figura 10: Framework AIGov4Gov em caso de terceirização do desenvolvimento e da sustentação do sistema de IA.
Fonte: Elaboração própria

No caso de se optar pela terceirização apenas do desenvolvimento (figura 11), alguns detalhes requerem a atenção da organização contratante. Primeiramente, no edital, propõe-se requisitos claros de como ocorrerão evoluções do sistema de IA, a partir dos resultados da monitoração automática, da supervisão humana e da coleta de *feedback*. A omissão de tais especificações pode comprometer a realização das práticas previstas na fase de operação, e até na quebra de confiança na organização pública. Adicionalmente, a comunicação dos resultados da monitoração automática, da supervisão humana e do *feedback* dos usuários requer profissionais capacitados à análise e especificação para fundamentar a demanda contratual para evolução do sistema.

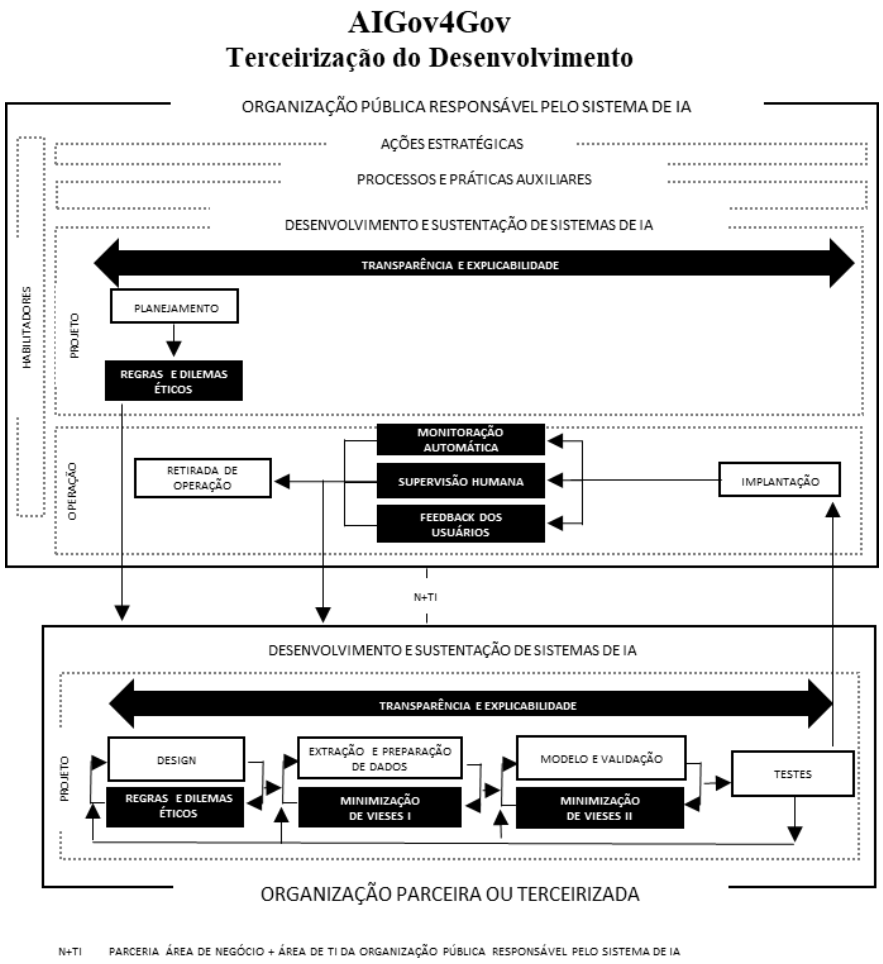

Figura 11: Framework AIGov4Gov em caso de terceirização apenas do desenvolvimento.
Fonte: Elaboração própria.

As organizações da amostra que se enquadravam tanto nos modelos apresentados nas figuras 10 e 11 viabilizaram a participação de profissionais das áreas de negócio e de TI da organização pública na coordenação e articulação com a empresa contratada. Este requisito passa a ser fundamental, visto que a execução do processo de desenvolvimento e sustentação e sistemas de IA por uma organização externa introduz complexidade aos processos auxiliares do modelo de governança, pois algumas ações e informações sobre dados (governança, qualidade e proteção de dados), riscos e segurança precisam ser repassadas à organização contratada, tempestivamente, assim como os próprios dados.

Também foi observado o treinamento de fiscais de contrato e cientistas de dados em desenvolvimento de sistemas de IA, para serem capazes de especificarem os editais e fiscalizarem tais contratos, comportamento que explica os resultados das análises das proposições 3 e 4; razão pela qual o AIGov4Gov manteve os treinamentos ainda no caso de contratações.

As parcerias entre organizações encontradas na amostra, em sua maioria, ocorreram com institutos de pesquisas e centros de tecnologia do próprio governo. Nesses casos, são órgãos especializados no desenvolvimento de sistemas de IA para si, e para as demais agências do governo. Nesses casos, também se fez necessária a atuação de gestores da TI e da área de negócio da organização demandante na articulação entre as duas organizações públicas para que a integração entre os processos e práticas de estratégicos e táticos com o processo de desenvolvimento seja viabilizada, e assim, garantida uma governança de IA.

A amostra revelou organizações cientes de que existem questões éticas a serem observadas quando se decide produzir sistemas de IA, e em movimento gradual para implantação ou adaptação de ações para tal propósito. Contudo, a maioria não possui uma clareza quanto à real dimensão do que representa o termo "governança de IA", haja vista esta expressão não ter sido mencionada por qualquer participante durante as entrevistas. Prevendo que este cenário pudesse ocorrer, a pesquisa foi beneficiada pelo fato dos instrumentos de coleta de dados terem sido construídos para captar os primeiros movimentos no sentido de boas práticas para desenvolvimento de sistemas de IA considerando princípios éticos.

Sumariamente, identificou-se na amostra estudada desafios semelhantes aos indicados pela academia, e um movimento contínuo no sentido de ajustar os seus modelos de governança e de gestão para que princípios éticos sejam atendidos.

O estudo permitiu ainda observar o potencial que agências de governos têm para promover a governança de IA ao definirem diretrizes estratégicas para todas as organizações sob sua responsabilidade, ou de recomendar padrões de boas práticas na implantação da governança de IA para todas as instâncias de governo; apesar do longo tempo necessário para implantá-las. Outro ponto observado foi a complexidade da governança e gestão pública para produção de sistemas de IA que considerem princípios éticos, quando as parcerias e contratações precisam ser estabelecidas, especialmente quando processos e práticas externos ao ato de desenvolver sistemas precisam ser envolvidos em todo o ciclo de vida de tais sistemas.

Os fatores citados reforçam as preocupações de Zuiderwijk et al. (2021) e Mäntymäki et al. (2022) quanto à necessidade de atenção à governança de IA dentro das organizações públicas, quando o tema "regular a inteligência artificial" ou "ética na inteligência artificial" é debatido; pois, é nas organizações onde devem ser materializadas as discussões realizadas no âmbito filosófico e legal sobre esses temas. Se as dificuldades

não forem vencidas tempestivamente dentro das organizações públicas, na sua missão de servir exemplarmente à sociedade, não haverá governança de IA em perspectiva nacional.

3.7.4 Limitações e agenda

Apesar do amplo e sistemático processo de construção da amostra, a presente pesquisa realizou as análises nos países e organizações que publicaram iniciativas de produção de sistemas de IA disponibilizaram os contatos, aceitaram participar da pesquisa e seguiram o rito estabelecido para ela. Logo, não correspondem às proporções em que é realizada a produção global de sistemas de IA na esfera pública, apesar dos esforços em incluir representantes de todos os países com destacada produção no ranking global.

Como o foco era captar a existência de práticas e processos, não se aprofundou em cada processo ou em modelos de maturidade desses processos. Similarmente, aspectos financeiros envolvidos nos projetos e sustentação de sistemas de IA não foram analisados para que as questões éticas possam ser tratadas adequadamente nas organizações públicas.

No contexto legislativo, além de lei dirigida a regular inteligência artificial ou tratamento de dados nos aspectos éticos e lei para proteção de dados pessoais; não foram consideradas análises de outras leis que possam ser impactadas ou ser habilitadoras da implantação da governança de IA.

Como um estudo exploratório e descritivo, a presente pesquisa abre caminhos para uma agenda de mais pesquisas que se aprofundem nos achados de cada proposição e de como ocorrem as combinações encontradas. No nível estratégico, a investigação de como a aplicação dos guias de princípios éticos recebe a influência da cultura de um governo ou de uma organização pública. No nível tático, a investigação de como tem ocorrido a gestão da inovação no serviço público de maneira a atender os princípios éticos. E, aprofundando-se no ciclo de vida de sistemas de IA, a elaboração de um modelo de maturidade para o processo de desenvolvimento e sustentação de sistemas de IA.

3.7.5. Contribuições

Ao ter atendido à lacuna destacada por Mäntymäki et al. (2022a) e Mikalef et al. (2022) sobre um conhecimento empírico de como as organizações têm interpretado e incorporado em seus processos e práticas, boas práticas para o desenvolvimento de sistemas de IA, a presente pesquisa se apresenta como inovadora enquanto conteúdo.

Ao ter utilizado *crisp-set* QCA, *fuzzy* QCA e análise de conteúdo, a pesquisa atendeu às lacunas apresentadas por Zuiderwijk et al. (2021), tanto em conteúdo apresentando práticas do setor público para a governança da IA, quanto em método que adote pesquisas empíricas contemplando combinações de abordagens exploratórias e análises qualitativa-quantitativas, a presente pesquisa se apresenta como inovadora em conteúdo e método.

Resumem-se, portanto, as seguintes contribuições aos gestores e pesquisadores:

a) Identificação de como os processos e práticas dirigidos à aplicação de princípios éticos na produção de sistemas de IA, têm sido combinados e internalizados nos modelos de governança e de gestão de organizações públicas;
b) Identificação de como habilitadores da implantação de governança da IA têm sido utilizados por organizações públicas;
c) Identificação de soluções para superar dificuldades encontradas na implantação da governança de IA em organizações públicas;
d) *Framework* para governança de IA em organizações públicas, em que processos e práticas se articulam em nível estratégico, tático e operacional na produção de sistemas de IA que considerem princípios éticos;
e) Para gestores de organizações com intenção de estabelecer parcerias com organizações públicas, o conhecimento de requisitos que possam ser exigidos por organizações públicas em contratos para o desenvolvimento e sustentação de sistemas de IA que considerem princípios éticos.

4. FORTALECIMENTO DA GOVERNANÇA NACIONAL DE INTELIGÊNCIA ARTIFICIAL PELO FOCO NAS ORGANIZAÇÕES PÚBLICAS

RESUMO

A complexidade para implantar governança de inteligência artificial (IA) tem movimentado academia, governo, indústria e a sociedade organizada. A extensa rede de *stakeholders* que se forma tanto na produção de um sistema de IA , quanto ao longo de sua vida, aumenta a dificuldade de toda articulação necessária para o êxito de tal implantação. Considerando o governo como influente *stakeholder* na governança de IA de uma nação, a presente pesquisa investigou se o robustecimento da governança de IA nacional poderia ser atingido por meio do foco na implantação da governança de IA das organizações públicas desse país. Como método, procurou-se a existência de compatibilidade entre um modelo funcional de governança de IA em âmbito nacional, e um modelo de implantação de governança de IA em organizações públicas que estavam em estágios mais avançados nesse desafio, construído a partir de pesquisa empírica de abordagem global. A compatibilidade de tais modelos foi reconhecida, o que permite que se considere a estratégia de governo para fomento e habilitação de organizações públicas a implantarem sua própria governança de IA, como uma ação para potencializar a implantação de governança de IA em uma nação.

Palavras-chave: Governança de IA, Ética na IA, IA no Governo, IA em organizações públicas, Governança Pública de IA

ABSTRACT

The complexity of implementing an artificial intelligence (AI) governance has moved academia, government, industry and the organized civil society. The extensive stakeholders' network structured both in the production of an AI system and throughout its life, increases the difficulty of all necessary articulation for such implementation success. Considering the government as an influential stakeholder in a country's AI governance, this research investigated whether strengthening AI governance nationally could be achieved by focusing on the implementation of AI governance in public organizations in a country. The method consisted of investigating the compatibility between a functional AI governance model at the national level and an AI governance model in public organizations that were at most advanced stages of this challenge, based on global empirical research. The compatibility of such models was found, which implies considering the government strategy for promoting and enabling public organizations to implement their own AI governance, as an action to enhance the establishment of AI governance in a nation.

Key words: AI Governance, Ethics on AI, AI in Government, AI in public organizations, AI Public Governance.

4.1 Introdução

A implantação de governança da inteligência artificial (IA) requer engajamento nacional e global (Fjeld et al. 2020; AI HLEG 2019b). A multidisciplinaridade exigida para enfrentamento do desafio (Bonnemais et al. 2018; Leitner & Stiefmueller 2019; Ma et al. 2018), a diversidade de percepções e interesses dos *stakeholder*s, somados à velocidade com que a tecnologia evolui, ampliam a complexidade e urgência por ações coordenadas e efetivas (Wright & Schultz 2018; Abraham et al. 2018).

Apesar de dezenas de países terem anunciado formalmente uma estratégia nacional para investimento em inteligência artificial como solução para vários problemas por eles enfrentados (tabela 3), a regulação da IA ainda é tema de muitas discussões em parlamentos (OECD 2022c). Na academia, apesar da publicação de muitos guias com princípios éticos de natureza filosófica (Morley et al. 2020b), ainda considera a governança de IA uma área carente de estudos que possam trazer, de maneira mais tangível, o que precisa ser feito (Taeihagh 2021).

Devido às longas cadeias produtivas de sistemas de IA, envolvendo vários países, e a globalização da economia, torna-se imperativa a implantação da governança de IA globalmente (Boden et al. 2017; Nevejans 2016; Jackson 2019a), de maneira a se poder rastrear a complexa composição dos produtos e serviços com IA embutida (Zuiderwijk et al. 2021; Gasser & Schmitt 2019). Importa registrar que a regulação da IA é ação que requer participação de muitos *stakeholders* do serviço público, o que gera expectativas de que nesse universo haja compreensão sobre a necessidade de mecanismos para aplicação de boas práticas do uso e desenvolvimento de sistemas de IA (Hickman 2020; Mäntymäki et al. 2022a), de maneira a se implantar governança de IA (Zuiderwijk et al. 2021).

Pelo exposto, formula-se a seguinte pergunta de pesquisa: Como impulsionar a governança de IA nacional por meio do foco na governança de IA das organizações públicas?

4.2 Governança de IA: contexto nacional e organizacional

Em pesquisa sistemática realizada na literatura especializada entre 2009 e 2020, de Almeida et al. (2021) reuniram as principais características de 21 modelos distintos que se propunham a representar práticas para uso e desenvolvimento de sistemas de IA que considerem princípios éticos. Um modelo agrupador das principais características do conjunto selecionado foi elaborado para representar a articulação de principais

stakeholders em um país no propósito de regular a IA de uma maneira contínua: o *Artificial Intelligence Regulation* – AIR.

No AIR, há vários fluxos estabelecendo as atribuições e parcerias necessárias entre os *stakeholders* a uma regulação sustentável, em uma articulação para viabilizar a governança de IA de um país. Fato que não pode passar desapercebido é de que, além das funções próprias de cada esfera de governo previstas no AIR - Executivo, Legislativo e Judiciário; cada organização pública produtora de sistemas de IA precisa criar mecanismos para que tais sistemas sejam confiáveis. Isso implica a implantação de iniciativas para viabilizar a governança de IA em cada organização pública.

A partir da percepção citada, em segunda pesquisa empírica com vinte e oito organizações públicas, distribuídas em dezessete países, que possuem sistemas de IA operacionais, identificaram-se combinações de processos e práticas apresentadas pela literatura especializada como necessários à implantação da governança de IA. O estudo resultou no *framework* AIGov4Gov como uma proposta de implantação de governança de IA em organização pública, por meio da combinação de processos e práticas distribuídas nos níveis estratégico, tático e operacional.

No nível estratégico, somam-se: estratégia de IA, política para sistemas de IA, guia de princípios éticos aplicáveis a sistemas de IA, processo para a governança de IA, e uma estrutura que possa receber as atribuições deliberativas referentes à governança de IA (comitê, cargo ou unidade). Há uma subordinação da estratégia de IA em relação à estratégia de TI, e esta, em relação à estratégia corporativa.

Com o propósito de viabilizar a implantação das diretrizes estratégicas, além de práticas voltadas ao processo de desenvolvimento de sistemas de IA, foram previstos processos e práticas auxiliares à governança de IA: processo para governança de dados, processo para gestão da qualidade de dados, processo para gestão de dados pessoais, processo para gestão de riscos em sistemas de IA, processo para auditoria em sistemas de IA, prática de monitoração de mudanças de ambiente, processo para gestão da segurança de sistemas de IA. Para a construção do sistema em si, o AIGov4Gov prevê a representação formal dos princípios e dilemas éticos nas regras de negócio, práticas para minimização de vieses e práticas para prover transparência no desenvolvimento de sistemas de IA, durante o projeto do sistema. Após ser colocado em operação, foram previstas práticas de monitoração automática, de supervisão humana e de coleta de *feedback* dos usuários.

A pesquisa empírica que fundamentou o AIGov4Gov identificou a vantagem na implantação da governança de IA por organizações públicas inseridas em governos que haviam estabelecido diretrizes estratégicas e padrões com boas práticas às organizações sob sua responsabilidade. O achado sinalizou o impacto positivo quando ocorrem ações partindo de agências de governo para estabelecer diretrizes e padrões dirigidos à governança de IA. O achado permite que se formule a seguinte **proposição**: Para fortalecer a governança de IA nacionalmente, governos usam estratégias de habilitação das organizações públicas a implantar práticas e processos que as auxiliem a produzir sistemas de IA confiáveis, conduzindo-as a implantarem sua própria governança de IA.

Procurou-se, então, analisar a proposição apresentada, usando os estudos que fundamentaram os *frameworks* AIR e o AIGov4Gov.

4.3 Método

Diante da inexistência de dados empíricos para a formulação do AIR e da existência de dados da pesquisa que construiu o AIGov4Gov; considerou-se apenas a amostra de vinte e oito organizações públicas utilizada nesta última, cujas características encontram-se nas tabelas 7 e 8.

Para verificar como a governança de IA em organizações públicas poderia potencializar a implantação da governança de um país, procurou-se investigar se o modelo AIGov4Gov, originado de combinações de práticas e processos de organizações públicas que estão em estágio mais avançado na governança de IA, pode ser considerado funcional dentro do AIR.

A verificação da compatibilidade entre os modelos AIGov4Gov e o AIR foi realizada por meio da busca de atores, processos e práticas do AIGov4Gov que se estendam ao contexto externo à organização pública responsável pelo sistema de IA. Procurou-se, então, identificar a existência de tais elementos no AIR. No sentido oposto, buscou-se identificar se os fluxos previstos no AIR para uma organização pública produtora de sistemas de IA, existem no AIGov4Gov. As verificações ocorreram considerando os documentos disponibilizados pelas organizações e governos da amostra utilizada no estudo que fundamentou o AIR. Nessa direção, a análise ocorreu em duas fases rotuladas de “fluxos do AIR de/para organizações públicas produtoras de sistemas de IA” e “fluxos do AIGov4Gov de/para o contexto externo”.

4.4 Resultados e Discussão

Os resultados de cada fase de análise são apresentados separadamente, e posteriormente, a união das análises das fases.

4.4.1 Fluxos do AIR de/para organizações públicas produtoras de sistemas de IA

O *stakeholder* "GOV INSTITUTIONS" existente no AIR foi utilizado para representar organizações públicas em um contexto amplo, haja vista o fato de que os fluxos do AIR de/para instituições "LEGILATIVE", "EXECUTIVE" e "JUDICIARY" foram aqueles estritos à missão de cada esfera de poder na regulação da IA: o Legislativo na elaboração de leis, o Executivo como agência regulatória, e o Judiciário nos julgamentos. Portanto, o único fluxo que envolve "GOV INSTITUTIONS" foi o de políticas públicas para sociedade, uma vez que não havia, até então, estudos do que ocorre dentro de uma organização pública. Contudo, houve a previsão de que, enquanto produtoras de sistemas de IA, as organizações públicas estariam sujeitas a uma parte dos fluxos desenhados para o *stakeholder* "INDUSTRIES AND SERVICES PROVIDERS"

No poder Executivo, a "AGENCY", *stakeholder* com características de uma agência regulatória para a IA, reune atribuições de uma *sandbox*, de uma entidade de certificação e de auditoria nas organizações produtoras de sistemas de IA.

Adicionalmente, em de Almeida et al. (2021), previu-se que uma organização pública produtora de sistemas de IA submeter-se-ia aos fluxos de avaliação, certificação e auditoria que a agência regulatória estabelece com qualquer organização que produza sistemas de IA, todos previstos no AIR. Portanto, para que se considere o que ocorre nas organizações públicas, basta que se crie no AIR instâncias do "GOV INSTITUTIONS" dentro do Executivo, do Legislativo e do Judiciário.

4.4.2 Fluxos do AIGov4Gov de/para o contexto externo

Há três tipos de fluxos previstos pelo AIGov4Gov de comunicação para fora da organização responsável pelo sistema de IA: a) com agências do governo que estabelecem diretrizes estratégicas; b) com agências do governo responsáveis por guias e padrões de boas práticas necessárias à governança de IA; c) com organização contratada ou parceira para desenvolvimento de sistemas de IA.

4.4.2.1 O estabelecimento de diretrizes estratégicas

A possibilidade de um governo estabelecer diretrizes estratégicas para todas as organizações públicas a ele subordinadas, foi prevista no AIGov4Gov. Compreende-se

como diretrizes, o subconjunto das ações estratégicas composto por: estratégia de IA, política ou normativos para sistemas de IA, e princípios éticos para sistemas de IA (figura 8).

As organizações da amostra cujos governos já haviam definido princípios éticos, declararam ter adotado tais princípios, poupando-lhes tempo e recursos financeiros, além de garantir um alinhamento pleno aos princípios de sua esfera de governo.

Quanto à estratégia de IA, quando existente, as organizações consideravam somente aquela realizada dentro do seu domínio, mesmo que fossem diretrizes estratégicas de IA inclusas nas suas estratégias de TI. Em apenas dois casos, as organizações consideraram como sua estratégia de IA, a mesma estratégia definida pelo seu governo. Esses últimos casos tinham em comum o fato de seus dirigentes reportarem-se diretamente a ministros de Estado ou ocuparem uma posição no comitê estratégico de sua esfera de governo.

Em relação à política dirigida a sistemas de IA, quando existente, foram identificadas, em proporção equilibrada, dois cenários: organizações que consideraram tais normativos do governo como suficientes para que pudessem atuar no uso e produção de sistemas de IA; e organizações que construíram novos normativos alinhados aos do governo, porém, com abordagem mais específica ao seu contexto (CNJ 2020; LIAA-3R 2022).

4.4.2.2. Guias e padrões de boas práticas para governança de IA

A diversidade de temas envolvidos na governança de IA é representada pela variedade de processos e práticas suscetíveis a padronização (figura 8).

A análise dos documentos disponibilizados pelos governos da amostra permitiu a identificação de diferentes abordagens de construção e disseminação de tais padrões (tabela 22). Alinhada à análise realizada durante as QCA e entrevistas que fundamentaram o AIGov4Gov, a análise de conteúdo dos documentos seguiu as mesmas categorias, tendo identificado guias para padronização de boas práticas nas seguintes categorias: “dados”, “riscos”, “vieses”, ”transparência”, “segurança”, “contratações”, “pessoas”. Acrescentaram-se as categorias “abordagens gerais” e ”abordagens de aplicações específicas”.

Tabela 22: Padrões de governo aplicáveis à governança de IA

Guias com padrões para boas práticas úteis à governança de IA
Governos da amostra

Tópico	Referência
Dados	Australian Government (2019a) Australian Government (2019b) Norwegian Data Protection Authority (2018) Government of United Kingdom (2017) Government of United Kingdom (2020a) Government of United Kingdom (2020e) Government of United Kingdom (2021a) Government of United Kingdom (2021d) Government of United Kingdom (2021f) Government of United Kingdom (2022e) Information Commissioner's Office (2020) Information Commissioner's Office (2021) Ekspertgruppen om dataetik (2018) Balahur et al. (2022)
Riscos	German Federal Ministry for Economic Affairs and Energy (2020) AI HLEG (2019b) Government of Canada (2020a) Government of Canada (2021b) Government of United Kingdom (2022a)
Vieses	Government of United Kingdom(2019b) Government of United Kingdom (2020d) Government of United Kingdom (2021a) Government of United Kingdom (2021d) Government of United Kingdom (2022c) Leslie (2019) National Institute of Advanced Industrial Science and Technology (2022)
Transparência	Norwegian Data Protection Authority (2018) Government of United Kingdom (2021a) Government of United Kingdom (2021b)
Segurança	European Union Agency for Cyber Security (2021) National Institute of Advanced Industrial Science and Technology (2022) Government of Canada (2019a) Government of Canada (2019b) Government of Canada (2021b)
Contratações	Government of United Kingdom (2020b) Government of United Kingdom (2020c) Government of Canada (2022c)
Pessoas	Leslie (2019) Australian Government (2019a)
Abordagens gerais	Australian Government (2019a) Government of Canada (2020b) Government of Canada (2021a) Government of Canada (2022a) Government of United Kingdom(2019b) Government of United Kingdom(2020f) Government of United Kingdom (2022a) Government of United Kingdom (2022d)
Reconhecimento facial	Council of Europe (2021) European Data Protection Board (2022)

Fonte: Elaboração própria.

4.4.2.2.1 Guias e padrões sobre dados

Os processos direcionados a dados, previstos pelo AIGov4Gov, correspondem a processo de governança de dados que, por sua vez, fornece diretrizes e deliberações para processo de gestão de qualidade de dados e gestão da proteção de dados pessoais.

Entre os países da amostra, os guias referentes a tais processos apresentaram as seguintes abordagens.

O governo da Austrália priorizou padrões para proteção de dados pessoais com um guia de princípios dirigidos à privacidade (Australian Government 2019b), no qual o foco é a conformidade para a lei sobre privacidade de dados. Confirmando a pesquisa que fundamentou o AIGov4Gov, a gestão qualidade de dados é abordada como necessária para a gestão dos dados pessoais para que se viabilize o cumprimento da lei para proteção de privacidade naquele país.

Na Noruega, as boas práticas para dados dirigidas à IA são apresentadas em documento emitido pela Autoridade de Proteção de Dados Pessoais do Governo, *Datatilsynet*, o que naturalmente implicou em conotação para a gestão de dados pessoais, porém, contemplando ações para gestão da qualidade dos dados (Norwegian Data Protection Authority 2018). Apesar de não ser o foco principal, o guia aborda práticas dirigidas a minimizar vieses de dados pessoais.

No Reino Unido, as padronizações para dados foram construídas em módulos ao longo do tempo, por várias agências do governo, além do Centro de Ética de Dados (United Kingdom 2019a), de maneira a compor um conjunto interrelacionado e robusto. Um *framework* para governança de dados (Government of United Kingdom 2022e) orienta, de forma articulada, gestão da qualidade de dados, gestão da proteção de dados pessoais e a ética em dados. A qualidade de dados possui *framework* específico (Government of United Kingdom 2020e), e a gestão da proteção de dados pessoais é padronizada por meio de um guia geral para organizações públicas e privadas (Government of United Kingdom 2021d), e um questionário para avaliação de impacto na proteção de dados (Government of United Kingdom 2021f), em alinhamento à conformidade com a lei de proteção de dados pessoais. A dependência dos resultados dos sistemas de IA em relação aos dados foi representada pelo *framework* que integra ética, transparência e *accountability* para sistemas de IA, com integrações aos guias anteriores (Government of United Kingdom (2021a); e, na mesma direção, um framework específico para ética em dados (Government of United Kingdom 2020a). O guia de compartilhamento de dados em nuvens (Government of United Kingdom 2017) e o guia

para movimentação de dados internacionais (Information Commissioner's Office 2021) são apontados em vários padrões citados.

Patrocinados pela Comissão Europeia, Balahur et al. (2022) elaboraram guia para qualidade de dados com foco na padronização para minimizar vieses. Vários padrões ISO foram incorporados pelo documento.

Na Dinamarca, o governo lançou o guia para ética em dados (Ekspertgruppen om dataetik 2018) com abordagens em governança de dados, qualidade de dados, e proteção de dados pessoais, permitindo uma preparação para os ajustes na legislação que exige política de ética em dados para grandes organizações, e aplicada ao governo (Danish Government 2020).

4.4.2.2.2. Guias e padrões sobre riscos

A abrangência da aplicação de gestão de riscos é muito ampla. Diretamente aplicável à gestão de riscos em sistemas de IA, foram identificados na amostra padrões sobre categorias de riscos e avaliação de riscos.

O governo alemão , por meio de seu Ministério da Economia e Energia, realizou levantamento sobre os padrões internacionais que poderiam ser adotados e/ou adaptados para a produção de sistemas de IA na Alemanha, de maneira a atender a boas práticas, inclusive as questões éticas. Um modelo de categorização de riscos (German Federal Ministry for Economic Affairs and Energy 2020) foi apresentado, de maneira sintonizada com o entendimento da União Europeia (European Commission 2021a), como basilar para a escolha das ações de mitigação, sendo, inclusive apresentada como critério para aplicabilidade de lei para IA.

A avaliação de riscos que projetos de sistemas de IA possam apresentar às pessoas individualmente, à sociedade, ao governo e ao meio ambiente, tem sido uma peça-chave para o Canadá, o Reino Unido e a União Europeia. São formulários utilizados para identificar o nível de risco e as ações da política, normativos ou legislação a serem aplicados a cada caso.

No caso do Reino Unido (Government of United Kingdom 2022a), a avalição envolve, entre vários temas, o propósito do projeto, arquitetura que explique o funcionamento da IA, origem e como ocorre a coleta de dados, dados pessoais, dados de pacientes da área médica, dados de crianças ou adultos em situação de vulnerabilidade. Analisam-se o alcance do impacto na população; o tipo de potencial ameaça e vieses; segurança de dados. confidencialidade, controle de qualidade; uso de padrões que

permitam a gestão de dados, a reprodução dos procedimentos e auditoria; a existência de supervisão humana; e o tipo de transparência dada aos resultados, aos dados, e aos métodos e ferramentas utilizadas.

No caso do Canadá (Government of Canada 2020a), questiona-se a classificação da informação (Government of Canada 2021b), propósito dos sistemas; *stakeholders* envolvidos; se o sistema irá tomar decisão que será feita por um ser humano; impactos ao meio ambiente; quem coletou os dados para treinamento; quem coletou os dados para o uso do sistema; se há governança para os dados, estratégia e planejamento; segurança de dados; proteção de dados pessoais; se o sistema é de auxílio à decisão ou se fará sozinho a decisão; documentação dos testes para minimização de vieses; trilhas de auditoria; existência de monitoração, supervisão humana, coleta de *feedback* dos usuários.

O modelo de *framework* para IA confiável da Comissão Europeia é fundamentado em classes de riscos, que se associam às correspondentes recomendações (AI-HLEG 2019b).

4.4.2.2.3. Guias e padrões sobre vieses

Frequentemente presente em textos dirigidos à ética na IA, guias com orientações diretas sobre como minimizar a ocorrência de vieses foram encontrados em guias dos governos do Reino Unido e do Japão.

Sob a denominação de "gestão da qualidade de *machine learning"*, o governo japonês, por meio do Instituto de Ciência e Tecnologia Avançada, apresenta de forma aprofundada, técnicas para identificar e minimizar vários tipos de vieses (National Institute of Advanced Industrial Science and Technology 2022).

No Reino Unido, proposta de padronização ocorreu em módulos, envolvendo o Centro de Ética em Dados (Government of United Kingdom 2019a) e outras agências do governo britânico. Iniciou-se com conceituação e forma de identificação de vieses (Government of United Kingdom 2020d); e avançou-se para um modelo integrado e mais abrangente sobre ética na IA (Government of United Kingdom 2021a), incluindo a participação do Instituto Alan Turing (Leslie 2019) de maneira a apresentar aos servidores públicos diversos tipos de aplicações em que a existência de viés tem alta probabilidade de ocorrer, com instruções de como identificá-los. Destaca-se a preocupação em apresentra atributos comuns entre legislação sobre equidade e sobre proteção de dados pessoais, assim como atributos que só dizem respeito a uma abordagem e não à outra (Government of United Kingdom 2021d).

4.4.2.2.4 Guias e padrões sobre transparência

Compondo um princípio ético básico para enfrentar a opacidade algorítmica (Tutt 2017; Butterworth 2018; Buiten 2019), guias para prover transparência ao processo de desenvolvimento de sistemas e IA foram identificados nos governos da Noruega e do Reino Unido.

O governo da Noruega enfatiza que a proteção de dados pessoais não exclui a necessidade de transparência dos sistemas de IA, e apresenta algumas técnicas de como viabilizá-la (Norwegian Data Protection Authority 2018). Com o foco no auxílio às organizações públicas sob sua responsabilidade, o Reino Unido mantém um conjunto de guias que abordam desde instruções de transparência e *accountability* a serem observadas em um projeto de IA (Government of United Kingdom 2021a), um checklist do que precisa ser documentado durante o processo de desenvolvimento de sistemas de IA, até um conjunto de padrões de transparência algorítmica (Government of United Kingdom 2021b).

4.4.2.2.5 Guias e padrões sobre segurança

A disponibilização de guias sobre segurança dos sistemas de IA foi identificada na União Europeia, Japão e Canadá.

Em uma extensa abordagem de prevenção quanto a ataques cibernéticos, a União Europeia apresenta em seu guia, tipos de vulnerabilidades e ataques específicos a sistemas de IA para cada fase do ciclo de vida desses sistemas. E, para cada caso, ações são recomendadas (European Union Agency for Cyber Security 2021).

O guia do governo japonês aborda segurança em três aspectos. Na perspectiva da privacidade dos dados; na relação entre falhas nos sistemas e problemas de segurança física daqueles envolvidos direta ou indiretamente com o produto de IA embutida; e, na relação entre dados maliciosamente enviesados e ataques cibernéticos (National Institute of Advanced Industrial Science and Technology 2022).

O governo canadense, em um modelo modular de diretrizes para sistemas de IA, direciona ao guia de gestão de identidade digital (Government of Canada 2019a) e de gestão de segurança da informação (Government of Canada 2019b), com o uso de padrão para classificação do nível de segurança da informação (Government of Canada 2021b).

4.4.2.2.6 Guias e padrões sobre contratações

Em uma abordagem de destaque, o AIGov4Gov previu a existência de contratação de organizações, ou o estabelecimento de parcerias para o desenvolvimento e sustentação de sistemas de IA. Guias foram encontrados no governo britânico e no governo canadense para garantir que os princípios éticos sejam considerados, apesar da complexidade que ele introduz numa governança de IA (Hickok 2022).

No Reino Unido, a orientação inicia desde a previsão da contratação na estratégia de IA, a avaliação da capacidade dos dados de servir aos propósitos do sistema planejado, a existência de requisitos para garantir um nível de explicabilidade dos resultados do sistema, de maneira a considerar ações em todo o ciclo de vida dos sistemas de IA (Government of United Kingdom 2020b; Government of United Kingdom 2020c). São orientações derivadas de parceria estabelecida com o World Economic Forum, fundamentada na avaliação de riscos da contratação. Entre os temas abordados na avaliação, podem ser citados: estratégia de IA, estratégia digital, políticas para uso da IA, políticas de inovação, políticas de dados, políticas de tecnologias, políticas de acesso à informação, políticas para proteção de dados pessoais, processo para gestão de dados, processo que garanta a identificação e critérios para atribuição da propriedade dos dados, identificação e planos para lidar com problemas de qualidade de dados, práticas para lidar com vieses em dados, critérios para formação de equipes de projeto com perfil diversificado, nível de autonomia esperado ao sistema, estratégias de transferência de conhecimento da tecnologia, e práticas para garantir a segurança cibernética (World Economic Forum 2020a; World Economic Forum 2020b; World Economic Forum 2020c).

A proposta do guia canadense baseia-se em avaliação prévia de empresas que, se aprovadas, compõem uma lista de autorizadas a concorrer em editais para sistemas de IA do governo. Acompanha-se publicamente as avaliações e evolução da situação das empresas da lista (Government of Canada 2022c).

4.4.2.2.7 Guias e padrões sobre capacitação e perfil de pessoas

Prevista no AIGov4Gov como habilitadora da implantação da governança de IA, a capacitação de pessoas foi identificada em guias do governo britânico e australiano. No primeiro caso, por meio do Instituto Alan Turing (Leslie 2019), a proposta apresenta as competências necessárias para as atribuições previstas para a produção de sistemas IA confiáveis. No caso australiano, as orientações são dirigidas aos perfis necessários em

todas os campos de conhecimento envolvidos na produção responsável de sistemas de IA (Australian Government 2019a).

4.4.2.2.8 Guias de abordagens gerais

A multiplicidade de temas envolvidos nos níveis organizacionais para a governança de IA tem requerido guias gerais, apresentando conceitos e problemas que precisam ser evitados. Documentos de orientação geral para a ética na IA foram identificados em vários governos em uma demonstração da necessidade de conscientização e engajamento de múltiplos *stakeholders* (Government of United Kingdom2019b; Government of United Kingdom 2020f; Government of United Kingdom 2022a; Australian Government 2019a; Government of Canada 2022a).

A existência de padrões para criação de serviços digitais em alguns governos tem sido referenciada pelos padrões próprios de sistemas de IA (Government of Canada 2020b; Government of United Kingdom 2022d)

4.4.2.2.9 Abordagens de aplicações específicas

A Comissão Europeia (European Data Protection Board 2022) e o continente europeu (Council of Europe 2021) elaboraram padrões para uso e desenvolvimento ético de sistemas e IA para reconhecimento facial.

4.4.2.3 Interações com organização contratada ou parceira para desenvolvimento de sistemas de IA

As variações do AIGov4Gov para contratações, com organizações públicas ou privadas, para o desenvolvimento e sustentação de sistemas de IA, criam a necessidade de inserção no AIR, de fluxos entre a organização pública responsável pelo sistema de IA, e os *stakeholders* “INDUSTRIES AND SERVICE PROVIDERS” e “ACADEMIA”.

4.4.3 Compatibilização dos modelos

A união das análises permitiu a identificação de alguns fluxos implícitos no AIR para o AIGov4Gov, uma vez que o AIR não foi construído de maneira a considerar o que ocorre dentro das organizações públicas. Similarmente, era implícita a existência de organizações públicas produtoras de sistemas de IA em cada uma das esferas de governo previstas no AIR. Em consequência, não eram visíveis agências do governo para estabelecer diretrizes estratégicas e/ou para construir guias e padrões de boas práticas à

governança de IA em cada esfera de governo. Tais agências não devem ser confundidas com a Agência Regulatória prevista no AIR.

Adicionalmente, em cada instância de organização pública, devem existir fluxos necessários à troca de informações decorrentes de contratos e parcerias com outras organizações para o desenvolvimento e sustentação de sistemas de IA, sejam privadas ou outras organizações públicas.

Pelo exposto, a visão do AIR integrada ao AIGov4Gov requer que sejam explicitadas as relações descritas em um modelo próprio para que se possa compreender como a governança de IA de uma organização pública pode ser integrado à governança de IA de um país. A essa visão específica da integração relatada, denominou-se de AIGov4Gov no AIR, cujo desenho é apresentado na figura 12.

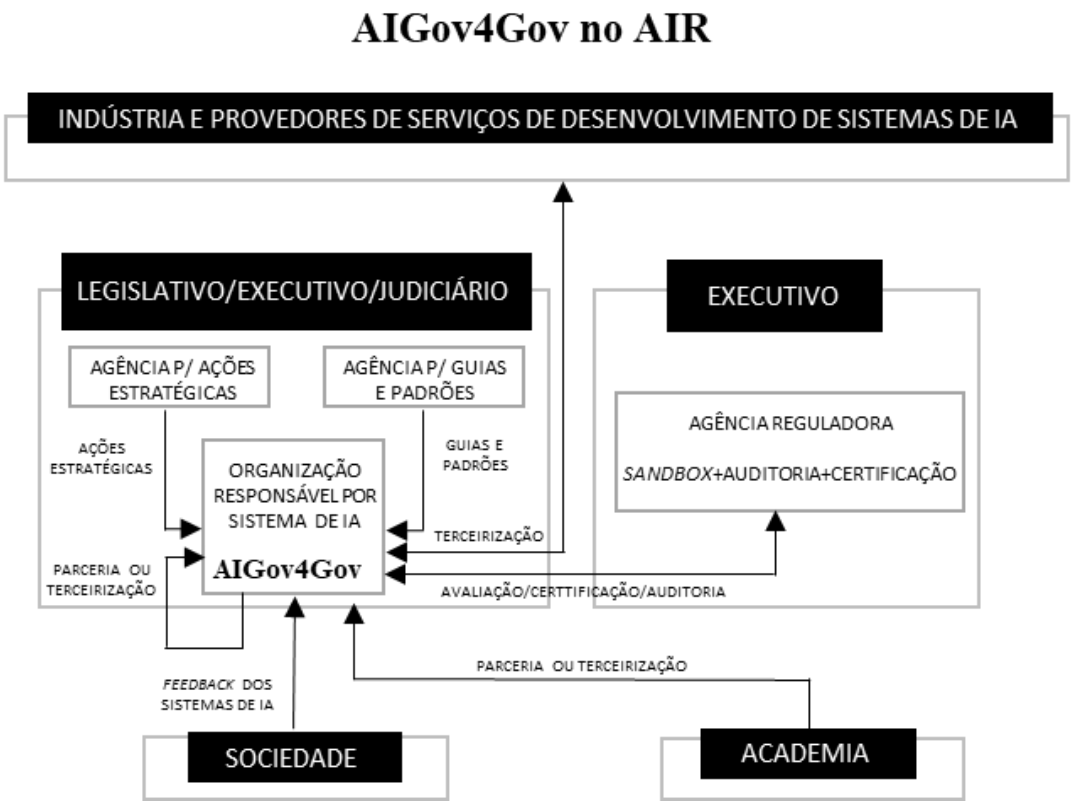

Figura 12: Localização do *framework* AIGov4Gov no *framework* AIR.
Fonte: Elaboração própria.

Na amostra utilizada para a pesquisa que fundamentou o AIGov4Gov, foram identificados casos em que a agência geradora de diretrizes estratégicas (estratégia, política e princípios éticos) é a mesma que cria os padrões. Contudo, faz-se importante sua distinção, como identificado em casos onde o nível de padronização é mais avançado requerendo órgãos com atribuições e especialidades em cada tema.

A Agência Reguladora corresponde à “AGENCY” prevista no AIR, e permanece com todos os fluxos previstos para todos os *stakeholders* daquele *framework*. Apenas para efeito de simplificação, tais fluxos foram omitidos da figura 12, para que pudesse ser evidenciada a integração do que ocorre dentro de uma organização pública genérica e os *stakeholders* diretamente relacionados a ela.

4.5 Conclusões

As verificações de compatibilização entre o modelo de Governança de IA em um país, AIR (de Almeida et al. 2021), e o modelo de governança de IA para uma organização pública, AIGov4Gov, indicaram que tais modelos são integráveis, portanto, compatíveis.

A confirmação permite que se gere a expectativa de que , uma vez implantados processos e práticas aderentes ao AIGov4Gov nas organizações públicas, seja possível integrar o modelo nacional de governança de IA aos correspondentes modelos dessas organizações governamentais, poupando tempo e recursos humanos; e, em consequência, potencializando os resultados da governança de IA de um país.

4.6 Limitações, contribuições e agenda

O presente estudo utilizou a mesma amostra que fundamentou o AIGov4Gov, composta de vinte e oito organizações públicas. Amostras maiores poderiam, em teoria, promover resultados mais próximos da realidade.

Pesquisadores e gestores públicos se beneficiam da pesquisa pela apresentação de proposta de integração de cenários complexos de implantação de governança de IA. A proposta permite que se avance em *soft laws*, processos e práticas para a Governança de IA; enquanto tratativas mais longas, envolvendo instâncias políticas e variáveis nacionais e internacionais são ajustadas. E, após adequação da legislação, a integração dos modelos favorece o robustecimento de ações sustentáveis e contínuas para produção de sistemas de IA considerando princípios éticos.

Propõem-se pesquisas que permitam investigar como a sociedade está percebendo e reagindo aos movimentos dos demais *stakeholders* em torno da governança de IA nacionalmente e globalmente.

REFERÊNCIAS

Aaronson, S. A., & Leblond, P. (2018). Another Digital Divide: The Rise of Data Realms and its Implications for the WTO. *Journal of International Economic Law*, *21*(2), 245–272. https://doi.org/10.1093/jiel/jgy019

Aasi, P., Rusu, L., & Han, S. (2014). The Influence of Culture on IT Governance: A Literature Review. *47th Hawaii International Conference on System Sciences*, 4436-4445. doi: 10.1109/HICSS.2014.546

Abdollahi, B., & Nasraoui, O. (2018). Transparency in Fair Machine Learning: the Case of Explainable Recommender Systems. In: Zhou, J., Chen, F. (eds) *Human and Machine Learning. Human–Computer Interaction Series*. Springer, Cham. https://doi.org/10.1007/978-3-319-90403-0_2

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*(July), 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Aburachid, L.M.C. & Greco, P.J. (2011). Validação de conteúdo de cenas do teste de conhecimento tático no tênis. *Estudos de Psicologia*. Campinas, 28(2); 261-267.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access Review*, 6, 52138-52160

African Union (2018). The Digital Transformation Strategy for Africa. Disponível em https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf Acessado em 5 de novembro de 2022.

Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019, June). Protecting World Leaders Against Deep Fakes. In CVPR workshops, l(1).

Ahn, M. J. & Chen, Y. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, Volume 39, Issue 2. https://doi.org/10.1016/j.giq.2021.101664.

AI HLEG (2019a). A Definition of AI: Main Capabilities and Disciplines. Definition developed for the purpose of the AI HLEG's deliverables. *European Comission*. Disponível em https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines Acessado em 5 de novembro de 2022.

AI HLEG (2019b). Ethics Guidelines for Trustworthy AI. *European Comission*. Disponível em https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Acessado em 5 de novembro de 2022.

AI4People (2018). Ethical framework for a good society: opportunities, risks, principles, and recommendations. *Atomium – European Institute for Science, Media and*

Democracy. Disponível em http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf . Acessado em 5 de novembro de 2022.

Aiken, C. (2021). Classifying AI Systems. Center for Security and Emerging Technology. Georgetown University. Disponível em https://cset.georgetown.edu/publication/classifying-ai-systems/ Acessado em 5 de novembro de 2022.

Alshahrani, A., Dennehy, & D., Mäntymäki, M. (2021). An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Government Information Quarterly*, Volume 39, Issue 4. https://doi.org/10.1016/j.giq.2021.101617

Amigoni F., & Schiaffonati, V. (2018). Ethics for Robots as Experimental Technologies. *IEEE Robotics & Automation Magazine*, March 25, 30-36. doi: 10.1109/MRA.2017.2781543.

Amsterdam Data Science (2018). Ained – National AI Strategy. Disponível em https://amsterdamdatascience.nl/news/ained-a-national-ai-strategy-for-the-netherlands-is-published/. Acessada em 5 de novembro de 2022.

Anderson M., & Anderson S.L. (2018) Geneth: a general ethical dilemma analyzer. De Gruiter. Paladyn, J. Behav. Robot. (9) 337–357. https://doi.org/10.1515/pjbr-2018-0024

Andrews, L. (2018). Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'. *Public Administration*, *97*(2), 296–310. https://doi.org/10.1111/padm.12534

Arrieta, A.B., Díaz-Rodríguez, N., Ser, J.D., Bennetot, A., Tabik, S, Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R, Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. (58), 82-115. https://doi.org/10.1016/j.inffus.2019.12.012.

Ashaye, O. R., & Irani, Z. (2019). The role of stakeholders in the effective use of e-government resources in public services, International Journal of Information Management, (49), 253-270. https://doi.org/10.1016/j.ijinfomgt.2019.05.016.

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for Artificial Intelligence and Digital technologies. *International Journal of Information management*, (62). https://doi.org/10.1016/j.ijinfomgt.2021.102433

Ashokan, A., & Haas, C. (2021). Fairness metrics and bias mitigation strategies for rating predictions, Information Processing & Management, Volume 58, Issue 5, 102646, ISSN 0306-4573, https://doi.org/10.1016/j.ipm.2021.102646.

Aurora AI (2019). Aurora AI - Towards a human Centric Society. Disponível em https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf Acessado em 5 de novembro de 2022;

Australian Government (2019a). Artificial Intelligence – Australia's Ethics Framework. *Department of Industry, Innovation and Science*. Disponível em https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework Acessado em 5 de dezembro de 2022.

Australian Government (2019b). Australian Privacy Principles Guidelines. *Offiice of Australian Information Commissioner.* Disponível em https://www.oaic.gov.au/__data/assets/pdf_file/0009/1125/app-guidelines-july-2019.pdf Acessível em 05 dezembro 2022.

Australian Government (2021). Australian's Artificial Intelligence Action Plan. Disponível em https://www.industry.gov.au/publications/australias-artificial-intelligence-action-plan. Acessado em 5 de dezembro de 2022.

Awad, E., Dsouza, S., Bonnefon, J. F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing moral machines. Communications of the ACM, 63(3), 48-55.

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, *61*(6), pp. 54–61. http://dx.doi.org/10.1145/3209581.

Balahur, A., Jenet, A., Torres, I., Charisi, V., Ganesh , A., Griesinger, C.B., Maurer, P., Mian, L., Salvi, M., Scalzo, S., Sol er Garrido, J., Taucer, F. & Tolan, S. (2022). Data quality requirements for inclusive, non-biased and trustworthy AI. Putting-Science-Into-Standards. *Publications Office of the European Union.* Luxembourg. doi:10.2760/365479, JRC131097

Baterman, W., & Powles, J. (2021) Legal Audit of AI in the Public Sector. Hmi Technology Policy Paper. Disponível em https://static1.squarespace.com/static/5c105643ec4eb7d1a8c68c9c/t/628ae3c3e1a7ce73c7d7cb01/1653269450230/Legal+Audit+of+AI+in+the+Public+Sector.pdf. Acessado em 5 de novembro de 2022.

Baumane-Vitolina, I., Cals I., & Sumilo E. (2016) Is Ethics Rational? Teleological, Deontological and Virtue Ethics Theories Reconciled in the Context of Traditional Economic Decision Making, *Procedia Economics and Finance*, (39), pp. 108-114, ISSN 2212-5671, https://doi.org/10.1016/S2212-5671(16)30249-0.

Beltran, N. (2020). Artificial Intelligence in Lethal Automated Weapon Systems-What's the Problem?: Analysing the framing of LAWS in the EU ethics guidelines for trustworthy AI, the European Parliament Resolution on autonomous weapon systems and the CCW GGE guiding principles.

Bench-Capon, T., & Modgil, S. (2017). Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence & Law Review*, 25, 29-64.

Bendszus, R. (2022). 'Make-or-Buy'in the Era of Artificial Intelligence: Insights from AI Use Cases in the German Public Sector (Doctoral dissertation, Hertie School, Berlin).

Benfeldt, O.; Persson, J.S.; & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Inf Syst Front* 22, 299–313. https://doi.org/10.1007/s10796-019-09923-z

Benjamins, V.R. & García I. S. (2020). Towards a framework for understanding societal and ethical implications of Artificial Intelligence. *Vulnerabilidad y cultura digital* by Dykinson. 87-98, 2020

Besio, C., & Pronzini, A. (2014). Morality, ethics, and values outside and inside organizations: An example of the discourse on climate change. Journal of Business Ethics, 119, 287-300.

Betarelli-Júnior, A.A. & Ferreira, S.F. (2018). Introdução à Análise Qualitativa Comparativa e aos Conjuntos Fuzzi (FSQCA). Brasilia. Enap.

Black, J. (2002) Critical reflections on regulation. *Aust. J. Legal Philos.* 27, pp. 1–35. Disponível em http://www.austlii.edu.au/au/journals/AUJlLegPhil/2002/1.pdf. Acessado em 5 de novembro de 2022.

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B. & Winfield, A. (2017). Principles of robotics: regulating robots in the real world. *Connection Science*, 29:2, 124-129.

Bogucki, A., Engler A., Perarnaud, C., & Renda, A. (2022) The AI Act and Emerging EU Digital Acquis. CEPS in depth analysis. September 2022. Disponível em file:///C:/Users/patri/Downloads/CEPS-In-depth-analysis-2022-02_The-AI-Act-and-emerging-EU-digital-acquis%20(2).pdf Acessado em 24 de fevereiro de 2023.

Bonnemains, V., Saurel, C. & Tessier, C. (2018). Embedded ethics: some technical and ethical challenges. *Ethics Information Technology*. (20), 41.

Bonsón, E., Lavorato, D., Lamboglia, R., & Mancini,D. (2021). Artificial intelligence activities and ethical approaches in leading listed companies in the European Union. *International Journal of Accounting Information Systems*. (43) 1467-0895. https://doi.org/10.1016/j.accinf.2021.100535.

Borgesius, F. Z. (2018). Discrimination, artificial intelligence, and algorithmic decision-making. *Directorate General of Democracy. Council of Europe.*

Bovens, L. (2009). The Ethics of Nudge. In: Grüne-Yanoff, T., Hansson, S.O. in Preference Change. *Theory and Decision Library*, (42). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-2593-7_10

Boyd, R. & Holton, R.J. (2018). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation? *Journal of Sociology*. 54 (3), 331-345. https://doi.org/10.1177/1440783317726591

Breier, J., Baldwain, A., Balinsky, & Liu, Y. (2020). Risk Management for Machine Learning Security. arXiv:2012.04884v1 [cs.CR]

BSA (2021). Report to NIST on developing an Artificial Intelligence Risk Management. https://www.bsa.org/files/policy-filings/09142021nistairmf.pdf Acessado em 13 de junho 2022.

Buiten, C. M. (2019). Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation, 10*(1), 41-59.

Butterworth, M. (2018). The ICO and artificial intelligence: The role of fairness in the GDPR framework. *Computer Law & Security Review*, 34, 257-268.

Bussola Tech (2022) https://bussola-tech.co/ Acessado em 31 de outubro de 2022.

Bynum, T.W. (2006). Flourishing Ethics. *Ethics Inf Technol* 8**,** 157–173. https://doi.org/10.1007/s10676-006-9107-1

C4IR Brasil (2022). Guia de Contratações Públicas de Inteligência Artificial. Centro para a 4ª Revolução Industrial. Disponível em https://ideiagov.sp.gov.br/guia-de-contratacoes-publicas-de-inteligencia-artificial/ Acessado em 11 de novembro de 2022.

California General Assembly (2020). AB-2261 Facial recognition technology. Disponível em https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB2261 Acessado em 29 de dezembro de 2022.

Calzada, I., & Almirall, E. (2020). Data ecosystems for protecting European citizens' digital rights. *Transforming Government: People, Process and Policy*, 14(2), 133–147. https://doi.org/10.1108/TG-03-2020-0047

Campitelli, G. & Gobet, F. (2010). Herbert Simon's Decision-Making Approach: Investigation of Cognitive Processes in Experts. Review of General Psychology. 2010;14(4):354-364. doi:10.1037/a0021256

Carretero, A., Gualo, F., Caballero, I., Piattini, M. (2017). MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. *Elsevier BV*, *54*, 139–151. https://doi.org/10.1016/j.csi.2016.11.008

Carter D. (2020). Regulation and ethics in artificial intelligence and machine learning technologies: Where are we now? Who is responsible? Can the information professional play a role? *Business Information Review*. 2020;37(2), 60-68.

Cave, S., Nyrup, R., Vold, K., & Weller, A. (2019). "Motivations and Risks of Machine Ethics," in *Proceedings of the IEEE*, vol. 107, no. 3, 562-574.

Cerka P., Grigiene, J., & Sirbikite, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 31(3), 376-389.

Cerka, P., Grigiene, J. & Sirbikyte, G. (2017). Is it possible to grant legal personality to artificial intelligence software systems? *Computer Law & Security Review*, 33(5), 685-699.

Chen, S. (2021). Regulating autonomous vehicles: Liability paradigms and value Choices. AI, Data and Private Law. 147-172. Research Collection School Of Law.Available at: https://ink.library.smu.edu.sg/sol_research/3403

Chen, T., Liu, J., Xiang, Y., Niu, W., Tong, E., & Han, Z. (2019). Adversarial attack and defense in reinforcement learning-from AI security view. Cybersecurity, 2, 1-22.

Chen, X. & Deng, Y. (2022). An Evidential Software Risk Evaluation Model. *Mathematics*, 10, 2325. https://doi.org/10.3390/ math10132325

China Academy of Information and Communication Technology (2021) White Paper on Trustworthy Artificial Intelligence. Disponível em https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/ Acessado em 24 de fevereiro de 2022.

Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., ... & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing, 101, 107050.

Chmielewski, P. (2018). Ethical Autonomous Weapons?: Practical, Required Functions; *IEEE Technology and Society Magazine*, 37(3), 48-55. doi: 10.1109/MTS.2018.2857601.

Cihon p., Maas M.M., & Kempo L. (2020) Should Artificial Intelligence Governance be centralized?: Design Lessons from History. AAAI/ACM Conference on AI, Ethics, and Society (AIES '20). Association for Computing Machinery, New York, NY, USA, 228–234. https://doi.org/10.1145/3375627.3375857

Codá, R.C, Farias, J. S., & Dias, C. (2022): Interactive Value Formation and Lessons Learned from Covid-19: The Brazilian Case, Journal of Quality Assurance in Hospitality & Tourism, DOI: 10.1080/1528008X.2022.2135057

Cogliaanese C. (2020). Evironmental soft law as a Governance Srategy. Coglianese, Cary, Environmental Soft Law as a Governance Strategy (2020). *Jurimetrics*, 61, p. 19, U of Penn Law School, Public Law Research Paper No. 21-05. https://ssrn.com/abstract=3775088

Copeland, B.J. (2000) The Turing Test. *Minds and Machines*, 10, 519–539. https://doi.org/10.1023/A:1011285919106

Commission de Surveillance du Secteur Financier (2018). Artificial Intelligence: opportunities, risks and recommendations for the financial sector. Luxembourg. https://www.cssf.lu/en/Document/white-paper-artificial-intelligence-opportunities-risks-and-recommendations-for-the-financial-sector/ Acessado em 13 de abril 2022.

Conitzer, V., Sinnott-Armstrong, W., Borg, JS, Deng, Y., & Kramer, M. (2017). Moral Decision Making for Artificial Intelligence. *AAAI Publication, 31º Conference on Artificial Intelligence*.

Conselho Nacional de Justiça (2020). Resolução Nº 332 de 21/08/2020 – Disponível em https://atos.cnj.jus.br/atos/detalhar/3429 Acessado em 5 de dezembro de 2022.

Cornforth, C. (2003). The governance of public and non-profit organisations: What do boards do? London: Routledge. http://oro.open.ac.uk/15872/

Correia, A., & Água, P. B. (2021). A corporate governance perspective on IT governance. In S. Hundal, A. Kostyuk, & D. Govorun (Eds.), Corporate governance: A search for emerging trends in the pandemic times (pp. 107–114). https://doi.org/10.22495/cgsetpt19

Council of Europe (2018). European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment. *European Commission for the Efficiency of Justice (CEPEJ*), p. 12. Disponível em https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c Acessado em 23 de fevereiro de 2023.

Council of Europe; European Commission, European Union Agency for Fundamental Rights, InterAmerican Development Bank; Organisation for Economic Co-operation and Development, United Nations, Unesco, World Bank Group (2021). GlobalPolicy.AI. Disponível em https://globalpolicy.ai/en/ Acessado em 13 de abril 2022

Council of Europe (2021). Guidelines on facial recognition. Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data. Convention 108. Disponível em https://rm.coe.int/guidelines-facial-recognition-web-a5-2750-3427-6868-1/1680a31751 Acessado em 20 de dezembro de 2022.

Danish Government (2019). National Strategy for Artificial Intelligence. *Ministry of Finance and Ministry of Industry, Business and Financial Affairs*. Disponível em https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/ Acessado em 25 de fevereiro de 2023.

Danish Government (2020). Lov om ændring af årsregnskabsloven. (Krav om rapportering af dataetik) Disponível em https://www.retsinformation.dk/eli/lta/2020/741 Acessado em 7 de dezembro de 2022.

Das A. (2020). Opportunities and Challenges in Explainable Artificial Intelligence 9XAI): A Survey. arXiv:2006.11371 [cs.CV] https://doi.org/10.48550/arXiv.2006.11371

Dazeley, R., Vamplew, P., Foale, C., Young, C., Aryal, S., Cruz, F. (2021). Levels of explainable artificial intelligence for human-aligned conversational explanations. *Artificial Intelligence*, 299, 103525, ISSN 0004-3702. https://doi.org/10.1016/j.artint.2021.103525.

de Almeida, P.G.R., dos Santos, C.D., & Farias, J.S. (2021) Artificial Intelligence Regulation: a framework for governance. *Ethics Inf Technol* 23**,** 505–525. https://doi.org/10.1007/s10676-021-09593-z

De Haes, S., & Van Grembergen, W. (2004). IT governance and its mechanisms. Information systems control journal, 1, 27-33.

de Oliveira, T. F. (2019). Avaliação das Práticas de Auditoria Interna da Secretaria Federal de Controle Interno da CGU sob a Ótica da Auditoria Baseada em Riscos. *Controladoria Geral da União*, Brasil. Disponível em https://revista.cgu.gov.br/Revista_da_CGU/article/view/73/pdf_60.

De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. Patterns, 3(6), 100489.

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77, 1-14, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2015.11.012.

Dias, O.C. (2011). Análise Qualitativa Comparativa (QCA) Usando Conjuntos Fuzzy – Uma Abordagem Inovadora Para Estudos Organizacionais no Brasil. *XXXV Encontro da ANPAD*. Rio de Janeiro.

Dignum, V. (2019). AI is multidisciplinar. *AI Matters*, 5(4), 19-21. https://doi.org/10.1145/3375637.3375644

Dignum,V. (2022). Relational Artificial Intelligence. https://doi.org/10.48550/arXiv.2202.07446

Djeffal, C. (2018). Sustainable AI Development (SAID): On the Road to More Access to Justice. http://dx.doi.org/10.2139/ssrn.3298980

Djeffal, C. (2022). Democracy, AI Regulation and the draft EU AI Act. *Transatlantic Policy Quarterly*. Disponível em http://turkishpolicy.com/article/1106/democracy-ai-regulation-and-the-draft-eu-ai-act Acessado em 20 de dezembro de 2022.

Domnich, A., & Anbarjafari, G. (2021). Responsible AI: Gender bias assessment in emotion recognition. arXiv:2103.11436 [cs.CV]. https://doi.org/10.48550/arXiv.2103.11436

Donahoe, E. & Metzger, M. M. (2019). Artificial Intelligence and Human Rights. *Journal of Democracy* 30(2), Johns Hopkins University Press, Retrieved Jun 12, 115-126.

Doneda, D. & Almeida. V.A.F. (2016). What Is Algorithm Governance? *IEEE Internet Computing*, 20(4), 60-63. doi: 10.1109/MIC.2016.79.

Donge, W. V., Bharosa, N., Janssen, M.F.W.H.A. (2022). Data-driven government: Cross-case comparison of data stewardship in data ecosystems. *Government Information Quarterly*, 39(2). https://doi.org/10.1016/j.giq.2021.101642.

Dubai Government (2019). Smart Dubai. Artificial Intelligence Principles and Ethics. Disponível em https://smartdubai.ae/initiatives/ai-principles-ethics Acessado em 5 de novembro de 2022.

Duijm, N. J. (2015). Recommendations on the use and design of risk matrices. Safety Science, 76, 21-31. ISSN 0925-7535. https://doi.org/10.1016/j.ssci.2015.02.014

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan Y., Dwivedi R., Edwards J., Eirug A., Galanos V., P. Ilavarasan V., Janssen M., Jones P., Kar A.K., Kizgin H., Kronemann B., Lal B., Lucini B., Medaglia R., Meunier-FitzHugh K. L. , Le Meunier-FitzHugh L.C., Misra S., Mogaji E., Sharma S.K., Singh J.B., Raghavan V., Raman R., Rana N.P., Samothrakis S., Jak Spencer, Tamilmani K., Tubadji A., Walton P., Michael D., & Williams M.D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Eggers, W. D., Schatsky, D., & Viechnicki, P. (2017). AI augmented government: using cognitive technologies to redesign public sector work. *Deloitte Center for Government Insights*.

Eggers, S., & Sample, C. (2020) Vulnerabilities in Artificial Intelligence and Machine Learning Applications and Data. Idaho National Laboratory. US. https://inldigitallibrary.inl.gov/sites/sti/sti/Sort_57369.pdf. Acessado em 13 junho 2022.

Eke, D.O., Chintu, S.S., & Wakunuma, K. (2023a). Towards Shaping the Future of Responsible AI in Africa. In: Eke, D.O., Wakunuma, K., Akintoye, S. (eds) Responsible AI in Africa. *Social and Cultural Studies of Robots and AI*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-08215-3_8

Eke, D.O., Wakunuma, K., & Akintoye, S. (2023b). Introducing Responsible AI in Africa. In: Eke, D.O., Wakunuma, K., Akintoye, S. (eds) Responsible AI in Africa. Social and Cultural Studies of Robots and AI. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-08215-3_1

Ekonomikos ir Inovaciju Ministerija (2018). Lithuanian Artificial Intelligence Strategy – A vision of the future. Disponível em http://kurklt.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf . Acessado em 11 de novembro de 2022.

Ekspertgruppen om dataetik (2018). Data i menneskets tjeneste Anbefalinger fra Ekspertgruppen om dataetik . Disponível em

https://em.dk/media/13315/ekspertgruppens-afrapportering-inkl-anbefalinger_final-a.pdf Acessado em 5 de dezembro de 2022.

Eisenhardt, K. M. (1989) Agency Theory: An Assessment and Review. *Academy of Management Review*. 14(1), 57-74.

Erlina, E., Nasution, A. A., Yahy, I., & Atmanegara, A. W. (2020). The role of risk based internal audit in improving audit quality. Erlina, Abdillah Arif Nasution, Idhar Yahya and Agung Wahyudhi Atmanegara, The Role of Risk Based Internal Audit in Improving Audit Quality, International Journal of Management, 11(12), 299-310.

EU GDPR (2016). General Data Protection Regulation. *European Parliament* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 Acessado em 27 Julho 2022.

European Commission (2018a). Artificial Intelligence for Europe. Disponível em https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN Acessado em 6 de novembro de 2022.

European Commission (2018b) AI Watch. Disponível em https://ai-watch.ec.europa.eu/about_en Acessado em 25 de fevereiro de 2023.

European Commission (2021a). Regulation of the European Parliament and of the Council Laying down harmonized rules on artificial intelligence. Artificial Intelligence Act and amending certain union legislative acts. Brussels. Disponível em https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF. Acessado em dezembro de 2021.

European Commission (2021b). Selected AI cases in the public sector. *Joint Research Centre* [Dataset] PID: http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a Acessado em 31 de outubro de 2022.

European Commission (2022a). Proposal for a Regulation of the European Parliament and the Council on European data governance (Data Governance Act) COM/2020/767 final

European Commission (2022c). New rules to improve road safety and enable fully driverless vehicles in the EU Disponível em https://ec.europa.eu/commission/presscorner/detail/en/IP_22_4312 Acessado em 29 de dezembro de 2022.

European Commission (2022c). First regulatory sandbox on Artificial Intelligence presented. Disponível em https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented Acessado em 24 de fevereiro de 2023.

European Data Protection Board (2022). Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement. Version 1.0 Disponível em https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en Acessado em 20 de dezembro de 2022.

European Parliament (2022a). Draft Report on the proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))

European Parliament (2022b). Regulatory divergences in the draft AI act: Differences in public and private sector obligations, Study, European Parliamentary Research Service (EPRS), Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729507/EPRS_STU(2022)729507_EN.pdf

European Union Agency for Cyber Security (2021). Securing Machine Learning Algorithms. ISBN: 978-92-9204-543-2 – DOI: 10.2824/874249.

European Union Agency for Cyber Security (2022). Risk Management Standards – Analysis of standardisaton requeriments in supporting of cybersecurity policy. Disponível em https://www.enisa.europa.eu/publications/risk-management-standards Acessado em 20 de dezembro de 2022.

Eyert F., Irgmaier, F., & Ulbricht, L. (2020). Extending the framework of algorithmic regulation. The Uber case. https://doi.org/10.1111/rego.12371

Fernandes, G., Domingues, J., Tereso, A., & Pinto, E. (2021). A Stakeholders' Perspective on Risk Management for Collaborative. *University-Industry R&D Programs.* Procedia Computer Science, 181, 110-118, ISSN 1877-0509. Https://doi.org/10.1016/j.procs.2021.01.110

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 1.

Floridi, L. (2003). On the Intrinsic Value of Information Objects and the Infosphere. Ethics and Information Technology, 4(4), 287–304.

Floridi, L. (2006). Information Ethics: Its Nature and Scope. In W.J. van den Hoven & J. Weckert, editors, *Moral Philosophy and Information Technology*. Cambridge University Press.

Floridi, L. (2018). Soft ethics, the governance of the digital and the General Data Protection Regulation. *Philosophical Transactions Series A Mathematical, Physical, and Engineering Sciences*, 376(2133), Article 20180081. https://doi.org/10.1098/rsta.2018.0081

Floridi L., Cowls J., King T., & Taddeo M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics* 26(3):1771-1796. doi: 10.1007/s11948-020-00213-5. Epub 2020 Apr 3. PMID: 32246245; PMCID: PMC7286860.

Firth-Butterfield, K. (2017). Artificial Intelligence and the Law: more questions than answers. *Scitech Lawyer*, 14, 28-31.

Freitas, V.S., Neto, F.B. (2016). Qualitative Comparative Analysis (QCA): usos e aplicações do método. *Revista Política Hoje*, 2ª Edição, 24 , 103-117.

Fundação para Ciência e Tecnologia (2021). Research in Data Science and Artificial Intelligence applied to Public Administration. Disponível em https://www.fct.pt/media/docs/Brochura_ResearchinDataScienceandAIappliedtoPA.pdf Acessado em 5 de novembro de 2022.

Future of Life Institute (2017). An Open Letter to the United Nations Convention on Certain Conventional Weapons - Future of Life Institute. https://futureoflife.org/autonomous-weapons-open-letter-2017/ Acessado em 13 de abril 2022

Future of Life Institute (2019). Ansilomar AI Principles. https://futureoflife.org/ai-principles/. Acessado em 13 de abril 2022.

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58–62. https://doi.org/10.1109/mic.2017.4180835

Georgieva, T., Timan, T. & Hoekstra, M. (2022), Regulatory divergences in the draft AI act: Differences in public and private sector obligations, Study, European Parliamentary Research Service (EPRS), Brussels. https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729507/EPRS_STU(2022)729507_EN.pdf.

Gerkensmeier, B., & Ratter, B.M.W. (2018). Governing coastal risks as a social process—Facilitating integrative risk management by enhanced multi-stakeholder collaboration. *Environmental Science & Policy*. 80, 144-151,ISSN 1462-9011. https://doi.org/10.1016/j.envsci.2017.11.011

German Federal Ministry for Economic Affairs and Energy (2020). German Standardization Roadmap on Artificial Intelligence. Disponível em https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf Acessado em 6 de dezembro de 2022.

German Federal Government (2017). Ethical guidelines for self-driving cars https://www.bmvi.de/SharedDocs/EN/publbications/report-ethics-commission.pdf?__blob=publicationFile

German Federal Government (2019) Opinions of the data ethics commission

https://assets.contentstack.io/v3/assets/blt3de4d56151f717f2/blt300ce23c9789e0f3/5e5cfe13fa08326331360f93/191023_DEK_Kurzfassung_en_bf.pdf Acessado em 13 de abril 2022.

German Federal Government (2020). Artificial intelligence strategy of the German Federal Government. https://www.ki-strategie-deutschland.de/home.html. Acessado em 13 de abril 2022

German Federal Government (2021). Germany road traffic act. https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl121s3108.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl121s3108.pdf%27%5D__1671810398003_Acessado em 22 de dezembro de 2022.

Gobierno de Chile (2020). Política Nacional de Inteligencia Artificial. Ministério de Ciencia, Tecnologia, Conocimiento e Innovación. Disponível em https://www.minciencia.gob.cl/uploads/filer_public/bc/38/bc389daf-4514-4306-867c-760ae7686e2c/documento_politica_ia_digital_.pdf Acessado em 11 de novembro de 2022.

Goenka, N., & Tiwari, S. (2021). Deep learning for Alzheimer prediction using brain biomarkers. Artif Intell Rev 54, 4827–4871. https://doi.org/10.1007/s10462-021-10016-0

Gong, Y., Yang, J., &Shi, X. (2020) Towards a comprehensive understanding of digital transformation in government: Analysis of flexibility and enterprise architecture, Government Information Quarterly, 37(3), 101487, ISSN 0740-624X, https://doi.org/10.1016/j.giq.2020.101487.

González, F., Ortiz, T., & Ávalos, R.S. (2020). Responsible use of AI for public policy: data science toolkit. OECD e IDB. https://publications.iadb.org/publications/english/document/Responsible-use-of-AI-for-public-policy-Data-science-toolkit.pdf

Gouvernment de France (2021). Stratégie Nationale pour l'Intelligence Artificielle. Disponível em https://www.intelligence-artificielle.gouv.fr/fr Acessado em 5 de novembro de 2022.

Government of Austria (2018). AIM AT 2030 – Artificial Intelligence Mission Austria 2030. Disponível em file:///C:/Users/patri/Downloads/aimat_ua%20(3).pdf Acessado em 6 de novembro de 2022.

Government of Canada (2019a). Directive on Identity Management. Disponível em https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=16577. Acessado em 24 de fevereiro de 2023.

Government of Canada (2019b). Directive on Security Management. Disponível em https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32611 Acessado em 24 de fevereiro de 2023.

Government of Canada (2020a). Algorithmic Impact Assessment available at https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html accessed 15 December 2020

Government of Canada (2020b). Guidelines on Service and Digital. Disponível em https://www.canada.ca/en/government/system/digital-government/guideline-service-digital.html#ToC4_5 Acessado em 24 de fevereiro de 2023.

Government of Canada (2021a). Directive on Automated Decision-making https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

Government of Canada (2021b). Levels of Security. National Security and Defense. Disponível em https://www.tpsgc-pwgsc.gc.ca/esc-src/protection-safeguarding/niveaux-levels-eng.html Acessado em 5 de dezembro de 2022.

Government of Canada (2022a). Responsible use of artificial intelligence – Our guiding principles. Disponível em https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 Acessível em 20 dez 2022.

Government of Canada (2022b). Pan-Canadian Artificial Intelligence Strategy. Disponível em https://ised-isde.canada.ca/site/ai-strategy/en Acessado em 20 de dezembro de 2022.

Government of Canada (2022c). List of interested Artificial Intelligence (AI) suppliers. Disponível em https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/list-interested-artificial-intelligence-ai-suppliers.html Acessado em 24 de fevereiro de 2023.

Government of Spain (2020). National Strategy for Artificial Intelligence. Disponível em https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf Acessado em 5 de novembro de 2022.

Government of India (2020a) Artificial Intelligence – Use Case Compendium. Ministry of Housing and Urban Affairs. Disponível em https://dsc.smartcities.gov.in/uploads/resource/resourceDoc/Resource_Doc_1663751038_AI-Use-Case-Compendium-Book.pdf Acessado em 4 de novembro de 2022.

Government of India (2020b). Datamart Cities. Empowering Cities Through Data. *Ministry of Housing and Urban Affairs*. Disponível em https://dsc.smartcities.gov.in/uploads/strategy/Strategy_doc_1626620947.pdf Acessado em 5 de novembro 2022.

Government of Ireland (2021). AI – Here for Good. A National Artificial Intelligence Strategy for Ireland. Disponível em https://www.gov.ie/en/publication/91f74-national-ai-strategy/. Acessado em 5 de novembro de 2021.

Government of the Grand Duchy of Luxembourg (2018). Artificial Intelligence: a strategic view for Luxembourg. Disponível em https://digital-luxembourg.public.lu/ Acessado em 5 de novembro de 2022.

Government the Republic of Estonia (2019). Kratt – Estonian Artificial Intelligence Deployment. Disponível em https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f_27a618cb80a648c38be427194affa2f3.pdf Acessado em 5 de novembro de 2022.

Government of Singapore (2019). National Artificial Intelligence Strategy – Advancing our Smart Nation Journey. Disponível em https://www.smartnation.gov.sg/initiatives/artificial-intelligence/#:~:text=By%202030%2C%20we%20see%20Singapore,to%20our%20citizens%20and%20businesses. Acessado em 5 de novembro de 2022.

Government of Sweden (2020). AI Sweden. Disponível em https://www.ai.se/en/about-0/strategic-areas . Acessado em 5 de novembro de 2022.

Government of United Kingdom (2017). Public sector use of the cloud. Disponível em https://www.gov.uk/guidance/public-sector-use-of-the-public-cloud Acessado em 5 de dezembro de 2022.

Government of United Kingdom (2019a). Centre for Data Ethics (CDEI) 2 Year Strategy - GOV.UK. https://www.gov.uk/government/publications/the-centre-for-data-ethics-and-innovation-cdei-2-year-strategy/centre-for-data-ethics-cdei-2-year-strategy. Acessado em 13 de abril 2022.

Government of United Kingdom(2019b). Understanding artificial intelligence ethics and safety. Disponível em https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety. Acessado em 2 de novembro de 2022.

Government of United Kingdom (2020a). Data Ethics framework. *Government Digital Service.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/923108/Data_Ethics_Framework_2020.pdf Acessado em 13 de abril 2022.

Government of United Kingdom (2020b). Guidelines for AI procurement. https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement Acessado em 13 de abril 2022.

Government of United Kingdom (2020c). Guidelines for AI procurement. A summary of best practices addressing specific challenges of acquiring Artificial Intelligence in the public sector. UK Office for Artificial Intelligence. Disponível em https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/990469/Guidelines_for_AI_procurement.pdf Acessado em 2 de dezembro de 2022.

Government of United Kingdom (2020d). Review into Bias in Algorithmic Decision-making. Disponível em https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm

ent_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf Acessado em 4 de dezembro de 2022.

Government of United Kingdom (2020e). The Government Data Quality Framework. Disponível em https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework-guidance Acessado em 26 de fevereiro de 2023.

Government of United Kingdom(2020f). A guide to using artificial intelligence in the public sector. Disponível em https://www.gov.uk/government/publications/a-guide-to-using-artificial-intelligence-in-the-public-sector Acessado em 26 de fevereiro de 2023.

Government of United Kingdom (2021a). Ethics, Transparency and Accountability Framework for Automated Decision-Making. Disponível em https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making Acessado em 4 de dezembro de 2022.

Government of United Kingdom (2021b). Algorithmic transparency template. Disponível em https://www.gov.uk/government/publications/algorithmic-transparency-template/algorithmic-transparency-template. Acessado em 4 de dezembro de 2022.

Government of United Kingdom (2021c). Algorithmic Transparency Standard. Disponível em https://www.gov.uk/government/collections/algorithmic-transparency-standard Acessado em 4 de dezembro de 2022.

Government of United Kingdom (2021d). Using personal data in your business or other organization. Disponível em https://www.gov.uk/guidance/using-personal-data-in-your-business-or-other-organisation#data-protection-and-gdpr Acessado em 5 de dezembro de 2022.

Government of United Kingdom (2021e). National AI Strategy. Disponível em https://www.gov.uk/government/publications/national-ai-strategy Acessado em 20 de dezembro de 2022.

Government of United Kingdom (2021f). Data Protection Impact Assessments. Disponível em https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/ Acessado em 24 de fevereiro de 2024.

Government of United Kingdom (2022a). Ethics Self-Assessment Tool. *UK Statistics Autority*. Disponível em https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/ Acessado em 2 de dezembro de 2022.

Government of United Kingdom (2022b) Standard. Food Standards Agency: Food Hygiene Rating Scheme – AI. Disponível em https://www.gov.uk/government/publications/food-standards-agency-food-hygiene-

rating-scheme-ai/food-standards-agency-food-hygiene-rating-scheme-ai A cessado em 29 de dezembro de 2022.

Government of United Kingdom (2022c). Equality Impact Assessment. Disponível em https://www.gov.uk/government/consultations/emergency-evacuation-information-sharing/equality-impact-assessment Acessado em 24 de fevereiro de 2023.

Government of United Kingdom (2022d). Service Standard. Disponível em https://www.gov.uk/service-manual/service-standard Acessado em 24 de fevereiro de 2023.

Government of United Kingdom (2022e). Data Sharing Governance Framework. Disponível em https://www.gov.uk/government/publications/data-sharing-governance-framework/data-sharing-governance-framework Acessado em 26 de fevereiro de 2023.

Grembergen, W. (2002). Introduction to the Minitrack" IT Governance and Its Mechanisms". In 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07).

Gu, T., Dolan-Gavitt, B, & Garg, S. (2019). BadNets: Identifying Vulnerabilities in Machine Learning Model Supply. rXiv:1708.06733v2 [cs.CR]

Gutierrez, C.I., & Marchant, G. (2021). A Global Perspective of Soft Law Programs for the Governance of Artificial Intelligence. Sandra Day O'Connor College of Law. Arizona State University.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Análise multivariada de dados. Bookman Editora. p. 415.

Hagendorff, T. (*2019*). The Ethics of AI Ethics - An Evaluation of Guidelines. *CoRR, abs/1903.03425.*

Hanna, M. J., & Kimmel, S. C. (2017). Current US federal policy framework for self-driving vehicles: Opportunities and challenges. *Computer*, 50(12), 32-40.

Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Abu Bakar, N. A. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management, 45,* 25–43. https://doi.org/10.1016/j.ijinfomgt.2018.10.007

Heracleois, L., & Lan, L. L. (2012). Agency Theory, Institutional Sensitivity and Inductive Reasoning: Towards a Legal Perspective. *Journal of Management Studies*. 49(1), 223-239.

Hernández-Nieto, R. (2002). Contributions to Statistical Analysis. Mérida: Universidad de Los Andes.

Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: the EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. European Business Organization Law Review, 22, 593-625.

Hickok, M. (2022). Public procurement of artificial intelligence systems: new risks and future proofing. AI & society, 1-15. https://doi.org/10.1007/s00146-022-01572-2

Hildebrandt, M. (2018). Algorithmic regulation and the rule of law. *Philosophy Transactions of the Royal Society*, 376.

Holmström, J. (2022). From AI to digital transformation: The AI readiness framework, Business Horizons, 65(3), 329-339, https://doi.org/10.1016/j.bushor.2021.03.006.

Hopster, J. (2021). What Are Socially Disruptive Technologies?. *Technology in Society* 67, 101750. https://doi.org/10.1016/j.techsoc.2021.101750.

Hopster, J., & Maas, M. M.(2022). Triaging the Technology Triad: Disruptive AI, Regulatory Gaps and Value Change.

Holder, C., Khurana, V., Harrison, F., & Jacobs, L. (2016a). Robotics and law: Key legal and regulatory implications of the robotics age (Part I of II). *Computer Law & Security Review*, 32(3), 383-402.

Holder C., Khurana V., Hook J., Bacon G., & Day R. (2016b). Robotics and law: key legal and regulatory implications of the robotics age (Part II of II). *Computer Law Secure Rev*iew; 32:557–576.

House of Lords (2018). AI in the UK: ready, willing and able?. Select Committee on Artificial Intelligence, Report of Session 2017-19. 13 March 2018.

Huawei (2018). AI Security White Paper. https://www-file.huawei.com/-/media/corporate/pdf/trust-center/ai-security-whitepaper.pdf Acessado em 13 de junho de 2022.

Hungarian Ministry for Innovation and Technology (2020). Hungary's Artificial Intelligence Strategy. Disponível em https://ai-hungary.com/files/e8/dd/e8dd79bd380a40c9890dd2fb01dd771b.pdf Acessível em 5 de novembro de 2022.

Husch B., & Teiden A. (2017). Regulating autonomous vehicles. *National Conference of State Legislature*. 25(13). https://www.ncsl.org/research/transportation/regulating-autonomous-vehicles.aspx Acessado em 13 de abril 2022.

IEEE (2019a). Ethically Aligned Design. Committees of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2nd version. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf Acessado em 13 de abril 2022.

IEEE (2019b). The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS). Disponível em https://standards.ieee.org/industry-connections/ecpais/ acessado em 13 de abril 2022.

IEEE (2020a). A Call to Action for Business Using AI - Ethically Aligned Design for Business. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead/ead-for-business.pdf Acessado em 13 de abril 2022.

IEEE (2020b). P7010 - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems. https://standards.ieee.org/ieee/7010/7718/ Acessado em 13 de abril de 2022.

IEEE (2021a). P7000 - Model Process for Addressing Ethical Concerns During System Design. https://standards.ieee.org/ieee/7000/6781/ Acessado em 13 de abril 2022.

IEEE (2021b). P7001 - Transparency of Autonomous Systems. https://standards.ieee.org/ieee/7001/6929/ Acessado em 13 de abril 2022.

IEEE (2021c). P7005 - Standard for Transparent Employer Data Governance. https://standards.ieee.org/ieee/7005/7014/ Acessado em 13 de abril 2022.

IEEE (2021d). P7007 - Ontological Standard for Ethically Driven Robotics and Automation Systems. https://standards.ieee.org/ieee/7007/7070/ Acessado em 13 de abril 2022.

IEEE (2022a). P7002 - Data Privacy Process. https://standards.ieee.org/ieee/7002/6898/ Acessado em 13 de abril 2022.

IEEE Computer Society (2022b). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems https://standards.ieee.org/industry-connections/ec/autonomous-systems/ Acessado em 13 de abril de 2022.

IFAD (1999) Good Governance: an overview. *Executive Board – Sixty-Seventh Session*, Rome, 8-9.

Illinois General Assembly (2008). Illinois Biometric Information Privacy Act (BIPA). Disponível em https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57 Acessado em 29 dezembro 2022.

Imai, T. (2019). Legal regulation of autonomous driving technology: Current conditions and issues in Japan. IATSS Research. 43(4), 263-267. https://doi.org/10.1016/j.iatssr.2019.11.009.

Information Commissioner's Office (2020). Big Data, artificial intelligence, machine learning and data protection. Version 2.2. https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf Acessado em 27 de junho de 2022

Information Commissioner's Office (2021). International data transfers. Disponível em https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/international-data-transfers/ Acessado em 5 de dezembro de 2022.

Information Commissioner's Office (2022). Regulatory Sandbox. Disponível em https://ico.org.uk/for-organisations/regulatory-sandbox/ Acessado em 24 de fevereiro de 2023.

Insurance Institute for Highway Safety (2022). Autonomous vehicle laws. Disponível em https://www.iihs.org/topics/advanced-driver-assistance/autonomous-vehicle-laws. Acessado em 29 de dezembro de 2022.

Inter-Parliamentary Union (2022). https://www.ipu.org/innovation-hub/open-data-hub Acessado em 31 de outubro de 2022.

IPS-X (2021). IPS-X Survey. European Cases. https://ipsoeu.github.io/ips-explorer/ Acessado em 31 de outubro de 2022.

ISO (2018). ISO 31:000 – Risk Managment – guideline. Disponível em https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en Acessado em 20 de dezembro de 2022.

ISO (2021a). ISO/IEC 24027 - Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making. https://www.iso.org/standard/77607.html Acessado em 13 de abril 2022.

ISO (2021b). ISO/IEC 24372 - Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems. https://www.iso.org/standard/78508.html Acessado em 13 de abril 2022

ISO (2021c). ISO/IEC 24668 - Information technology — Artificial intelligence — Process management framework for big data analytics. https://www.iso.org/standard/78368.html Acessado em 13 de abril 2022

ISO (2022a). ISO/IEC 38507 - Information Technology – Governance implications of the use of artificial intelligence by organizations. https://www.iso.org/standard/56641.html Acessado em 13 de abril 2022

ISO (2022b). ISO/IEC 23894 – Information Technology – Risk management. https://www.iso.org/standard/77304.html Acessado em 13 de abril 2022

ITGI (2003). Board Briefing on IT Governance. 2nd ed. Disponível em http://www.gti4u.es/curso/material/complementario/itgi_2003.pdf Acessado em 24 de fevereiro de 2023.

Jackson, B.W. (2019a). Artificial Intelligence and the Fog of Innovation: a deep-dive on governance and the liability of autonomous systems. 35 *Santa Clara High Tech*. L.J. 35.

Jackson B.W. (2019b). Cybersecurity, Privacy, and Artificial Intelligence: An Examination of Legal Issues Surrounding the European Union General Data Protection Regulation and Autonomous Network Defense, 21 MINN. J.L. SCI. & TECH. 169.

Jahn, K., & Kordyaka, B. (2019). The effects of robotic embodiment on intergroup bias: an experiment in immersive virtual reality. In Proceedings of the 27th European Conference on Information Systems (ECIS). Stockholm & Uppsala, Sweden, 8-14. ISBN 978-1-7336325-0-8 Research-in-Progress Papers. https://aisel.aisnet.org/ecis2019_rip/64

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Japanese Cabinet Office (2019). Social Principles of Human-Centric Artificial Intelligence. Council for Science, Technology and Innovation, (2019). https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf Acessado em 21 Julho 2022

Japanese Strategic Council for AI Technology (2017). Artificial Intelligence Technology Strategy. Disponível em https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf Acessível em 5 de novembro de 2022.

Jensen M., & Mechling W. H. (1976). Tehory of the firm: Managerial behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*. 3, 305-360.

Jing, H., Wei, W., Zhou, C., & He, X. (2021) An Artificial Intelligence Security Framework. *Journal of Physics: Conference Series*, Volume 1948, The 2021 2nd International Conference on Internet of Things. *Artificial Intelligence and Mechanical Automation* (IoTAIMA 2021), 14-16, Hangzhou, China

Jobin, A., Ienca, M., &Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat Mach Intell* 1**,** 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kaal, W. A., & Vermeulen, E. P.M. (2017). How to Regulate Disruptive Innovation - From Facts to Data. *Jurimetrics*. 57(2).

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93, 1449–1475.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. New York: *Cambridge University Press*.

Kale A., Nguyen T., Harris F.C., Li C., Zhang J., & Ma X. (2022). Provenance documentation to enable explainable and trustworthy AI: A literature review. *Data Intelligence*, 1-41. doi: https://doi.org/10.1162/dint_a_00119

Kazim, E., & Koshiyama, A.S. (2021). A high-level overview of AI ethics, *Patterns*, 2(9), 100314, ISSN 2666-3899, https://doi.org/10.1016/j.patter.2021.100314.

Kelly, E. (2011). Material Ethics of Value: Max Scheler and Nicolai Hartmann (Vol. 203). *Springer Science & Business Media*.

Khatri, V. (2016). Managerial work in the realm of the digital universe: The role of the data triad. *Business Horizons*, 59(6), 673–688. https://doi.org/10.1016/j.bushor.2016.06.001

Kim, M., Moon, J., & Kim, J. (2022). Analysis of Korean Road Traffic Regulations to Establish Legal Layer for Evaluating Safety of Autonomous Driving. *KSCE J Civ Eng*, 27, 313–321. https://doi.org/10.1007/s12205-022-1885-4

Kitsios, F., & Kamariotou, M. (2021). Artificial Intelligence and Business Strategy towards Digital Transformation: A Research Agenda. *Sustainability*, 13, https://doi.org/10.3390/su13042025

Kliegr, T., Bahník, S., & Fürnkranz, J. (2021). A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence*, 295, 103458, ISSN 0004-3702. https://doi.org/10.1016/j.artint.2021.103458.

Korjani, M.M. & Mendel, J.M. (2012). Fuzzy set Qualitative Comparative Analysis (fsQCA): Challenges and applications. *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, 1-6. doi: 10.1109/NAFIPS.2012.6291026.

Kozuka, S. (2019). A governance framework for the development and use of artificial intelligence: lessons from the comparison of Japanese and European initiatives. *Uniform. Law Review*., 24, 315–329**.**

Kraus, S., Durst, S., Ferreira, J.J., Veiga, P., Kailer,N., & Weinmann,A. (2022). Digital transformation in business and management research: An overview of the current status quo. International Journal of Information Management, 63,102466,ISSN 0268-4012. https://doi.org/10.1016/j.ijinfomgt.2021.102466.

Krippendorff, K. (2013). Content Analysis – An Introduction to Its Methodology. Sage. 3td Edition.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, *44*(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *IEEE 22nd Conference on Business Informatics (CBI)*, 201-210. doi: 10.1109/CBI49978.2020.00029.

Laboratório de Inteligência Artificial Aplicada da 3ª Região (2022). Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de IA no âmbito do LIAA-3R / Grupo de Validação Ético-Jurídica (GVEJ) do LIAA-3R, iLabTRF3, iJuspLab. - 2. ed., rev. e atual. - São Paulo.

Larsson S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society*, 1-23.

Laato, S., Birkstedt, T., Mäantymäki, M., Minkkinen, M., & Mikkonen, T. (2022). AI governance in the system development life cycle: Insights on responsible machine

learning engineering. In Proceedings of the *1st International Conference on AI Engineering: Software Engineering for AI* , 113-123.

Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, Power and Bias in Artificial Intelligence. *ArXiv, abs/2008.07341*.

Leftwich, A. (1993). Governance, Democracy and Development in the Third World. *Third World Quarterly*, 14, 605–24.

Leitner, C., & Stiefmueller, C. M. (2019). Disruptive technologies and the public sector: The changing dynamics of governance. In Public service excellence in the 21st century (pp. 237-274). Singapore: Springer Singapore.

Lenardon, J.P.A. (2017). The Regulation of Artificial Intelligence. *Master Thesis. Tilburg Institute for Law, Technology and Society*. Netherlands

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529

Lewis, T., & Yildirim, H. (2002). Learning by Doing and Dynamic Regulation. *The RAND Journal of Economics, 33*(1), 22-36.

BLK (2020). Korea opens way for development and commercialization of self-driving cars, with framework for designating road sections and test zones for "Level 3" and up. Lexology. Disponível em https://www.lexology.com/library/detail.aspx?g=ba58457c-d448-4f03-a033-9a93826c1bba Acessado em 29 de dezembro de 2022.

Li, H., Yu, L., Tian, S., Li, L., Wang, M., & Lu, X. (2017). Deep learning in pharmacy: The prediction of aqueous solubility based on deep belief network. Automatic Control and Computer Sciences, 51, 97-107.

Liang, L. & Acuna, D.E. (2020). Artificial mental phenomena: psychophysics as a framework to detect perception biases in AI models. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). *Association for Computing Machinery*, New York, NY, USA, 403–412. https://doi.org/10.1145/3351095.3375623

Lin, Y. T., Hung, T. W., & Huang, L. T. L. (2021). Engineering equity: How AI can help reduce the harm of implicit bias. Philosophy & Technology, 34(Suppl 1), 65-90.

Liu, R., Jia, C., Wei, J., Xu, G., & Vosoughi, S. (2022a). Quantifying and alleviating political bias in language models. Artificial Intelligence, 304, 103654.

Liu, Y. L., Huang, L., Yan, W., Wang, X., & Zhang, R. (2022b). Privacy in AI and the IoT: The privacy concerns of smart speaker users and the Personal Information Protection Law in China. Telecommunications Policy, 46(7), 102334.

Lloyd, K. (2018). Bias amplification in artificial intelligence systems. arXiv preprint arXiv:1809.07842.

Locher, M.A., & Bolander,B. (2019). Ethics in pragmatics, *Journal of Pragmatics*, 145, 83-90, ISSN 0378-2166, https://doi.org/10.1016/j.pragma.2019.01.011.

Ma, L., Zhang Z., & Zhang N. (2018). Ethical Dilemma of Artificial Intelligence and its Research progress. *IOP Conference Series: Materials Science and Engineering*. 392, 062188. https://doi.org/10.1088/1757-899X/392/6/062188

McGraw. G., Bonett, R., Shepardson, V., & Figueroa, H. (2020). The Top 10 Risks of Machine Learning Security. *Computer*, 53(6), 57-61. doi: 10.1109/MC.2020.2984868.

Makarius, E.E., Mukherjee, D., Fox, J.D., & Fox, A.K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. Journal of Business Research, 120, 262-273. https://doi.org/10.1016/j.jbusres.2020.07.045.

Makhlouf, K., Zhioua, S., & Palamidessi, C. (2021). Machine learning fairness notions: Bridging the gap with real-world applications. Information Processing & Management, 58(5), 102642.

Mantelero, A. (2018). AI & Big Data: A blueprint for human rights, social and ethical impact assessment. *Computer Law & Security Review*, 34(4), 754-772.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. AI and Ethics, 2(4), 603-609. https://doi.org/10.1007/s43681-022-00143-x

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022b). Putting AI Ethics into Practice: The Hourglass Model of Organizational AI Governance. arXiv:2206.00335 [cs.AI]. https://doi.org/10.48550/arXiv.2206.0033

Manzini, E. J. (2004). Entrevista semi-estruturada: análise de objetivos e de roteiros. *Seminário Internacional sobre Pesquisa e Estudos Qualitativos*, 2, 2004, Bauru. A pesquisa qualitativa em debate. Anais. Bauru: USC, isbn:85-98623-01-6.

Maluf S. (1995). Teoria Geral do Estado. 23ª ed., 205-208. *Editora Saraiva*. São Paulo.

Marchant, G. (2019). "Soft Law" Governance of Artificial Intelligence. *UCLA: The Program on Understanding Law, Science, and Evidence (PULSE)*. Disponível em https://escholarship.org/uc/item/0jq252ks Acessado em 24 de fevereiro de 2023.

Marda V. (2018). Artificial intelligence policy in India: a framework forengaging the limits of data-drivendecision-making.Phil.Trans.R.Soc.A376:20180087. http://dx.doi.org/10.1098/rsta.2018.0087

Matei, A., & Drumasu, C. (2015). Corporate Governance and public sector entities. Procedia Economics and Finance, 26, 495-504.

Mehrabi, N., Morstatter, F., Peng, N., & Galstyan, A. (2019). Debiasing community detection: The importance of lowly connected nodes. In *Proceedings of the 2019*

Medaglia, R., Gil-Garcia, J. R., & Pardo, T. A. (2021). Artificial Intelligence in Government: Taking Stock and Moving Forward. *Social Science Computer Review*, 1–18. https://doi.org/10.1177/08944393211034087

Meijerink, J., & Bondarouk, T. (2018). Uncovering configurations of HRM service provider intellectual capital and worker human capital for creating high HRM service value using fsQCA. Journal of business research, 82, 31-45.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big Data in Public Affairs. *Public Administration Review, 76*(6), 928–937. https://doi.org/10.1111/puar.12625

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data and Society*, *7*(2). https://doi.org/10.1177/2053951720948087

Microsoft (2020). Responsible AI Principles from Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3Aprimaryr6. Acessado em 13 de abril 2022.

Mika N., Nadezhda G., Jaana L., & Raija K., (2019). Ethical AI for the Governance of the Society: Challenges and Opportunities. *CEUR Workshop Proceedings*, 2505, 20-26. http://ceur-ws.org/Vol-2505/paper03.pdf Accessed 20 July 2020

Mikalef, P, Conboy, K, Lundström, J.E. & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side' of AI, *European Journal of Information Systems*, 31(3), 257-268, DOI: 10.1080/0960085X.2022.2026621

Millar, J. (2016). An Ethics Evaluation Tool for Automating Ethical Decision-Making in Robots and Self-Driving Cars, *Applied Artificial Intelligence*, 30(8), 787-809.

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2023). Co-shaping an ecosystem for responsible AI: five types of expectation work in response to a Technological Frame. Information Systems Frontiers, 25(1), 103-121.

Ministério da Ciência Tecnologia e Inovação do Brasil (2021). Estratégia Brasileira de Inteligência Artificial. Governo do Brasil. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_documento_referencia_4-979_2021.pdf Acessado em 13 de abril de 2022.

Ministério das Relações Exteriores (2022). De outros países no Brasil https://www.gov.br/mre/pt-br/assuntos/Embaixadas-Consulados-Missoes/de-outros-paises-no-brasil Acessado em 31 de outubro de 2022.

Ministero dello sviluppo economico (2019). Proposte per uma Strategia italiana per l'intelligenza artificiale. Disponível em https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf Acessado em 5 de novembro de 2022.

Ministry of Economic Affairs and Employment of Finland (2017). Suomen tekoälyaika

Suomi tekoälyn soveltamisen kärkimaaksi: Tavoite ja toimenpidesuositukset. Disponível em https://julkaisut.valtioneuvosto.fi/handle/10024/80849 Acessado em 20 de dezembro de 2022.

Ministry of Economic Affairs and Employment of Finland (2019). Leading the Way into the Era of Artificial Intelligence: Final Report of Finland's Artificial Intelligence Program 2019. Ministry of Economic Affairs and Employment of Finland. 133pp. http://urn.fi/URN:ISBN:978-952-327-437-2 Acessado em 13 de abril de 2022.

Misuraca, G., & van Noordt, C. (2020). Overview of the use and impact of AI in public services in the EU, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399

Moeini, M., & Rivard, S. (2019). Sublating tensions in the IT project risk management literature: a model of the relative performance of intuition and deliberate analysis for risk assessment. J. Assoc. Inf. Syst. 20. https://doi.org/10.17705/1jais.00535.

Mohammad, S. M. (2017). DevOps Automation and Agile Methodology. *International Journal of Creative Research Thoughts (IJCRT*), ISSN:2320-2882, 5(3), 946-949, August-2017, https://ssrn.com/abstract=3655581

Mökander J. & Floridi .L (2021). Ethics-based auditing to develop trust-worthy AI. Mind Mach 31,323–327. https:// doi. org/ 10. 1007/s11023- 021- 09557-8

Monetary Authority of Singapore (2019). Monetary Authority of Singapore. Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector. http://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf Accessed 20 July 2020

Moor J. H. (1985). What Is Computer Ethics? In T.W. Bynum, editor, *Computers and Ethics*, 263–275. Blackwell.

Moor J. R. (1998). Relativity and Responsibility in Computer Ethics. *Computers and Society*, 28(1), 14-21.

Moor J. (1999). Just Consequentialism and Computing. Ethics and Information Technology, 1, 65–69.

Morley, J., Machado, C.C.V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020a). The ethics of AI in health care: A mapping review, *Social Science & Medicine*, 260, 113172, ISSN 0277-9536, https://doi.org/10.1016/j.socscimed.2020.113172.

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020b). From what to how: an initial review of publicly available ai ethics tools, methods and research to translate principles into practices. *Sci Eng Ethics* 26, 2141–2168. https:// doi. org/ 10. 1007/ s11948- 019- 00165-5

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13-22. doi: 10.1177/160940690200100202.

Muller, C., Schöppl, N., & Peñalver, M. F. (2022a). AIA in-depth #1 – Objective, Scope, Definition – articles 1-4 & Annex I. ALLAI. Disponível em https://allai.nl/aia-in-depth-series-of-papers/ Acessado em 29 de dezembro de 2022

Muller, C., Schöppl, N., & Peñalver, M. F. (2022b.) AIA in-depth #2 – Prohibited AI Practices – article 5. ALLAI. Disponível em https://allai.nl/aia-in-depth-series-of-papers/ Acessado em 29 de dezembro de 2022.

Muller, C., Schöppl, N., & Peñalver, M. F. (2022c) AIA in-depth #3b - High-Risk AI Requirements – articles 9-15, 42, 43. ALLAI. Disponível em https://allai.nl/aia-in-depth-series-of-papers/ Acessado em 29 de dezembro de 2022.

Nagbøl, P. R., & Müller, O. (2020). X-RAI: a framework for the transparent, responsible, and accurate use of machine learning in the public sector. In IFIP EGOV-ePart-CeDEM conference. p. 259. CEUR Workshop Proceedings.

Nagbøl, P. R., Müller, O., & Krancher, O. (2021). Designing a risk assessment tool for artificial intelligence systems. In The Next Wave of Sociotechnical Design: *16th International Conference on Design Science Research in Information Systems and Technology*, DESRIST 2021, Kristiansand, Norway, August 4–6, 2021, Proceedings 16 (pp. 328-339). Springer International Publishing. https://doi.org/10.1007/978-3-030-82405-1_32

National Institute of Advanced Industrial Science and Technology (2022). Machine Learning Quality Management Guideline – 2nd English edition. Government of Japan – *Digital Architecture Research Center*. Disponível em https://www.digiarc.aist.go.jp/en/publication/aiqm/aiqm-guideline-en-2.1.1.0057-e26-signed.pdf Acessado em 4 de dezembro de 2022.

Navarro, S., Llinares, C., & Garzon, D. (2016). Exploring the relationship between cocreation and satisfaction using QCA. Journal of Business Research, 69(4), 1336–1339.

Neznamov A.V. (2020). Regulatory Landscape of Artificial Intelligence Advances in Social Science, Education and Humanities Research, 420, 201-204. XVII *International Research-to-Practice Conference 2020*. Atlantatis Press.

Nevejans, N. (2016). European Civil Law Rules in Robotics. Study requested by the European Parliament's Committee on Legal Affairs. *Policy Department Citizens' Right and Constitutional Affairs*.

Norwegian Data Protection Authority (2018). Datatilsynet. https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf Acessado em 13 de abril de 2022.

Norwegian Ministry of Local Government and Modernisation (2020). National Strategy for Artificial Intelligence. Disponível em

https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf Acessível em 20 de dezembro de 2022.

Phillips, P.J., Hanan, C.A., Fontana, P.C., Broniatowski, D.A. & Przybocki, M.A. (2021). Four Principles of Explainable Artificial Intelligence. National Institute of Standards and Technology. U.S. Department of Commerce.

NIST (2022). AI Risk Management Framework: first draft. https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf Acessado em 13 junho 2022.

National Policy Agency of Japan (2020). Revised Road Traffic Act – related to automated driving. Disponível em https://www.npa.go.jp/english/bureau/traffic/selfdriving.html Acessado em 29 de dezembro de 2022.

Nerur, S., & Balijepally, V. (2007). Theoretical reflections on agile development methodologies. Commun. ACM 50(3), 79–83. https://doi.org/10.1145/1226736.1226739

NITI Aayog (2018). National Strategy for Artificial Intelligence: #AI for All (Discussion Paper) https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf Accessed 30 July 2020.

NITI Aayog (2021). Responsible AI For All. https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf

NITI Aayog (2022). Responsible AI #AIForAll. Disponível em https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf Acessível em 20 de dezembro 2022.

Nolan J. (2013). The Corporate Responsibility to Respect Human Rights: Soft Law or Not Law? in S Deva and D Bilchitz (eds) *Human Rights Obligations of Business: Beyond the Corporate Responsibility to Respect.* Cambridge University Press.

Nordic Council of Ministers (2018). AI in the Nordic-Baltic region. Disponível em https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf Acessado em 29 de dezembro de 2022.

Ntoutsi, E., Fafaliosb P., Gadirajua U., Iosidisa V., Nejdla W., Vidalc M., Ruggierid S., Turinid F., Papadopoulose S., Krasanakise E., Kompatsiarise I., Kinder-Kurlandaf K., Wagnerf C., Karimif F., Fernandezg M., Alanig H., Berendth B., Kruegeli T., Heinzei C., Broelemannj K., Kasnecij G., Tiropanisk T., & Staab S. (2020). Bias in Dat-driven AI Systems – An Introductory Survey. arXiv:2001.09762v1 [cs.CY]

Ojo, A., Mellouli, S., & Ahmadi Zeleti, F. (2019). A Realist Perspective on AI-era Public Management. In *20th Annual International Conference on Digital Government Research,* 159-170. ACM.

Oneto, L. & Chiappa, S. (2020). Fairness in Machine Learning. arXiv:2012.15816 [cs.LG]

Oppy, G., & Dowe, D.L. (2011). The Turing Test. In Edward N. Zalta, editor, Stanford *Encyclopedia of Philosophy*. Stanford University. Disponível em http://plato.stanford.edu/entries/turing-test/.

Osborne, S. (1997). Managing the Coordination of Social Services in the Mixed Economy of Welfare: Competition, Cooperation or Common Cause? *British Journal of Management*, 8, 317–28.

Özdemir, V., & Hekim, N. (2018). Birth of industry 5.0: Making sense of big data with artificial intelligence,"the internet of things" and next-generation technology policy. *Omics: a journal of integrative biology*, 22(1), 65-76.

Organisation for Economic Co-operation and Development (2019). Recommendation of the Council on Artificial Intelligence. Disponível em https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#:~:text=The%20Recommendation%20is%20open%20to,to%20the%20COVID%2D19%20crisis. Acessado em 13 de abril 2022.

Organisation for Economic Co-operation and Development (2022a). Artificial Intelligence Observatory. Disponível em https://oecd.ai/en Acessado em 13 de abril 2022.

Organisation for Economic Co-operation and Development (2022b). United States – Local State and Federal Regulations on facial recognition technologies. Organization for Economic and Co-operation Development Artificial Intelligence Observatory. https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-26890 Acessado em 13 abril 2022.

Organisation for Economic Co-operation and Development (2022c). Policy Initiatives for Emerging AI-related regulation, Civil society. https://oecd.ai/en/dashboards/policy-initiatives?conceptUris=http:%2F%2Fai.oecd.org%2Fmodel%23Emerging_technology_regulation%7C%7Chttp:%2F%2Fai.oecd.org%2Ftaxonomy%2FtargetGroups%23TG16 Acessado em 16 junho 2022

OECD/CAF (2022), The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean, OECD Public Governance Reviews, OECD Publishing, Paris, https://doi.org/10.1787/1f334543-en.

Oxford Insight (2021). Government Artificial Intelligence Readiness Index – GAIRI. Disponível em https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/61ead0752e7529590e98d35f/1642778757117/Government_AI_Readiness_21.pdf Acessado em 13 de abril 2022.

ParlAmericas (2022) https://parlamericas.org

Partnership on AI to Benefit People and Society (2016). Disponível em https://www.partnershiponai.org/about/ Acessado em 24 de fevereiro de 2023.

Poel, I. V. (2016). An ethical framework for evaluating experimental technology. *Science and Engineering Ethics*, Springer, 22(3), 667–686.

Prakken, H. (2017). On the problem of making autonomous vehicles conform to traffic law. *Artificial Intelligence & Law*, 25, 341-363.

Presidencia de la Nación (2019). ARGENIA – Plan Nacional de Inteligencia Artificial. Argentina. Disponível em https://ia-latam.com/wp-content/uploads/2020/09/Plan-Nacional-de-Inteligencia-Artificial.pdf Acessado em 11 de novembro de 2022.

Ragin, C.C. (2008). Redesigning Social Inquiry: Fuzzy Sets and Beyond, 85- 97, Univ. of Chicago Press, Chicago, IL

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the *2020 conference on fairness, accountability, and transparency*, 33-44.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G. & Chin, M. H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Ann. Intern. Med.*, 169(12), 866–872. doi: 10.7326/M18-1990

Rahwan, I. (2017). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, Springer. 20, 5-14.

Rahwan, I., Cebrian, M., Obradovich, N. *et al.* (2019) Machine behaviour. *Nature* 568, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Republic of Korea Government (2019). National Strategy for Artificial Intelligence. Disponível em https://www.msit.go.kr/bbs/view.do?sCode=eng&mId=10&mPid=9&bbsSeqNo=46&nttSeqNo=9. Acessado em 6 de novembro de 2022.

Republica de la Colombia (2019). Política Nacional para la Transformacion Digital e Inleligência Artificial. Compes 3975. Disponível em https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3975.pdf. Acessado em 11 de novembro 2022.

República Portuguesa (2021). Estrategia Nacional de Inteligência Artificial. Disponível em https://portugaldigital.gov.pt/acelerar-a-transicao-digital-em-portugal/conhecer-as-estrategias-para-a-transicao-digital/estrategia-nacional-de-inteligencia-artificial/#:~:text=A%20Estrat%C3%A9gia%20Nacional%20de%20Intelig%C3%AAncia,e%20dos%20seus%20estados%2Dmembros. Acessado em 5 de novembro de 2022.

Rhahla, M., Allegue, S., & Abdellatif, T. (2021). Guidelines for GDPR compliance in Big Data systems. *Journal of Information Security and Applications*, *61*, 102896. https://doi.org/10.1016/j.jisa.2021.102896

Rhodes, R. (1997) Understanding Governance, Buckingham: Open University Press.

Richie, J., & Lewis, J. (2003). Qualitative research practice. *A guide for social science students and Researchers*.

Rihoux, B., & Ragin, C. (2008). Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques. London and Thousand Oaks, CA: Sage.

Roorda, S. A. H. (2021). Facial Recognition for Public Safety a supportive tool for the municipal decision-making process on using facial recognition for public safety, the FRPS risk governance method (Master's thesis, University of Twente).

Rose, J., Flack L.S.; Sæbø, Ø (2018) Stakeholder theory for the E-government context: Framing a value-oriented normative core, *Government Information* Quarterly, 35(3), Pages 362-374, ISSN 0740-624X, https://doi.org/10.1016/j.giq.2018.06.005.

Roselli, D.,Matthews, J., & Talagala, N. (2019). Managing Bias in AI. In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19). *Association for Computing Machinery*, New York, NY, USA, 539–544. https://doi.org/10.1145/3308560.3317590

Rousseau, J. (2016). The Social Contract. (202-230). ISBN: 978911495741. London: Sovereign.

Rubicondo, D., & Rosato, L. (2022) AI Fairness Addressing Ethical and Reliability Concerns in AI Adoption. IASON Essential Services for Financial Institutions. https://www.iasonltd.com/doc/rps/2022/ai_fairness_addressing_ethical_and_reliability_concerns_in_ai_adoption.pdf . Acessado em 12 de junho 2022.

Ruijer, E. (2021). Designing and implementing data collaboratives: A governance perspective. *Government Information Quarterly*, 38(4), 101612. https://doi.org/10.1016/j.giq.2021.101612

Russell, S., & Norvig, P. (1995). *Artificial Intelligence. A Modern Approach*. Prentice Hall, New Jersey. 4-5.

Ruttkamp-Bloem, E. (2023). Epistemic Just and Dynamic AI Ethics in Africa. In: Eke, D.O., Wakunuma, K., Akintoye, S. (eds) Responsible AI in Africa. Social and Cultural Studies of Robots and AI. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-031-08215-3_2

Rzeczypospolitej Polskiej (2020). Polityka dla rozwoju sztucznej inteligencji w Polsc. Disponível em https://www.gov.pl/web/govtech/polityka-rozwoju-ai-w-polsce-przyjeta-przez-rade-ministrow--co-dalej#:~:text=Rozw%C3%B3j%20AI%20w%20Polsce%20zwi%C4%99kszy,miejsca%20pracy%20w%20kluczowych%20sektorach. Acessado em 5 de novembro de 2022.

Brailsford, S.C., Potts, C.N., & Smith, B.M. (1999) Constraint satisfaction problems: Algorithms and applications, *European Journal of Operational Research*, 119(3), 557-581, ISSN 0377-2217, https://doi.org/10.1016/S0377-2217(98)00364-6.

Saarikko, T., Westergren, U. H., Blomquist,T. (2020). Digital transformation: Five recommendations for the digitally conscious firm. *Business Horizons*, 63(6), 825-839. https://doi.org/10.1016/j.bushor.2020.07.005.

Saldaña. J. (2013). The Coding Manual for Qualitative Researchers. Sage. 2nd Edition.

Senden, L. (2005) Soft law, self-regulation and co-regulation in European Law: Where Do They Meet? Electronic Journal of Comparative Law. 9(1).

Schoenauer., M., Bonnet., Y., Berthet., C., Cornut, A., Levin, F., & Rondepierre, B. (2018). For a Meaningful Artificial Intelligence: Toward a French and European Strategy. Mission assigned by the French Prime Minister. Disponível em https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf Acessado em 13 de abril 2022.

Schrader, D., & Ghosh, D. (2018). Proactively Protecting Against the Singularity: Ethical Decision Making AI. *IEEE Computer and Reliability Societies Review,* 16(3), 56-63.

Scheler, M. (1973). Formalism in ethics and non-formal ethics of values: A new attempt toward the foundation of an ethical personalism. Northwestern University Press.

Scherer, M.U. (2016). Regulating Artificial Intelligence Systems: Risks, Challenges, Competences and Strategies. *Harvard Journal of Law & Technology,* 29(2), 354-398.

Scupola, A., & Mergel, I. (2022). Co-production in digital transformation of public administration and public value creation: The case of Denmark. *Government Information Quarterly*. Volume 39(1). https://doi.org/10.1016/j.giq.2021.101650.

Sekretariat Nasional Kecerdasan Artifisial Indonesia (2020). Strategi Nasional untuk Kecerdasan Artifisial (STRANAS KA). Disponível em https://ai-innovation.id/server/static/ebook/stranas-ka.pdf. Acessado 11 de novembro de 2022.

Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute, 1(6).

Silveira M.B., Saldanha R.P., Leite J.C.C., Silva T.O.F.D., Silva T., & Filippin L.I. (2018). Construction and validation of content of one instrument to assess falls in the elderly. Einstein (Sao Paulo). 11;16(2):eAO4154. doi: 10.1590/S1679-45082018AO4154.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 59, 99–118.

Simon, H. A. (1956). Rational Choice and Structure of the Environment. *Psychological Review*., 63(2), 129–138. https://doi.org/10.1037/h0042769

Simonsson, M., & Johnson, P. (2006). Defining IT governance-a consolidation of literature. In *The 18th conference on advanced information systems engineering*, 6.

Scott, C. (2000). Accountability in the regulatory state. *Journal of law and society*, *27*(1), 38-60.

Sikstrom, L., Maslej, M.M., Hui, K., Findlay, Z., Buchman, D.Z., & Hill, S.L. (2022). Conceptualising fairness: three pillars for medical algorithms and health equity. *BMJ Health & Care Informatics* (IF). DOI: 10.1136/bmjhci-2021-100459

Shah, M.U. &. Guild, P.D. (2022). Stakeholder engagement strategy of technology firms: A review and applied view of stakeholder theory. Technovation, (114), ISSN 0166-4972. https://doi.org/10.1016/j.technovation.2022.102460.

Sharma, G.D, Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda, Sustainable Futures, 2, ISSN 2666-1888. https://doi.org/10.1016/j.sftr.2019.100004

Schneider, C. Q., & Wagemann, C. (2012). Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis. United Kingdom: Cambridge University Press.

Smuha, N. A. (2021). Beyond a human rights-based approach to AI governance: Promise, pitfalls, plea. *Philosophy & Technology*, 34(Suppl 1), 91-104.

Stahl, B.C. (2008). Researching Ethics and Morality in Information Systems: Some Guiding Questions. *ICIS*.

Stahl, B.C., Antoniou, J., Ryan, M. et al. (2022a) Organisational responses to the ethical issues of artificial intelligence. AI & Soc 37, 23–37. https://doi.org/10.1007/s00146-021-01148-6

Stahl, B.C., Rodrigues, R., Santiago, N., & Macnish, K. (2022b). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. Computer *Law & Security Review*, 45, 105661.

Stanford (2021) National Artificial Intelligence Strategies and Human Rights: a review. Global Digital Policy Incubator. Disponível em https://www.gp-digital.org/wp-content/uploads/2021/05/NAS-and-human-rights_2nd_ed.pdf. Acessado em 11 de novembro de 2021.

Stix, C. (2021). Foundations for the future: institution building for the purpose of artificial intelligence governance. *AI Ethics*. https://doi.org/10.1007/s43681-021-00093-w

Stop Killer Robots Coallition (2020). The Campaign To Stop Killer Robots. Disponível em https://www.stopkillerrobots.org/ Acessado em 13 de abril de 2022.

Strauß, S. (2021). Deep Automation Bias: How to Tackle a Wicked Problem of AI? Big Data Cogn. Comput., 5, 18. https://doi.org/10.3390/bdcc5020018

Stuurman, K., & Lachaud, E. (2022). Regulating AI. A label to complete the proposed Act on Artificial Intelligence, *Computer Law & Security Review*, 44, 105657, ISSN 0267-3649, https://doi.org/10.1016/j.clsr.2022.105657

Sweden Innovation Agency (2018). Artificial Intelligence in Swedish Business and Society. Disponível em https://www.vinnova.se/contentassets/29cd313d690e4be3a8d861ad05a4ee48/vr_18_09.pdf Acessado em 13 de novembro de 2022.

Switzerland Federal Council (2020). Guidelines on Artificial Intelligence for the Confederation. file:///C:/Users/patri/Downloads/leitlinien-ki_e.pdf

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good: An ethical framework will help to harness the potential of AI while keeping humans in control. *Science Review*, 361(6404), 751-752.

Taeihagh, A. (2021). Governance of artificial intelligence, *Policy and Society*, 40(2), 137–157, https://doi.org/10.1080/14494035.2021.1928377

Tomalin, M., Byrne, B., Concannon, S., Saunders, D., & Ullman, S.. (2021). The practical ethics of bias reduction in machine translation: why domain adaptation is better than data debiasing. *Ethics Inf Technol* **23,** 419–433. https://doi.org/10.1007/s10676-021-09583-1

Tangerding, E. (2021). Beyond Data Protection: Applying the GDPR to Facial Recognition Technology. essay: 87597. thesis. Disponível em http://essay.utwente.nl/87597/ Acessado em 20 de dezembro de 2022.

Tangi L., van Noordt C., Combetto M., Gattwinkel D., & Pignatelli F. (2022). AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-53058-9, doi:10.2760/39336, JRC129301

Tokmakov, M.A. (2019). Corporate Governance Modernization: Legal Trends and Challenges. *SHS Web of Conferences*, 71, P. 04011. Edp Sciences.

Torlig, E.G.S., Resende-Júnior, P.C., Fujihara, R.K. (2019). Proposição de uma Nova Orientação para Validação de Roteiros em Pesquisas Qualitativas. *XLIII Encontro da ANPAD* – EnANPAD 2019, São Paulo.

Torlig, E.G.S., Resende-Júnior, P.C., Fujihara, R.K., Demo, G., & Montezano, L. (2022). Validation Proposal for Quaitative Research Scripts (Vali-Quali). *Administração: Ensino e Pesquisa.* 23 (1)-4-29; Jan-Abr. DOI 10.13058/raep.2022.v23n1.2022

Toronto (2020). The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems. Disponível em https://www.torontodeclaration.org/ Acessado em 20 dezembro 2022.

The National Council for Artificial Intelligence(2019). Egypt National Artificial Intelligence Strategy. Disponível em https://mcit.gov.eg/Upcont/Documents/Publications_672021000_Egypt-National-AI-Strategy-English.pdf. Acessado em 11 de novembro de 2022.

Thiem, A. (2010). Set-relational fit and the formulation of transformational rules in fsQCA. COMPASSS Working Paper., no 2010–61. Houston: Department of Sociology, University of Houston-Downtown. Disponível em: <http://www.compasss.org/wpseries/ Thiem2010.pdf>

Thimbleby, H. (2008). Robot ethics? Not yet A reflection on Whitby's "Sometimes it's hard to be a robot". Interacting with Computers, 20(3), 338-341.

Tribunal de Contas da União (2021) TC 006.662/2021-8. https://portal.tcu.gov.br/data/files/1C/62/96/7E/06DF08102DFE0FF7F18818A8/006.662-2021-8-AC%20-%20Levantamento_Inteligencia_Artificial.pdf Acessado em 31 de outubro de 2022.

Tribunal de Contas da União (2022). Acordão nº1139/2022 – TCU – Plenário.

Tutt, A. (2017). An FDA for Algorithms. *Administrative Law Review*, 69(83), 83-123.

UNESCO (2019) Steering AI and advanced ICTs for knowledge societies: a Rights, Openness, Access, and Multi-stakeholder Perspective. https://unesdoc.unesco.org/ark:/48223/pf0000372132 Acessado em 13 de abril 2022.

UNESCO. (2020). First version of a draft text of a recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000373434

UNESCO (2021) Report of Social and Human Sciences Commission - SHS. Disponível em https://unesdoc.unesco.org/ark:/48223/pf0000379920

United States Government (2021). National Artificial Intelligence Initiative Act. Disponível em https://www.ai.gov/ Acessado em 5 de novembro de 2022.

University of Montreal (2018). Montreal Declaration for a Responsible Development of Artificial Intelligence. https://www.montrealdeclaration-responsibleai.com/the-declaration Acessada em 13 de abril 2022.

University of Ottawa (2017). Call for an International Ban on the Weaponization of Artificial Intelligence | Centre for Law. *Technology and Society*. Disponível em https://techlaw.uottawa.ca/bankillerai Acessada em 13 de abril 2022.

United Arab Emirad (2018) National Strategy for Artificial Intelligence 2031. Disponível em https://ai.gov.ae/wp-content/uploads/2021/07/UAE-National-Strategy-for-Artificial-Intelligence-2031.pdf Acessado em 5 de novembro de 2022.

United States Department of Defense (2018) Defense Innovation Board AI Principles. Disponível em https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF Acessado em 13 de abril 2022.

Vanhée l., & Borit, M. (2022) Viewpoint: Ethical by Designer – How to grow Ethical Designsers of Artificial Inelligence. Journal of Artificial Intelligence Research, 73, 619-631. https://doi.org/10.1613/jair.1.13135

Vandercruysse, L., Buts, C., & Dooms, M. (2020). A typology of smart city services: the case of data protection impact assessment. Cities, 104, 102731.

Vero (2019). Finnish Tax Administration's ethical principles for AI. Disponível em https://www.vero.fi/en/About-us/finnish-tax-

administration/operations/responsibility/finnish-tax-administrations-ethical-principles-for-ai/#:~:text=Our%20AI%20follows%20laws%20and%20regulations&text=The%20use%20of%20AI%20does,our%20partners%20carefully%20and%20responsibly. Acessado em 5 de dezembro 2022.

Vetrò, A., Torchiano, M., & Mecati, M. (2021). A data quality approach to the identification of discrimination risk in automated decision-making systems. *Government Information Quarterly*, Volume 38, Issue 4, ISSN 0740-624X. https://doi.org/10.1016/j.giq.2021.101619

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118-144, ISSN 0963-8687. https://doi.org/10.1016/j.jsis.2019.01.003.

Villaronga, E.F., & Heldeweg, M. (2018). Regulation, I presume? said the robot – Towards an iterative regulatory process for robot governance. *Computer Law & Security Review*, 21.

Vilminko-Heikkinen, R. & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, 47, 76-87, ISSN 0268-4012, https://doi.org/10.1016/j.ijinfomgt.2018.12.017.

Vining, R., McDonald, N., McKenna, L., Ward, M. E., Doyle, B., Liang, J., ... & Brennan, R. (2022). Developing a Framework for Trustworthy AI-Supported Knowledge Management in the Governance of Risk and Change. In HCI International 2022-Late Breaking Papers. Design, User Experience and Interaction: *24th International Conference on Human-Computer Interaction, HCII 2022*, Virtual Event, June 26–July 1, 2022, Proceedings, 318-333. Cham: Springer International Publishing.

Vergara, S. C. (2005). Métodos de pesquisa em administração. São Paulo: Atlas, 16-20.

Xue, M., Yuan, C., Wu, H., Zhang, Y., & Liu, W. (2020). Machine Learning Security: Threats, Countermeasures, and Evaluations. in *IEEE Access*, 8, 74720-74742, 2020, doi: 10.1109/ACCESS.2020.2987435.

Wallach, W., & Marchant, G. (2019). Toward the Agile and Comprehensive International Governance of AI and Robotics [point of view], in *Proceedings of the IEEE*, 107(3), 505-508. doi: 10.1109/JPROC.2019.2899422.

Weill, P., & Ross, J. W. (2004). It Governance on One Page. Available at SSRN: https://ssrn.com/abstract=664612 or http://dx.doi.org/10.2139/ssrn.664612

Weiss P. (1942). Morality and Ethics. *The Journal of Philosophy*, 39, 14, 381-385. Disponível em http://www.jstor.org/stable/2018625. Acessado em 5 de novembro de 2022.ssado em 11 novembro de 2022.

Wang, D., Khosla, A., Gargeya, R., Irshad, H.; Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. Preprint at https://arxiv.org/ abs/1606.05718

White House (2020). Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government. https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-promoting-use-trustworthy-artificial-intelligence-federal-government/

Wirtz, B.W., Weyerer, J.C., & Geyer, C. (2018a). Artificial Intelligence and the Public Sector—Applications and Challenges, *International Journal of Public Administration*. 42(7), 596-615.

Wirtz, B.W., & Müller, W.M. (2018b). An integrated artificial intelligence framework for public management, Public Management Review, 21(7), 1076-1100, DOI: 10.1080/14719037.2018.1549268

Wirtz, B.W.; Weyerer, J.C; Kehl,I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. Government Information Quarterly. ISSN 0740-624X. https://doi.org/10.1016/j.giq.2022.101685.

World Bank Group (2021) Harnessing Artificial Intelligence for Development in the Post-Covid-19 Era. A review of National AI Strategies and Policies. Disponível em https://www.worldbank.org/en/topic/digitaldevelopment/brief/harnessing-artificial-intelligence-for-development-in-the-post-covid-19-era Acessado em 5 de novembro de 2022.

World Economic Forum (2020a). Unblocking Public Sector AI – AI procurement in a Box: workbox.

World Economic Forum (2020b) Unblocking Public Sector AI. AI Procurement in a Box: Pilot case studies from the United Kingdom.

World Economic Forum (2020c) Unblocking Public Sector AI. AI Procurement in a Box: Project overview.

World Economic Forum (2020d) A Framework for Responsible Limits on Facial Recognition Disponível em https://www3.weforum.org/docs/WEF_Framework_for_action_Facial_recognition_2020.pdf Acessado em 20 de dezembro 2022.

Wright, S.A., & Schultz, A. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. *Business Horizons*, 61(6), 823-832.

Yapo A., & Weiss J. (2018). Ethical Implications of Bias in Machine Learning. *Hawaii International Conference on System Sciences.* 51st edition. 2018.

Yeung, K., Howes, A., & Pogrebna, G. (2019). AI governance by human rights-centred design, deliberation and oversight: An end to ethics washing. *The Oxford Handbook of AI Ethics*, 77-106.

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management, 57(2), 102025.

Zicari, R. V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., ... & Westerlund, M. (2021). Z-Inspection®: a process to assess trustworthy AI. IEEE Transactions on Technology and Society, 2(2), 83-97.

Zuiderwijk, A., Chen, Y., & Salem, F. (2021). Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. *Government Information Quarterly*, 38(3). ISSN 0740-624X. https://doi.org/10.1016/j.giq.2021.101577

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

This survey is part of Patricia G. R. de Almeida's* PhD thesis at the University of Brasilia, Department of Administration, supervised by Professor Dr. Carlos Denner**.

This study aims to research how public organizations are implementing or adapting their governance, management, and development models to produce safe Artificial Intelligence (AI) systems that follow ethical principles.

The questions, estimated to be answered within 15 minutes, are directed at professionals in charge of data science teams, AI development teams, AI projects teams, or AI policies teams.

The results will be published in scientific journals and websites after a thorough analysis of the entire sample.

There will be no mention of any organizations or respondents in the published papers.

*patricia.almeida@camara.leg.br
**carlosdenner@unb.br

*** Do you agree to answer the following questionnaire?**

- ◯ Yes
- ◯ No

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

This questionnaire is divided into 4 steps:

1. **Organization's information and context**
2. **Respondent's information**
3. **Organization's AI system information**
4. **Organization's governance, management, and development processes**

For the purposes of this survey:

- **"Organization"** refers to the public organization hereby represented by the respondent.
- **"AI system"** refers to a system that uses Artificial Intelligence technology to support decision-making or autonomous actions.

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

1- ORGANIZATION'S INFORMATION AND CONTEXT

*** 1.1 Country**

*** 1.2 Type of public organization**

*** 1.3 Organization size**

*** 1.4 Is the organization subject to any laws that regulate AI?**

*** 1.5 Is the organization subject to any government policy regarding AI systems?**

*** 1.6 Is the organization subject to any laws that regulate personal data and privacy?**

*** 1.7 For how long has the organization been developing, buying, or using AI systems?**

University of Brasília – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

2- RESPONDENT'S INFORMATION

*** 2.1 Your department**

*** 2.2 Your position within the organization**

*** 2.3 Your level of education**

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

3- ORGANIZATION’S AI SYSTEM INFORMATION

3.1 Please, provide information about the organization's 5 most frequently used AI systems

System 1

What is the system's aim? (Up to 250 characters)

The most relevant technological approach

Examples:

NLP - Natural Language Processing	*Chatbots (question-answering), speech recognition (speech to text), translation, summarization, semantic textual similarity, sentiment analysis, argument mining, text classification and topic modeling.*
CV - Computer Vision	*Face recognition, optical character recognition, facial emotion recognition, object detection.*

Resources for development

System 2

What is the system's aim? (Up to 250 characters)

The most relevant technological approach

Exemples:

NLP - Natural Language Processing	*Chatbots (question-answering), speech recognition (speech to text), translation, summarization, semantic textual similarity, sentiment analysis, argument mining, text classification and topic modeling.*
CV - Computer Vision	*Face recognition, optical character recognition, facial emotion recognition, object detection.*

Resources for development

System 3

What is the system's aim? (Up to 250 characters)

The most relevant technological approach

Exemples:

NLP - Natural Language Processing	*Chatbots (question-answering), speech recognition (speech to text), translation, summarization, semantic textual similarity, sentiment analysis, argument mining, text classification and topic modeling.*
CV - Computer Vision	*Face recognition, optical character recognition, facial emotion recognition, object detection.*

Resources for development

System 4

What is the system's aim? (Up to 250 characters)

The most relevant technological approach

Exemples:

NLP - Natural Language Processing	*Chatbots (question-answering), speech recognition (speech to text), translation, summarization, semantic textual similarity, sentiment analysis, argument mining, text classification and topic modeling.*
CV - Computer Vision	*Face recognition, optical character recognition, facial emotion recognition, object detection.*

Resources for development

System 5

What is the system's aim? (Up to 250 characters)

The most relevant technological approach

Exemples:

NLP - Natural Language Processing	*Chatbots (question-answering), speech recognition (speech to text), translation, summarization, semantic textual similarity, sentiment analysis, argument mining, text classification and topic modeling.*
CV - Computer Vision	*Face recognition, optical character recognition, facial emotion recognition, object detection.*

Resources for development

*** 3.2 Choose the type of AI system user**

Choose as many as you like

- [ ] Internal staff
- [ ] Citizens or other organizations
- [ ] Others (please specify)

*** 3.3 Where can the data used by the organization's AI systems be found?**

- ◯ Internally, including only the organization’s own data for training and inferencing purposes.
- ◯ Externally; regardless of how the organization’s own data is used, external data (from other organizations) is also used for training and/or inferencing purposes.

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

4- ORGANIZATION’S GOVERNANCE, MANAGEMENT, AND DEVELOPMENT PROCESSES

PLEASE, CONSIDER THE MEANINGS PROVIDED BELOW WHEN ANSWERING THE FOLLOWING QUESTIONS

ANSWER	MEANING
Yes, completely.	The organization owns or has implemented **all or most of the item's requirements**.
Yes, but only partially.	The organization owns or has implemented the f**ewer item's requirements**. Examples: limited resources, project still underway, project implemented while meeting only the basic requirements, programs distributed across several different projects.
No, but a formal decision has been made to implement it.	The organization has formally decided to implement the item. However, the organization has not initiated any projects to that effect yet.
No, and no formal decision has been made to implement it.	No, and no formal decision has been made to implement it.

*** 4.1 Does the organization have an AI Strategy (exclusively for AI or built into a Digital Transformation Strategy)?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.2 Does the organization have any policy related to AI systems?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.3 Does the organization have a person, council/committee, or department that is in charge of AI Governance?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.4 Does the organization have a code of ethics or ethical principles applied to AI systems?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.5 Has the organization established any Data Governance process or policy?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.6 Has the organization established any data quality management process?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.7 Has the organization established any personal data protection management process?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.8 Has the organization established any AI system risk management process that consider the entire lifecycle of AI systems?**

For the purposes of this questionnaire, "AI system lifecycle" refers to the entire period from project inception until the AI system is withdrawn from operation.

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.9 Has the organization established any AI system security management process that consider the entire lifecycle of AI systems in order to avoid cyberattacks?**

For the purposes of this questionnaire, "AI system lifecycle" refers to the entire period from project inception until the AI system is withdrawn from operation.

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.10 Does the organization audit its AI systems?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.11 Has the organization established any AI system development process?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.12 Does the organization provide training regarding data, AI-related risks and AI ethical principles to decision-makers?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.13 Does the organization provide training regarding data, AI development, and AI-related risks and AI ethical principles to** developers?

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.14 Does the organization provide training or communication program regarding data to users?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.15 Does the organization provide training regarding data and AI-related risks and AI ethical principles to internal auditors?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.16 At the planning stage of a project, does the organization identify AI system stakeholders while considering the entire AI system lifecycle?**

For the purposes of this questionnaire, "AI system lifecycle" refers to the entire period from project inception until the AI system is withdrawn from operation.

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.17 Does the organization adopt any practices for representing ethical rules and ethical dilemmas while AI systems are being developed?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.18 When developing AI systems, does the organization adopt any practices to mitigate data biases, such as identification of protected groups, identification of historical biases, mechanisms to avoid incorrect correspondences between ideal variables and available variables, or identification of incomplete attributes and missing values?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.19 Does the organization monitor how AI systems are performing after making them available to users?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.20 When looking for data biases or cognitive biases, does the organization implement human oversight for AI system outputs after they are made available to users?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.21 Does the organization monitor changes in the environment, such as societal trends, emergent practices, norms, and behaviors, that could impact AI system requirements?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

*** 4.22 After AI systems are made available to users, does the organization establish mechanisms for collecting user feedback about those systems?**

- ◯ Yes, completely.
- ◯ Yes, but only partially.
- ◯ No, but a formal decision has been made to implement it.
- ◯ No, and no formal decision has been made to implement it.

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

*** Name**

*** E-mail**

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

Please, feel free to leave your comments regarding one or more questions

University of Brasilia – Brazil
Department of Business Management
https://ppga.unb.br/

QUESTIONNAIRE

Please, click the "DONE" button to finish the form

Thank you very much!

Patricia Gomes Rêgo de Almeida

patricia.almeida@camara.leg.br

Tese - Anexo 2 – Roteiro de entrevistas realizadas

PORTUGUÊS		
ROTEIRO DE ENTREVISTAS		
INSTRUÇÕES	**Em alguns itens, a pergunta somente será feita se no questionário, a pergunta sobre esse item tiver sido respondida positivamente.** **Em outros itens, a pergunta somente será feita se no questionário, a pergunta sobre esse item tiver sido respondida negativamente.** **Sua organizaão significa a organização para a qual o entrevista representa.**	
CONSTRUTOS		**QUESTÕES**
GOVERNANÇA DA IA	**GOVERNANÇA, ESTRATÉGIA, POLÍTICAS, PRINCÍPIOS ÉTICOS**	**Na sua organização, quem define as diretrizes mais estratégicas sobre sistemas de IA?** Ex: quais unidades podem desenvolver sistemas de IA, quais problemas podem ser resolvidos pela IA, quais projetos terão IA, quando desenvolver internamente ou por meio de terceiros**.** **Quais os níveis de governança e/ou gestão envolvidos? Como ocorre?**
		Se há política ou normativo para sistemas de IA: Quais as definições e atribuições são estabelecidas pela política/norma de sistemas de IA na sua organização? **Se não há política ou normativo para sistemas de IA:** **Há outras políticas ou normativos gerais que a equipe de projeto do sistema de IA segue?** Ex: normas para gestão de projetos, normas para gestão de mudanças.
		Se há Declaração de Princípios Éticos que sejam aplicados aos sistemas de IA: Na sua organização, como foi a construção de diretrizes com princípios éticos para o uso da IA? As diretrizes foram aprovadas por qual nivel de governança?
		Na sua organização, quem autoriza o tratamento nos dados? Há uma lista dos proprietáros dos dados? **Se há processo ou política para governança de dados: O que a política ou processo de Governança de Dados define ou estabelece?**
	GESTÃO DE DADOS	**Se há processos para gestão de qualidade de dados: Quais unidades ou pessoas atuam para garantir boa qualidade nos dados da sua organização? Como funciona? A implantação ou projeto de gestão de qualidade de dados iniciou-se antes do desenvolvimento de sistemas de IA da organização?**
		Se há processos para gestão da proteção de dados pessoais: Quais unidades ou pessoas atuam para garantir a proteção de dados pessoais da sua organização? Como funciona? A implantação de gestão da proteção de dados pessoais iniciou-se antes do desenvolvimento de sistemas de IA da organização?
GESTÃO DO DESENVOLVIMENTO E SUSTENTAÇÃO DOS SISTEMAS DE IA	**FASES DE PROJETO**	**Se há práticas para atender a princípios éticos no desenvolvimento: Durante o desenvolvimento dos sistemas de IA, poderia explicar quais e como são realizadas as práticas para minimizar vieses? Há algo que você considera importante melhorar?**
		Na sua organização, quais as práticas são utilizadas para aumentar a transparência do processo de desenvolvimento de sistemas de IA?
		Se há sistemas de IA contratados: Na sua organização, como é feita a gestão do processo produtivo dos sistemas de IA desenvolvidos por terceiros? Após o desenvolvimento, quem assume a manutenção e evolução?
	FASES DO SISTEMA EM OPERAÇÃO	**Se há prática para monitoração de sistemas de IA em operação: Na sua organziação, após o sistema de IA ser colocado em operação, como é feita a monitoração? O que é observado?** **Se não há prática para monitoração de sistemas de IA em operação: Na sua percepção, quais fatores mais dificultam a implantação de monitoração de sistemas de IA em operação?**
CONSIDERAÇÕES FINAIS		**Deseja fazer algum comentário ou acrescentar informações?**

Tese - Anexo 2 – Roteiro de entrevistas realizadas

INGLÊS		
INTERVIEW SCRIPT		
INSTRUCTIONS		**For some items, the question will only be asked in case a positive answer was given in the questionnaire for the corresponding item.** **For other items, the question will only be asked in case a negative answer was given in the questionnaire for the corresponding item.** **"Your organization" refers to the organization represented hereby by the interviewee.**
CONSTRUCTS		**QUESTIONS**
AI GOVERNANCE	GOVERNANCE, STRATEGY, POLICIES, ETHICAL PRINCIPLES	**In your organization, who is in charge of devising the most strategic guidelines for AI systems?** E.g. which departments can develop AI systems, which problems can be solved with AI, which projects will use AI, whether AI systems are restricted to internal development or can be developed via outsourcing or partnerships. **Which governance/management levels are involved?**
		In case AI system policies ornorms exist: In your organization, which definitions and responsibilities are established by AI system policies/norms? **In case no AI system policy or norms exist: Does the AI system project team follow any other general policies or norms**? E.g. project management process, change management process.
		In case an Ethical Principles Declaration applies to AI systems: In your organization, how was the experience of drawing up the ethical principles guidelines for AI systems? Which governance level was responsible for approving them?
		In your organization, who authorizes data treatment? Is there a list of data owners? **In case a data governance process or policy exists: What does the Data Governance policy or process define or stipulate?**
	DATA MANAGEMENT	**In case data quality management processes exist: Which departments or employees are involved in data quality assurance? How does it work? Was the data quality management process or project implemented before the organization's AI systems were developed?**
		In case personal data protection management processes exist: Which departments or employees are involved in personal data protection assurance? How does it work? Was the personal data protection management process implemented before the organization's AI systems were developed?
AI SYSTEM DEVELOPMENT AND MAINTENANCE MANAGEMENT	IFECYCLE - PROJECT STAGE	**In case practices exist to meet the ethical principles during the development stage: While AI systems are being developed, which practices does your organization adopt for minimizing biases? How does it work? Is there anything you think should be improved?**
		What practices does your organization adopt to make the AI system development process more transparent?
		In case outsourced AI systems exist: How does your organization manage the production process of AI systems developed by third parties? Once developed, who is in charge of maintenance and improvement?
	LIFECYCLE - SYSTEM IN OPERATION	**In case practices are adopted to monitor AI systems in operation: After making AI systems available to users, how does your organization monitor them? What aspects of them are observed?** **In case no practices are adopted to monitor AI systems in operation: From your standpoint, what are the factors that create the most obstacles to monitoring AI systems in**
FINAL REMARKS		**Would you like to leave a comment or provide further information?**

Tabela 1 – Resultados do cálculo do Coeficiente de Validade de Conteúdo do questionário.

Resultados da Avaliação do Questionário

Questão	CVC - Clareza				CVV - Pertinência			
	Média	CVCi	Pei	CVCc	Média	CVCi	Pei	CVCc
1	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
2	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
3	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
4	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
5	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
6	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
7	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
8	4.250	0.850	0.004	0.846	4.250	0.850	0.004	0.846
9	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
10	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
11	4.500	0.900	0.004	0.896	4.500	0.900	0.004	0.896
12	4.000	0.800	0.004	0.796	4.250	0.850	0.004	0.846
13	4.250	0.850	0.004	0.846	4.000	0.800	0.004	0.796
14	4.250	0.850	0.004	0.846	4.250	0.850	0.004	0.846
15	4.500	0.900	0.004	0.896	4.750	0.950	0.004	0.946
16	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
17	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
18	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
19	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
20	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
21	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
22	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
23	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
24	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
25	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
26	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
27	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
28	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
29	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
30	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
31	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
32	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
33	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
34	4.750	0.950	0.004	0.946	5.000	1.000	0.004	0.996
35	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
36	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
37	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
38	5.000	1.000	0.004	0.996	5.000	1.000	0.004	0.996
		CVCt-clareza		0.957		CVCt-pertinencia		0.975

Fonte: Elaboração própria.

Tabela 1: Resultados da análise VALI-QUALI do roteiro de entrevista.

Resulados da Avaliação do Roteiro de Entrevista					
	DIMENSÃO CONTEÚDO		DIMENSÃO SEMÂNTICA		
	Alinhamento com o objetivo	Aderência ao construto	Clareza	Expectância qualitativa	
	Grau de alinhamento ao objetivo da pesquisa	Grau de aderência ao construto investigado	Grau de clareza da pergunta	Grau de expectativa qualitativa da resposta	
i	Médias				Qi
1	4.80	4.80	4.00	3.80	4.35
2	4.80	4.80	4.60	4.40	4.65
3	4.80	4.60	4.00	4.00	4.35
4	5.00	5.00	4.20	4.40	4.65
5	4.40	4.80	4.20	4.40	4.45
6	4.20	4.00	4.60	4.20	4.25
7	4.80	4.80	4.80	4.20	4.65
8	4.80	4.80	4.80	4.20	4.65
9	5.00	5.00	4.80	4.20	4.75
10	5.00	5.00	4.80	4.20	4.75
11	4.80	4.80	4.80	4.20	4.65
12	4.80	4.80	4.60	4.20	4.60
13	4.40	4.00	5.00	4.40	4.45
14	5.00	4.80	4.80	4.40	4.75
15	4.80	4.40	5.00	4.80	4.75
16	4.00	4.00	4.60	3.80	4.10
17	5.00	4.80	5.00	4.20	4.75

Fonte: Elaboração própria.

Tabela 1: Critérios de aceitação das questões em avaliação VALI-QUALI

Critérios de aceitação no VALI-QUALI	
Aprovação total	Pontuação média de Qi = 5,0
Modificação opcional	Pontuação média de 4,5 ≤ Qi < 5,0
Modificação necessária	Pontuação média de 2,5≤ Qi < 4,5
Exclusão	Pontuação média de 1 ≤Qi < 2,5

Fonte: Torlig et al. (2022)

Tese – Anexo 5 – Variáveis-conjunto utilizadas – descrição e fundamentação

Tabela 1: Variáveis-conjunto consideradas para o construto "Fatores Geradores de Expectativas"

Construto	Variável-conjunto	Fonte	Opções de valores
FATORES GERADORES DE EXPECTATIVAS	Lei para IA	Questionário - q 1.4	0;1
	Lei para proteção de dados pessoais	Questionário - q 1.6	0;1
	Política de governo com recomendações para sistemas de IA	Questionário - q 1.5	0;1

Fonte: Elaboração própria.

Tabela 2: Variáveis-conjunto consideradas para o construto "Governança de IA".

Governança de IA		
Variável-conjunto	Argumentação	Fonte Opções de Valor
Estratégia de IA	Embutida em uma estratégia de transformação digital ou em proposta própria, a existência de uma estratégia de IA para um país tem sido um fenômeno cada vez mais comum (Stanford 2021; Wirtz et al. 2018; Schoenauer et al. 2018; World Bank Group 2021, Oxford Insights 2021). Seu alcance geralmente ocorre para cada esfera de poder do sistema político de um país. Assim, no governo federal, apenas para ministérios, agências, e departamentos do poder Executivo; para Estados e municípios, instrumentos próprios são preparados. Como consequência, o sistema Judiciário e o Legislativo necessitariam, cada um, de sua própria estratégia para atingir serviços digitais construídos por meio de sistemas de IA, além de ações e pessoas articuladas para viabilizar tais entregas.	Questionário - q 4.1 0; 33; 67; 100
Política para sistemas de IA	Sob o formato de normas, regulamentos, portarias, resoluções, entre outros, políticas organizacionais podem estabelecer como e em que condições o provimento de sistemas de IA deve ocorrer (Marda 2018, LIAA-3R 2022).	Questionário - q 4.2 0; 33; 67; 100
Código ou declaração de princípios éticos para sistemas de IA	Conjunto de orientações que têm sido apresentadas pelas organizações públicas ou agências de governo, em suas respectivas esferas, designadas para que, centralizadamente, estabeleçam padrões norteadores ao uso e provimento de sistemas de IA nas organizações públicas.	Questionário - q 4.4 0; 33; 67; 100
Processo de governança de IA	Modelo utilizado na tomada das decisões mais estratégicas sobre o uso e provimento de sistemas de IA, quando sistematizado e devidamente publicizado, o processo de governança pode fazer parte da governança corporativa e/ou da governança de TI, ou ser um processo independente (ISO 2022a; Zuiderwijk et al. 2021).	Entrevista 0; 33; 67; 100
Estrutura/ pessoa responsável/ comitê para governança da IA	A operacionalização da sistemática de decisões sobre o uso e produção de sistemas de IA pode ser materializada por meio de uma unidade administrativa própria, comitê ou apenas a designação formal de uma pessoa para tal finalidade (Dignum 2022; Stix 2021);	Questionário - q 4.3 0; 33; 67; 100
Treinamento para tomadores de decisão	A percepção de que a capacitação de stakeholders em posição-chave é um habilitador à implantação efetiva de práticas de governança de IA, incluindo seus processos auxiliares, sugere a verificação se tais práticas estão sendo aplicadas aos gestores da organização (Ahn & Chen 2022; Benfeldt et al. 2020; Dignum 2022; Vanhée & Borit 2022).	Questionário - q 4.12 0; 33; 67; 100
Treinamento para desenvolvedores	Idem item anterior, porém, dirigido aos cientistas de dados e desenvolvedores de sistemas de IA.	Questionário - q 4.13 0; 33; 67; 100
Treinamento para usuários	Idem item anterior, porém, voltado aos usuários, e utilizando uma linguagem mais simples e abordagem voltada ao fornecimento e uso dos dados.	Questionário - q 4.14 0; 33; 67; 100
Treinamento para auditores internos	A percepção de que auditoria interna é um mecanismo de gestão de riscos (Erlina et al. 2020;de Oliveira 2019) e de preparação para auditorias externas pelos órgãos de controle (Zicari et al. 2021), a capacitação de auditores internos se apresenta como habilitadora para a governança de IA.	Questionário - q 4.15 0; 33; 67; 100

Fonte: Elaboração própria.

Tabela 3: Variáveis-conjunto consideradas para o construto “Processos e Práticas Auxiliares”

Processos e práticas auxiliares na governança de IA		
Variável-conjunto	**Argumentação**	**Fonte** **Opções de Valor**
Processo ou política para governança de dados	Sob a forma de normativo ou de um processo, é benéfico ao provimento de sistemas de IA, o estabelecimento de um modelo sistematizado de decisões sobre o que pode ser feito com os dados, quem pode decidir sobre eles, e quem deve cuidar em cada etapa do ciclo de vida dos dados (Medaglia et al. 2021; Haneem et al. 2019; Vetrò, 2021).	Questionário - q 4.5 0; 33; 67; 100
Processo para gestão da qualidade de dados	Sistemas de IA com resultados confiáveis dependem, entre outros fatores, de dados com qualidade, o que envolve conjuntos de ações para garantir que os dados estejam completos, íntegros e sirvam aos propósitos planejados para os sistemas de IA (Haneen et al. 2019; Khatri 2016; Rhahla et al. 2021).	Questionário - q 4.6 0; 33; 67; 100
Processo para gestão da proteção de dados pessoais	Quando dados sobre pessoas são utilizados pelos sistemas de IA, a utilização de processos para gestão de tais dados passa a ser necessária para garantir a privacidade dos indivíduos (Vandercruysse et al. 2020; Kuziemski 2020).	Questionário - q 4.7 0; 33; 67; 100
Processo para gestão de riscos de sistemas de IA	As práticas para Governança de IA têm sido fundamentadas na gestão de riscos em todo o ciclo de vida dos sistemas de IA. Especialmente na prestação de serviços públicos, cujos stakeholders são diversos e em constantes mudanças, a gestão de riscos assume uma posição de destaque contemplando elementos sociais além dos técnicos (Wirtz et al. 2022; Manterelo 2018; BSA 2021; NIST 2022).	Questionário - q 4.8 0; 33; 67; 100
Processo para gestão da segurança de sistemas de IA	O caráter autônomo, a amplitude dos serviços baseados em IA na administração pública, aliados a modelos de ataques cibernéticos específicos para determinadas técnicas de algoritmos de IA, impõem uma atenção contínua na segurança de tais sistemas durante todo o seu ciclo de vida (Jackson 2019b; Eggers & Sample 2020; European Union Agency for Cyber Security 2021).	Questionário - q 4.9 0; 33; 67; 100
Processo de Auditoria interna nos sistemas de IA	Governos têm preparado processos para avaliar o uso e a produção de sistemas de IA. Nesse contexto, as auditorias internas nas organizações públicas podem ter um papel muito importante tanto na preparação para as avaliações externas (Baterman & Powles 2021), quanto para a auto-regulação (Raji et al 2020; Wirtz et al. 2022; Zicari 2021).	Questionário - q 4.10 0; 33; 67; 100
Monitoração de mudanças de ambiente e tendências sociais	Diante das rápidas e constantes mudanças do ambiente social, regulamentos, leis, e demais variáveis de ambiente interno e externo à organização, a mitigação de riscos é robustecida por um acompanhamento dessas eventuais mudanças, com o fito de internalizar tais *insights* para retroalimentar o ciclo de vida dos sistemas de IA (González et al 2020).	Questionário - q 4.21 0; 33; 67; 100

Fonte: Elaboração própria.

Tabela 4: Variáveis-conjunto consideradas no construto "Gestão de Desenvolvimento e Sustentação da IA"

Práticas do Processo de Desenvolvimento de Sistemas de IA que visam Princípios Éticos+A1:C10		
Variável-conjunto	**Argumentação**	**Fonte** **Opções de Valor**
Identificação dos stakeholders	Apontada como tarefa basilar para várias práticas e processos dirigidos à governança da IA e riscos de uma maneira geral, a correta e ampla identificação dos *stakeholders* de um sistema de IA, em todo o seu ciclo de vida, tem sua realização defendida desde o início do projeto (Wright et al. 2018, De Silva & Alahakoon 2022).	Questionário - q 4.16 0; 33; 67; 100
Práticas para representação formal das regras e dilemas éticos	A necessidade de interpretação dos princípios éticos em cada caso específico de sistemas de IA, em linguagem compreensível por um perfil que não seja de desenvolvimento de sistemas, sugere a utilização de técnicas de representação formal de regras e de dilemas éticos, se estes existirem (Bench-Capon & Modgil; 2017; Bonnemains et al. 2018; Anderson & Anderson 2018; Zicari et al. 2021).	Questionário - q 4.17 0; 33; 67; 100
Práticas para mitigar vieses em sistemas de IA	O grande número de tipos de vieses impõe um cuidado especial em diferentes momentos ao longo do desenvolvimento do sistema de IA Práticas para mitigar tais vieses têm sido criadas e aperfeiçoadas (González et al 2020; Ashokan & Haas 2021; ISO 2021a; Baeza-Yates 2018; Oneto & Chiappa 2020); Makhlouf et al. 2021)	Questionário - q 4.18 0; 33; 67; 100
Transparência do processo de desenvolvimento de sistemas de IA	Além de documentação dos dados e do código, a necessidade de explicabilidade do resultado das decisões do sistema impõem conhecimento de todo o processo produtivo e resultados dos testes dos sistemas de IA (Arrieta et al 2020; AI HLEG 2019b, Das 2020; Dazeley et al 2021; Adadi & Berrada 2018; Phillips, et al. 2021).	Entrevista 0; 33; 67; 100
Monitoração automática	Quando em operação, a monitoração sistemática e automática do funcionamento do sistema de IA pode permitir o conhecimento de problemas não identificados durante o desenvolvimento, ou variações nos dados ou no ambiente que tenham descaracterizado o cenário idealizado para o funcionamento do sistema (Fjeld 2020; González et al 2020; De Silva & Alahakoon 2022; ISO 2022a).	Questionário - q 4.19 0; 33; 67; 100
Supervisão humana na busca por vieses	A supervisão humana dos resultados do sistema antes da tomada de decisões, ou após tais decisões, constituem uma medida de garantia da autonomia dos seres humanos em relação à tecnologia (Straub 2021; González et al 2020; Zicari et al. 2021; Dignum 2019; Hickman 2020).	Questionário - q 4.20 0; 33; 67; 100
Coleta de feedback dos usuários	A oferta de serviços digitais centrados no ser humano impõe a sua avaliação pelos seus usuários. Em sendo sistemas de IA, tal feedback oferece a oportunidade de se receber elementos associados à moralidade e à adequação do sistema aos propósitos para os quais foi projetado (Rahwan et al. 2019; Wright & Schultz 2018, Dignum 2019; Hickman 2020, de Almeida et al. 2021; AI HLEG 2019b; AI4People 2018).	Questionário - q 4.22 0; 33; 67; 100

Fonte: Elaboração própria.

Tabela 5: Variáveis-conjunto – características dos sistemas de IA da organização

Construto	**Variável-conjunto**	**Fonte**	**Opções de Pontuação**
CARACTERÍSTICAS DA ORGANIZAÇÃO E DOS SISTEMAS	Sistemas de IA contratados ou em parceria sem abertura de código	Questionário - q 3.1	0;1
	Sistemas de IA contratados ou em parceria com abertura de código	Questionário - q 3.1	0;1
	Sistemas de IA desenvolvidos internamente à organização	Questionário - q 3.1	0;1
	Dados internos à organização	Questionário - q 3.3	0;1
	Dados externos à organização	Questionário - q 3.3	0;1
	Organização tem acesso ao código de, pelo menos, 80% dos sistemas apresentados.	Questionário - q 3.1	0;1
	A organização iniciou o 1º projeto de sistemas de IA há mais de 3 anos.	Questionário - q 3.1	0;1
	Somente usuários internos acessam os sistemas de IA	Questionário - q 3.1	0;1

Fonte: Elaboração própria.

Tabela 1: Valores coletados do construto “Fatores Geradores de Expectativas”, transcritos de modo binário.

Fatores Geradores de Expectativas na Sociedade

Organização	Legislação para IA EXP1	Legsilação para proteção de dados pessoais EXP2	Políticas e normativos governamentais sobre IA EXP3
1	0	1	0
2	0	1	0
3	0	1	1
4	0	1	1
5	0	1	0
6	0	1	1
7	0	1	1
8	0	1	1
9	0	1	1
10	0	1	0
11	0	1	0
12	0	1	0
13	1	1	1
14	0	1	1
15	0	1	0
16	0	1	1
17	0	1	1
18	0	1	0
19	0	1	0
20	0	1	0
21	0	1	0
22	0	1	0
23	0	1	1
24	0	1	1
25	0	1	1
26	0	1	1
27	0	1	0
28	0	1	1

Fonte: Elaboração própria.

Tabela 2: Valores coletados do construto “Governança de IA” para variáveis do nível estratégico que formalizam a governança de IA.

Governança de IA

Ações Estratégicas

Organização	Estratégia de IA	Política ou norma para Sistemas de IA	Código ou Guia de Princípios Éticos	Processo para Governança de IA	Estrutura, Pessoa Responsável, Comitê para Governança de IA
	GOV1	GOV2	GOV3	GOV4	GOV5
1	67	33	0	0	0
2	67	67	0	100	100
3	100	100	100	100	100
4	0	0	67	0	0
5	0	67	100	100	67
6	67	67	67	67	100
7	33	67	0	0	0
8	100	67	33	67	100
9	67	67	100	100	67
10	0	0	0	0	100
11	0	0	67	0	33
12	0	0	0	0	0
13	100	100	100	100	100
14	67	67	100	0	67
15	0	67	0	100	0
16	33	33	33	0	67
17	0	0	0	0	0
18	0	0	0	0	0
19	67	67	67	100	100
20	67	67	33	67	67
21	67	67	67	0	100
22	67	100	33	67	100
23	67	67	67	0	0
24	33	0	100	0	0
25	67	0	67	0	67
26	67	67	100	0	67
27	67	0	100	100	100
28	33	33	33	100	33

Fonte: Elaboração própria.

Tabela 3: Valores coletados do construto "Governança de IA" para variáveis que alcançam o nível tático com foco em capacitação de pessoas.

	Governança de IA			
	Treinamento			
Organização	Treinamento para Tomadores de Decisões	Treinamento para Desenvolvedores de Sistemas de IA	Treinamento e/ou Campanhas de Comunicação para Usuários	Treinamento para Auditores Internos
	TDECISOR	TDESENV	TUSUARIO	TAUDITOR
1	0	0	67	0
2	33	33	67	0
3	0	0	0	0
4	0	0	67	67
5	67	33	100	67
6	33	33	67	67
7	67	0	67	67
8	67	100	100	100
9	67	67	100	67
10	0	0	100	0
11	67	67	67	67
12	0	0	0	0
13	100	100	100	100
14	0	67	67	0
15	0	0	67	67
16	67	0	0	0
17	0	0	0	0
18	0	0	100	0
19	0	67	0	0
20	67	67	0	33
21	33	33	33	67
22	100	100	100	100
23	0	0	0	0
24	67	100	67	100
25	67	67	0	0
26	67	67	0	0
27	100	100	100	0
28	67	67	67	0

Fonte: Elaboração própria

Tabela 4: Valores coletados do construto "Processos Auxiliares" para variáveis que representam processos auxiliares à governança de IA.

	Processos e Práticas Auxiliares					
Organização	Governança de Dados	Gestão da Qualidade de Dados	Gestão da Proteção de Dados Pessoais	Gestão de Riscos de Sistemas de IA	Gestão de Segurança de Sistemas de IA	Auditoria Interna nos Sistemas de IA
	PROGOVD	PROQUAD	PRODPES	PRORISC	PROSEG	AUDIT
1	33	0	33	0	33	0
2	67	67	67	67	67	0
3	100	100	100	100	67	100
4	100	67	100	0	0	0
5	100	67	100	33	100	67
6	67	67	100	67	67	67
7	67	67	100	0	0	0
8	67	100	100	67	100	100
9	100	67	100	100	100	67
10	0	0	100	100	33	100
11	67	67	100	0	100	0
12	33	0	100	0	0	0
13	67	67	100	100	100	100
14	0	0	100	0	0	0
15	67	67	67	0	0	67
16	67	67	67	0	67	0
17	0	0	100	0	0	0
18	0	67	67	0	0	0
19	100	100	100	33	67	67
20	100	100	100	33	67	100
21	33	33	33	33	33	67
22	100	100	100	100	100	100
23	67	100	100	0	0	67
24	100	33	100	0	0	100
25	67	100	100	0	67	0
26	67	0	100	0	67	0
27	100	67	100	100	0	0
28	67	67	100	33	67	100

Fonte: Elaboração própria

Tabela 5: Valores coletados para o construto “Desenvolvimento e Sustentação de Sistemas de IA” para variáveis que alcançam o nível tático e operacional.

	Desenvolvimento e Sustentação de Sistemas de IA						
	Projeto do Sistema de IA				**Sistema de IA em Operação**		
Organização	Identificação dos *Stakeholders*	Representação Formal de Regras e Dilemas Éticos	Mitigação de Viéses	Transparência do Processo Produtivo	Monitoração Automática	Supervisão Humana	Coleta de Feedback dos Usuários
	STAKEH	DILEMA	PVIESES	PTRANSP	MONAUTO	SUPHUMAN	FEEDBACK
1	0	0	0	0	67	33	67
2	100	33	67	33	67	33	67
3	100	67	67	100	100	100	100
4	0	0	0	0	0	0	0
5	67	33	67	100	67	67	67
6	0	67	67	0	100	100	100
7	0	67	0	0	67	0	67
8	67	33	100	100	100	67	0
9	100	100	100	100	67	100	100
10	0	0	100	67	100	100	100
11	0	0	100	67	67	100	0
12	67	0	0	67	67	0	100
13	100	100	100	100	100	100	100
14	0	67	100	100	67	67	67
15	100	67	0	67	67	0	67
16	67	0	0	67	100	67	67
17	0	0	0	0	0	0	0
18	0	0	0	67	67	0	67
19	100	67	67	0	67	67	67
20	100	67	67	0	100	67	100
21	67	67	67	67	33	0	33
22	100	100	100	0	100	100	100
23	0	67	100	67	67	100	0
24	100	100	100	67	100	100	100
25	67	67	67	0	67	100	100
26	67	67	100	67	0	67	67
27	100	100	67	100	100	67	100
28	100	67	100	0	100	100	67

Fonte: Elaboração própria

Tabela 6: Estatística descritiva das variáveis-conjunto coletadas – valores contínuos

Variável-conjunto	máx	méd	min	Não implantou, nem planeja	Não implantaou, mas, planeja implantar	Implantou em menor parcela	Implantou em maior parcela
Estratégia de IA	100	47	0	28.57%	14.29%	46.43%	10.71%
Política para sistema de IA	100	45	0	32.14%	10.71%	46.43%	10.71%
Código ou guia de princípios éticos	100	51	0	28.57%	17.86%	25.00%	28.57%
Processo para Governança de IA	100	42	0	53.57%	0.00%	14.29%	32.14%
Estrutura, comitê ou pessoa responsável pela IA	100	55	0	32.14%	7.14%	25.00%	35.71%
Proc. para governança de dados	100	64	0	14.29%	10.71%	42.86%	32.14%
Proc. para gestão da qualidade de dados	100	58	0	21.43%	7.14%	46.43%	25.00%
Proc. Para gestão da proteção de ados pessoais	100	91	33	0.00%	7.14%	14.29%	78.57%
Proc. para gestão de riscos	100	35	0	50.00%	17.86%	10.71%	21.43%
Proc. para gestão de segurança	100	47	0	35.71%	10.71%	32.14%	21.43%
Proc. Para auditoria nos sistemas de IA	100	45	0	46.43%	0.00%	25.00%	28.57%
Treinamento para tomadores de decisão	100	41	0	39.29%	10.71%	39.29%	10.71%
Treinamento para desenvolvedores de sistemas de IA	100	42	0	39.29%	14.29%	28.57%	17.86%
Treinamento para usuários	100	54	0	32.14%	3.57%	35.71%	28.57%
Treinamento para auditores	100	35	0	53.57%	3.57%	28.57%	14.29%
Identificação de stakeholders	100	56	0	35.71%	0.00%	25.00%	39.29%
Representação de regras e dilemas éticos	100	50	0	28.57%	10.71%	42.86%	17.86%
Práticas para minimizar vieses	100	61	0	28.57%	0.00%	32.14%	39.29%
Monitoração automática	100	72	0	10.71%	3.57%	46.43%	39.29%
Supervisão humana	100	61	0	25.00%	7.14%	28.57%	39.29%
Monitoração de mudanças de ambiente	100	54	0	28.57%	7.14%	39.29%	25.00%
Coleta de feedback	100	67	0	17.86%	3.57%	39.29%	39.29%
Práticas para transparência	100	50	0	35.71%	3.57%	35.71%	25.00%

Fonte: Elaboração própria

Tabela 7: Estatística descritiva das variáveis-conjunto coletadas – valores dicotômicos

Variável-conjunto	Organização está sujeita	Organização não está sujeita
Lei para IA	3.57%	96.43%
Lei para proteção de dados pessoais	100.00%	0.00%
Política governamental	53.57%	46.43%

Fonte: Elaboração própria

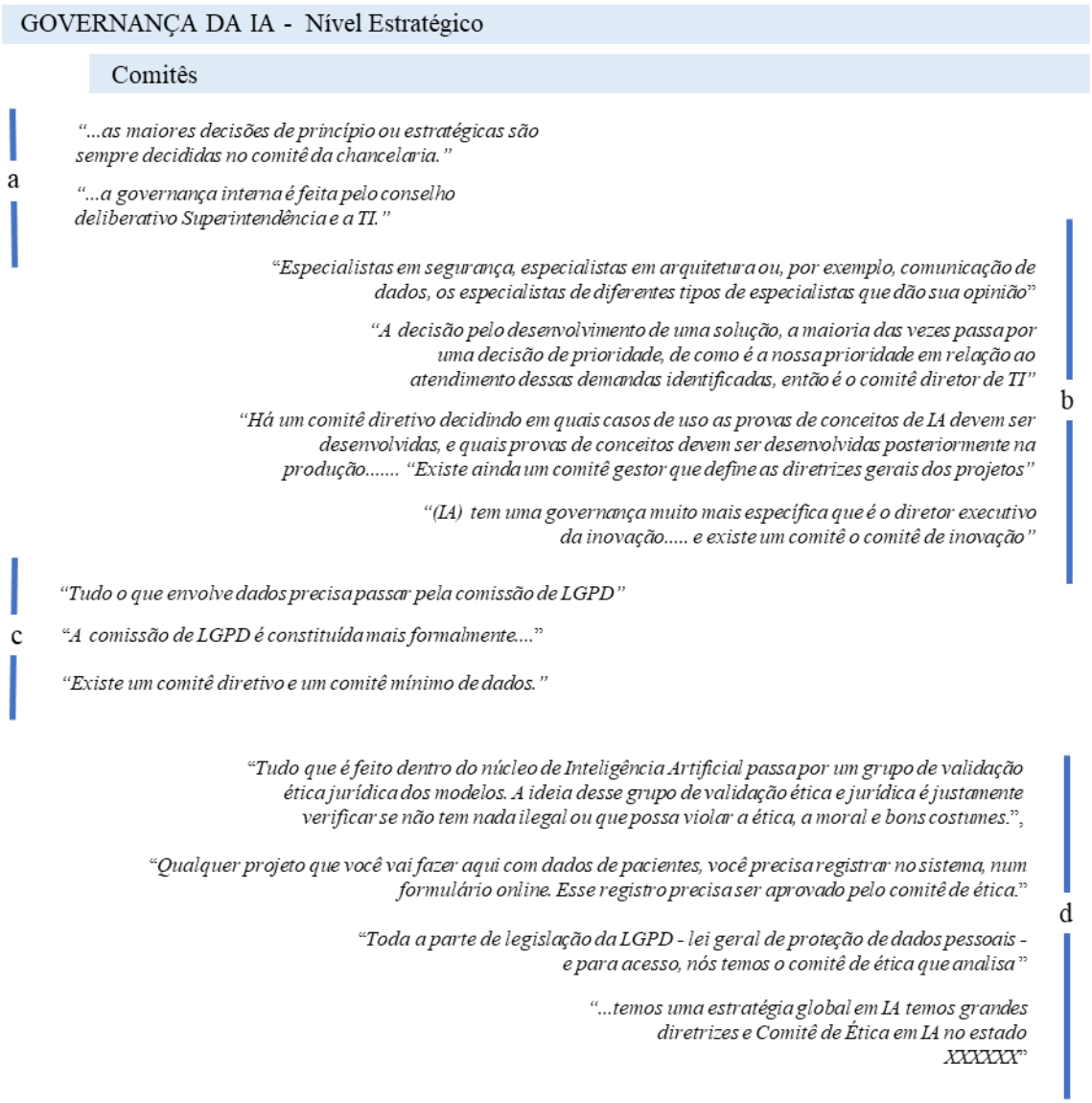

Figura 1: Trechos de entrevistas e documentos sobre comitês
Fonte: Elaboração própria

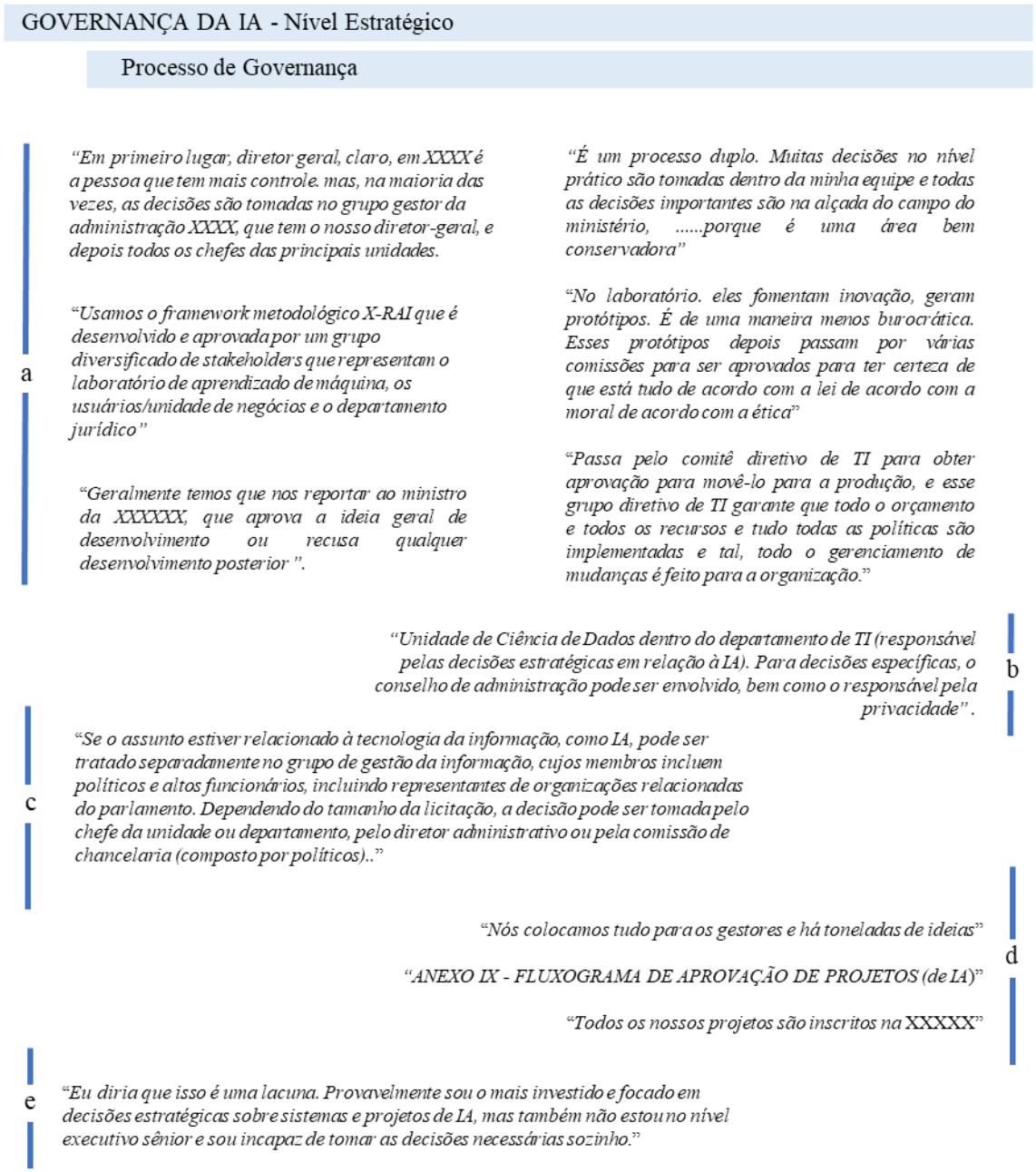

Figura 2: Trechos de entrevistas e documentos sobre processo de governança de IA
Fonte: Elaboração própria

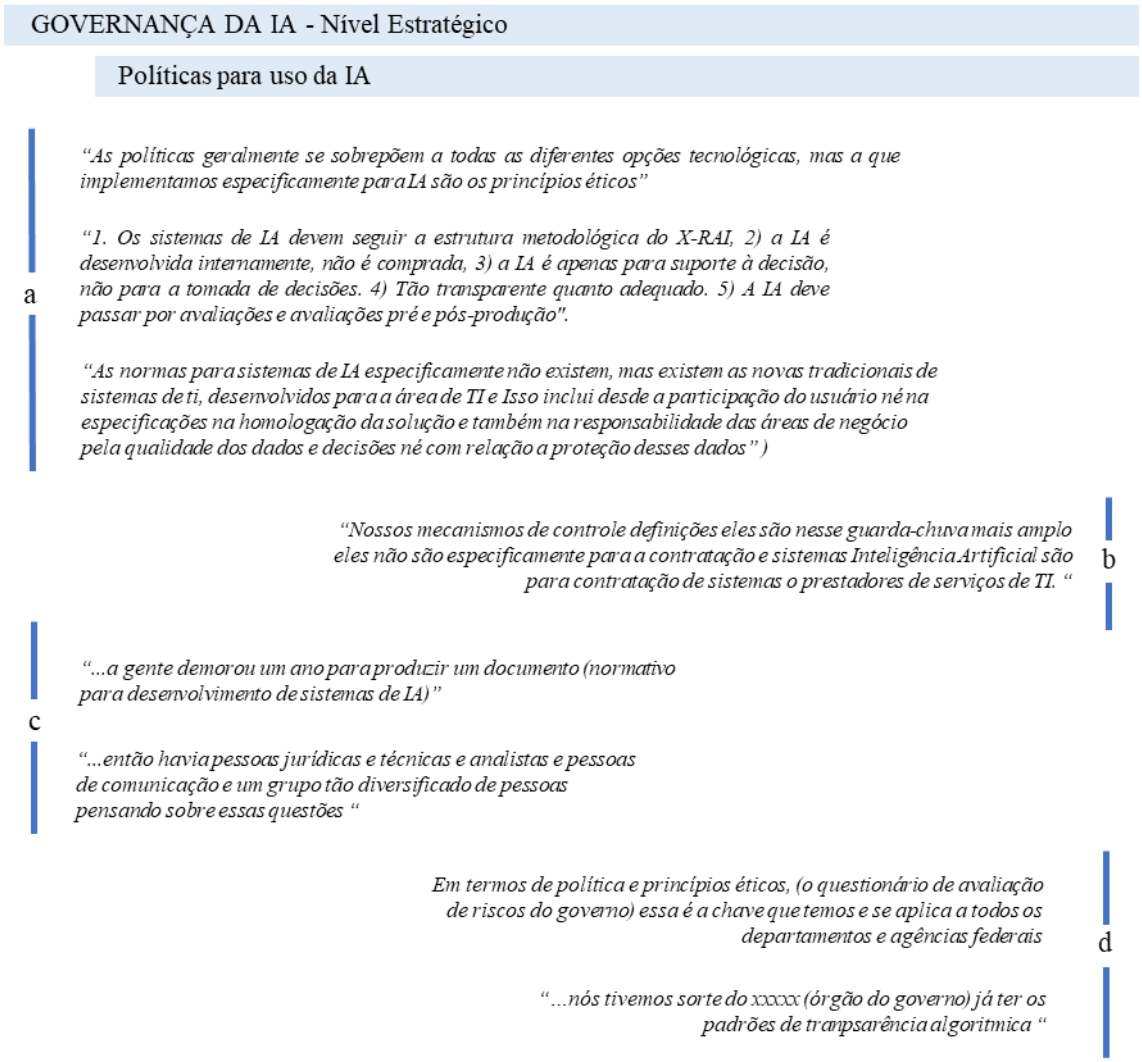

Figura 3: Trechos de entrevistas e documentos sobre políticas para uso da IA
Fonte: Elaboração própria

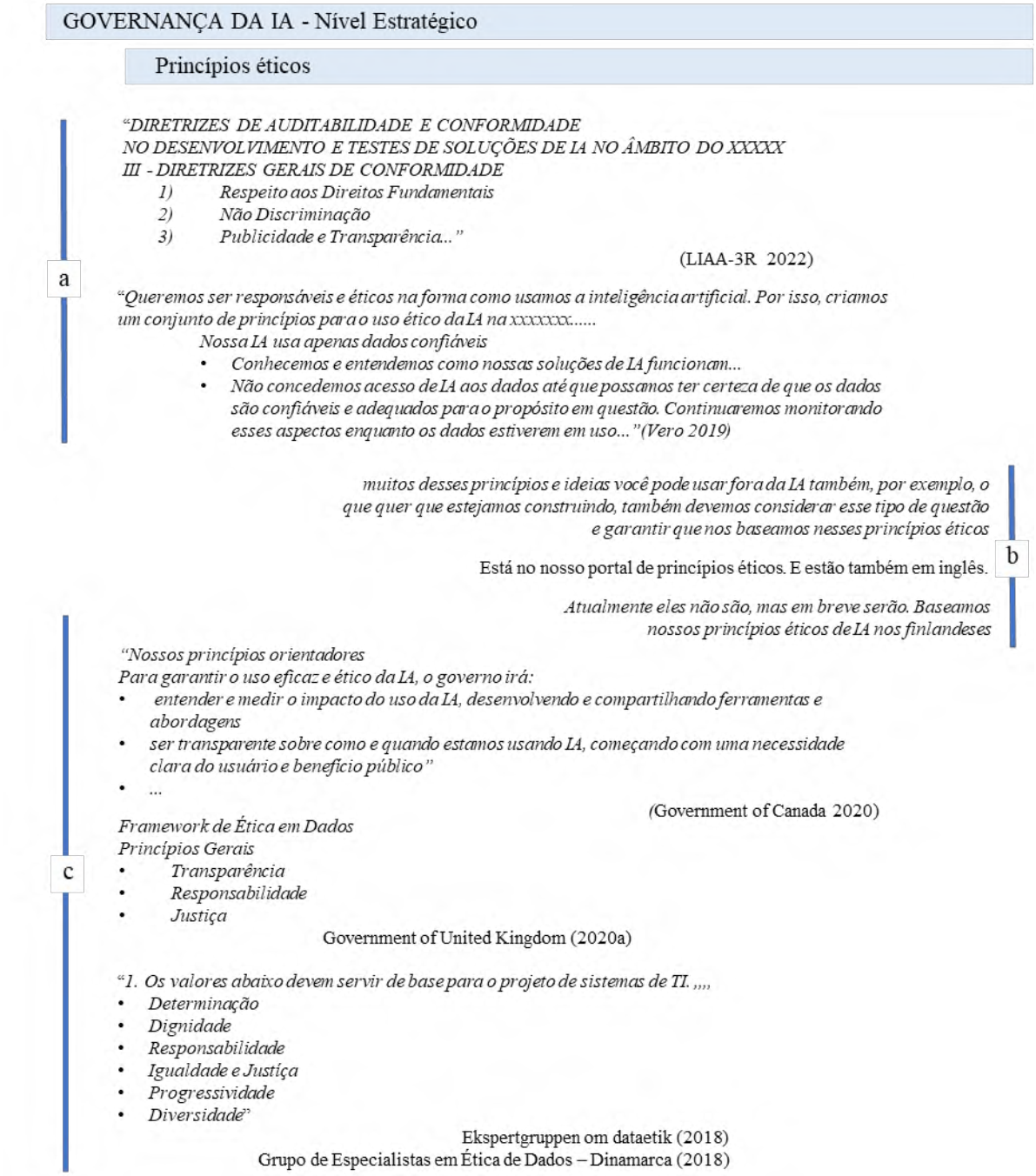

Figura 4: Trechos de entrevistas e documentos sobre princípios éticos
Fonte: Elaboração própria

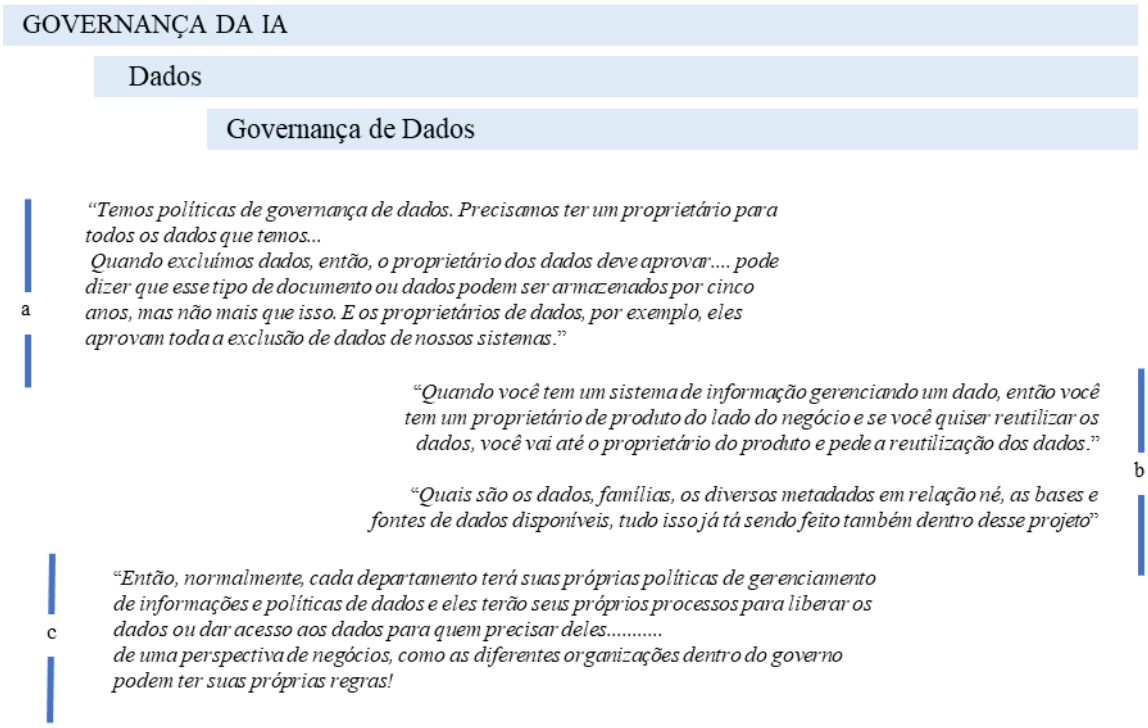

Figura 5: Trechos de entrevistas e documentos sobre governança de dados
Fonte: Elaboração própria

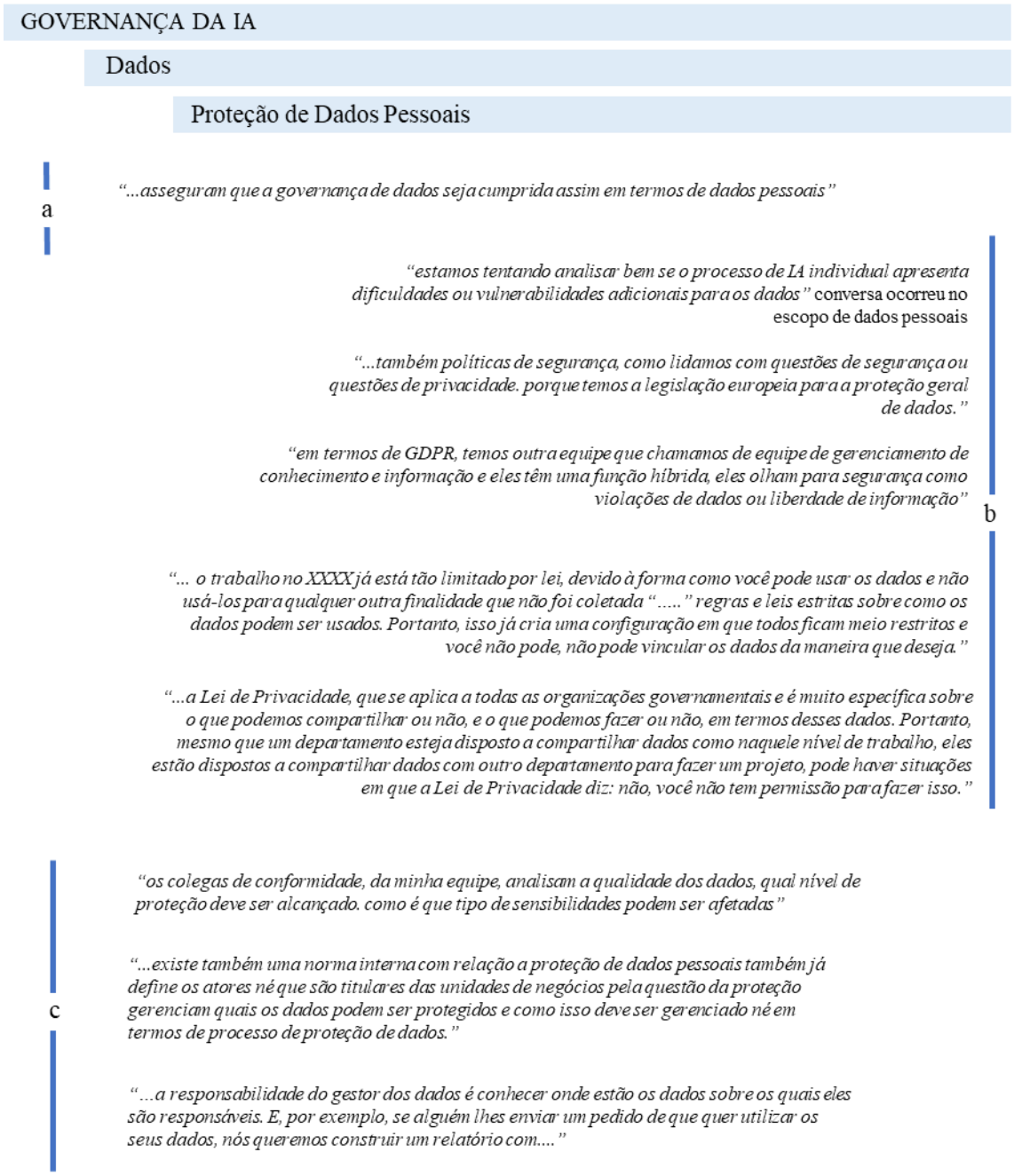

Figura 6: Trechos de entrevistas e documentos sobre proteção de dados pessoais
Fonte: Elaboração própria

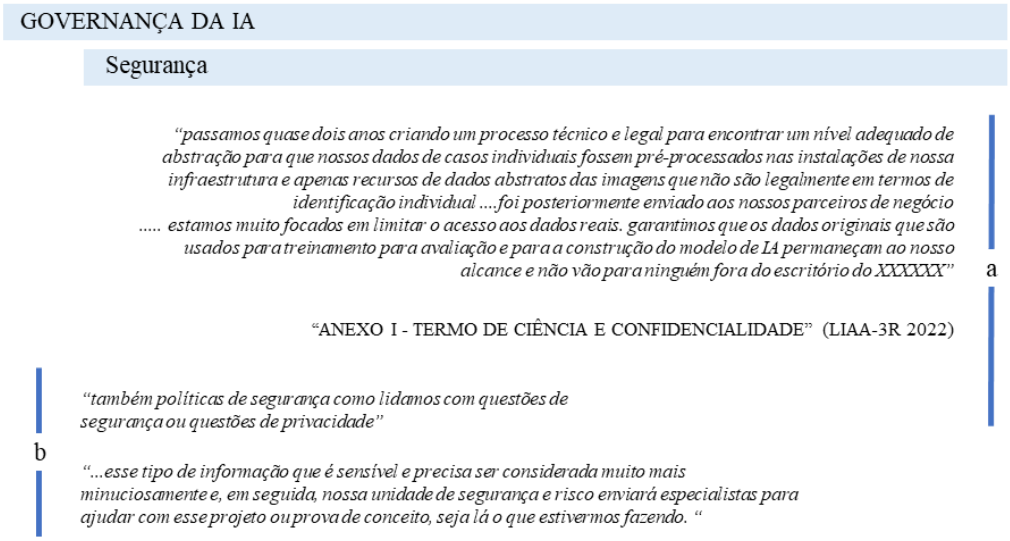

Figura 7: Trechos de entrevistas e documentos sobre segurança
Fonte: Elaboração própria

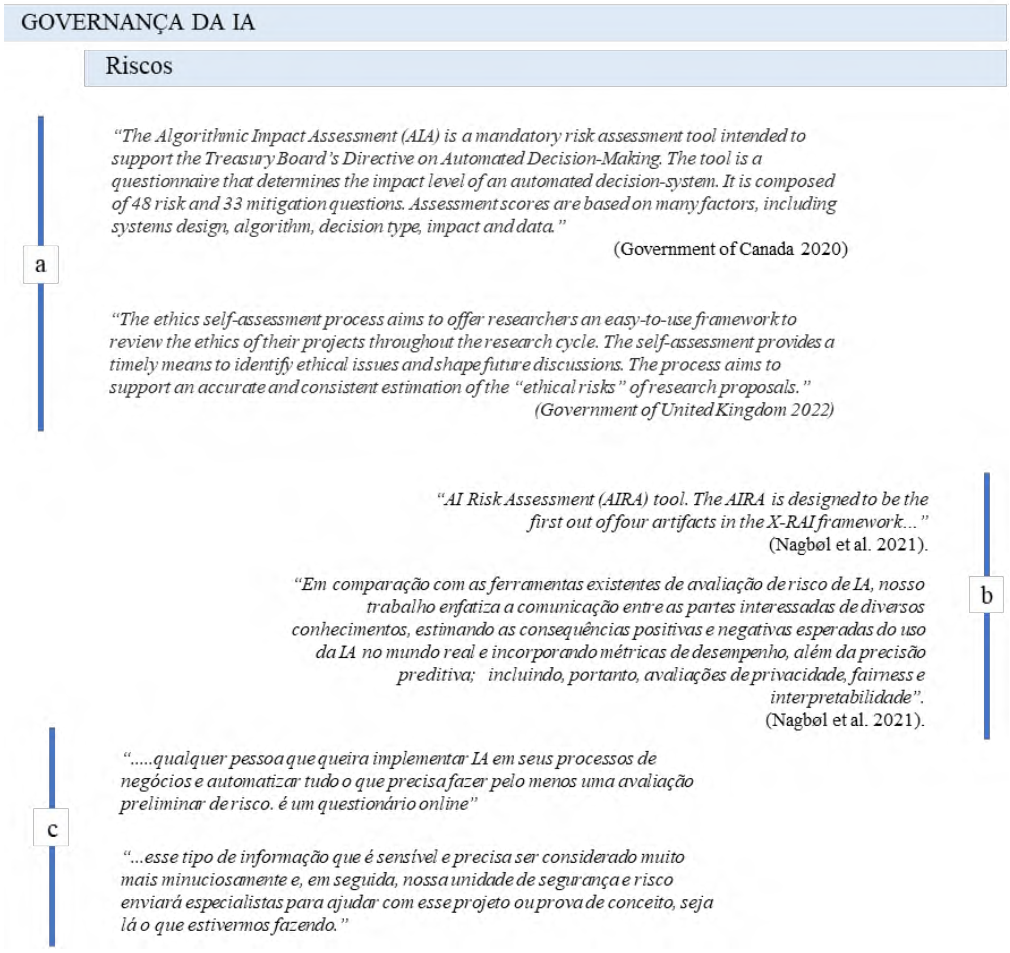

Figura 8: Trechos de entrevistas e documentos sobre riscos
Fonte: Elaboração própria

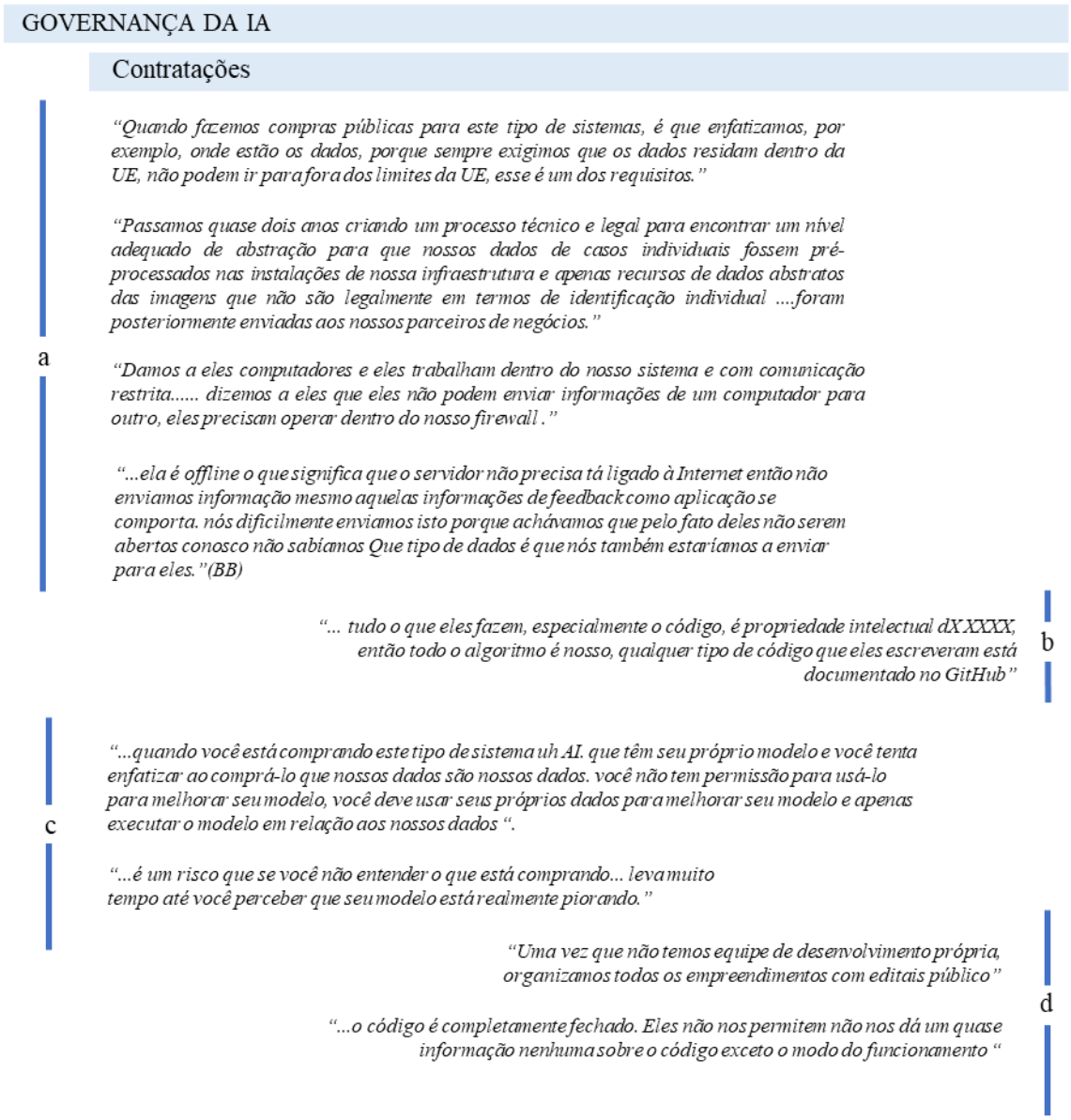

Figura 9: Trechos de entrevistas e documentos sobre contratações
Fonte: Elaboração própria

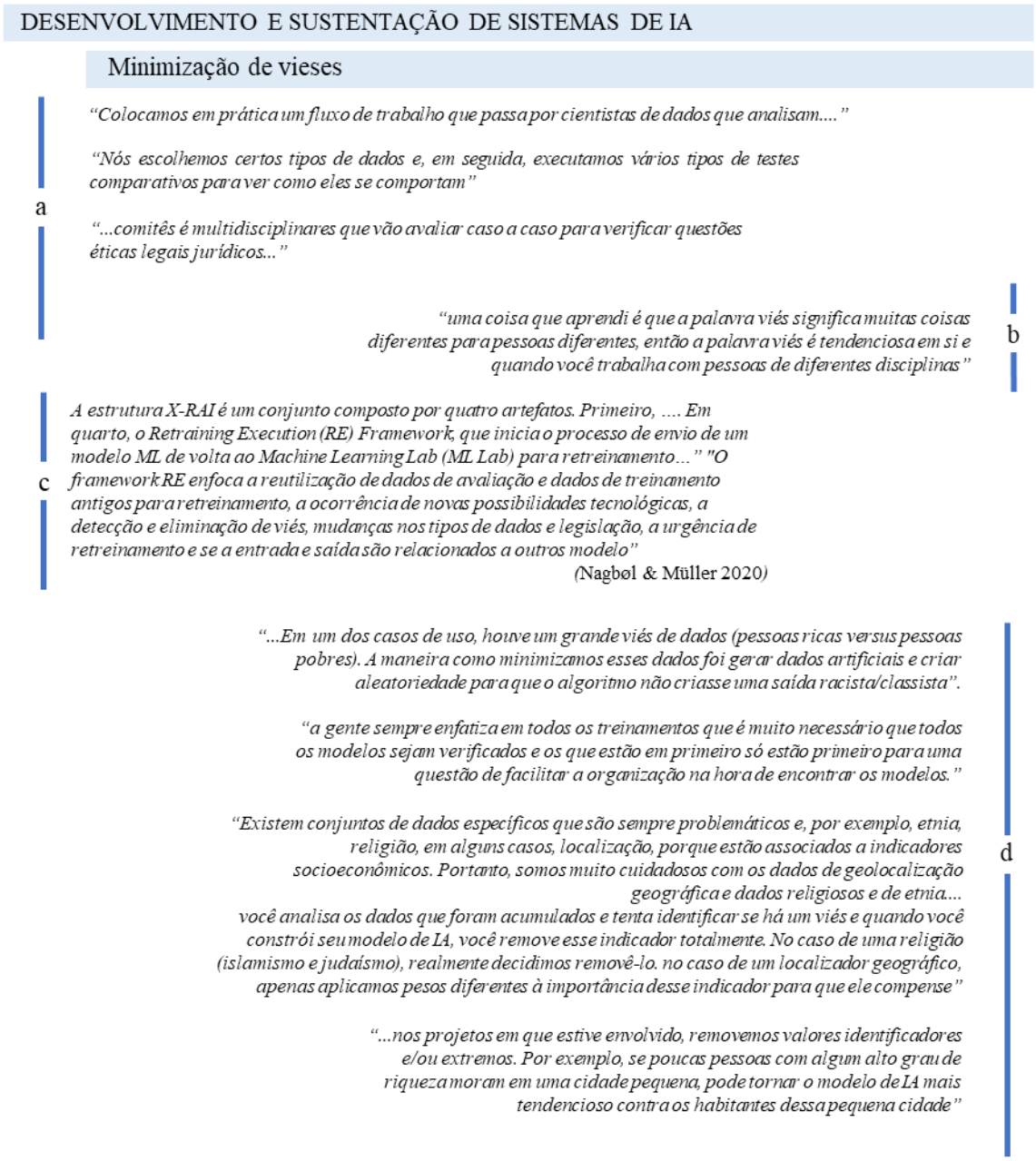

Figura 10: Trechos de entrevistas e documentos sobre minimização de vieses
Fonte: Elaboração própria

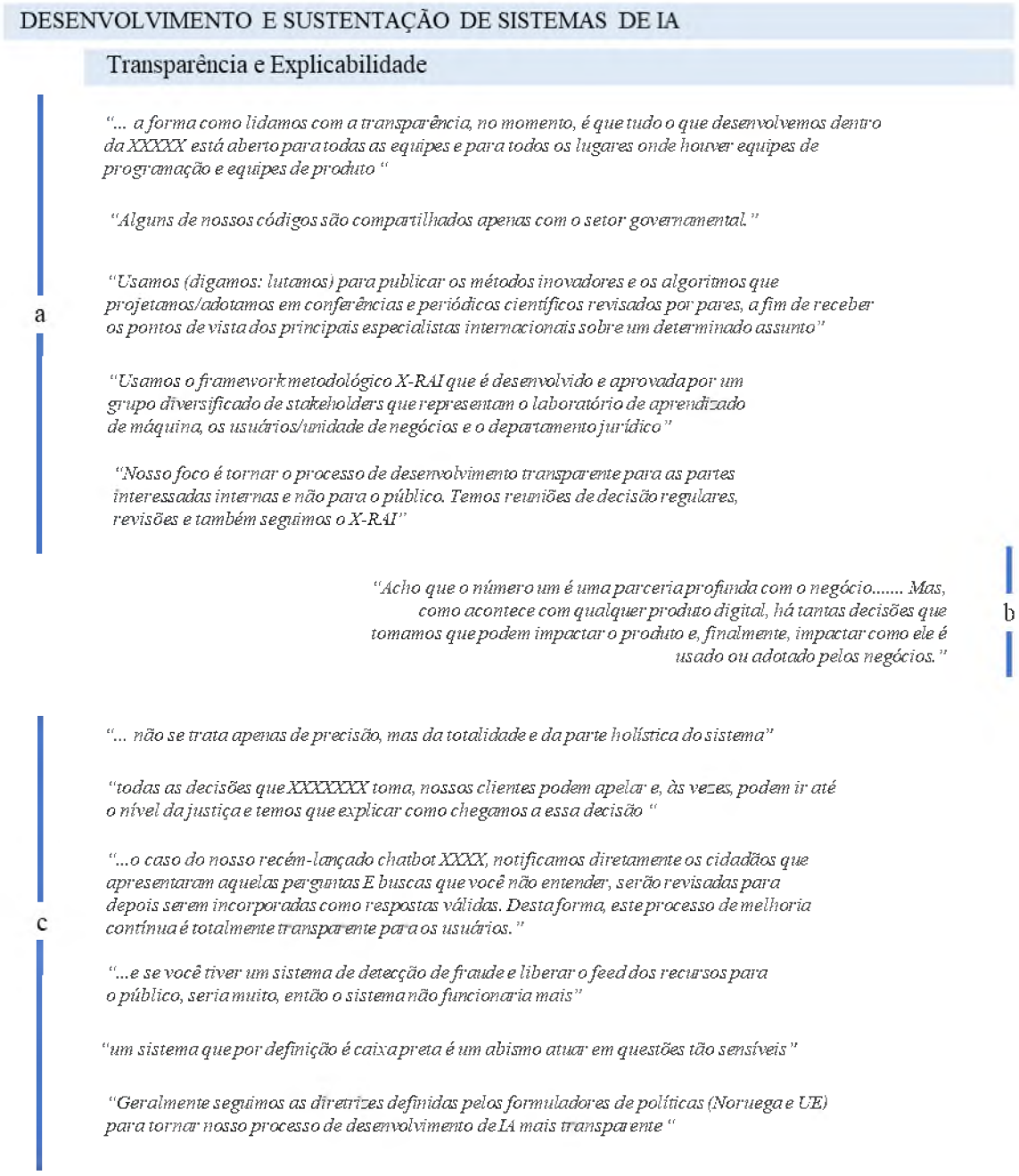

Figura 11: Trechos de entrevistas e documentos sobre transparência

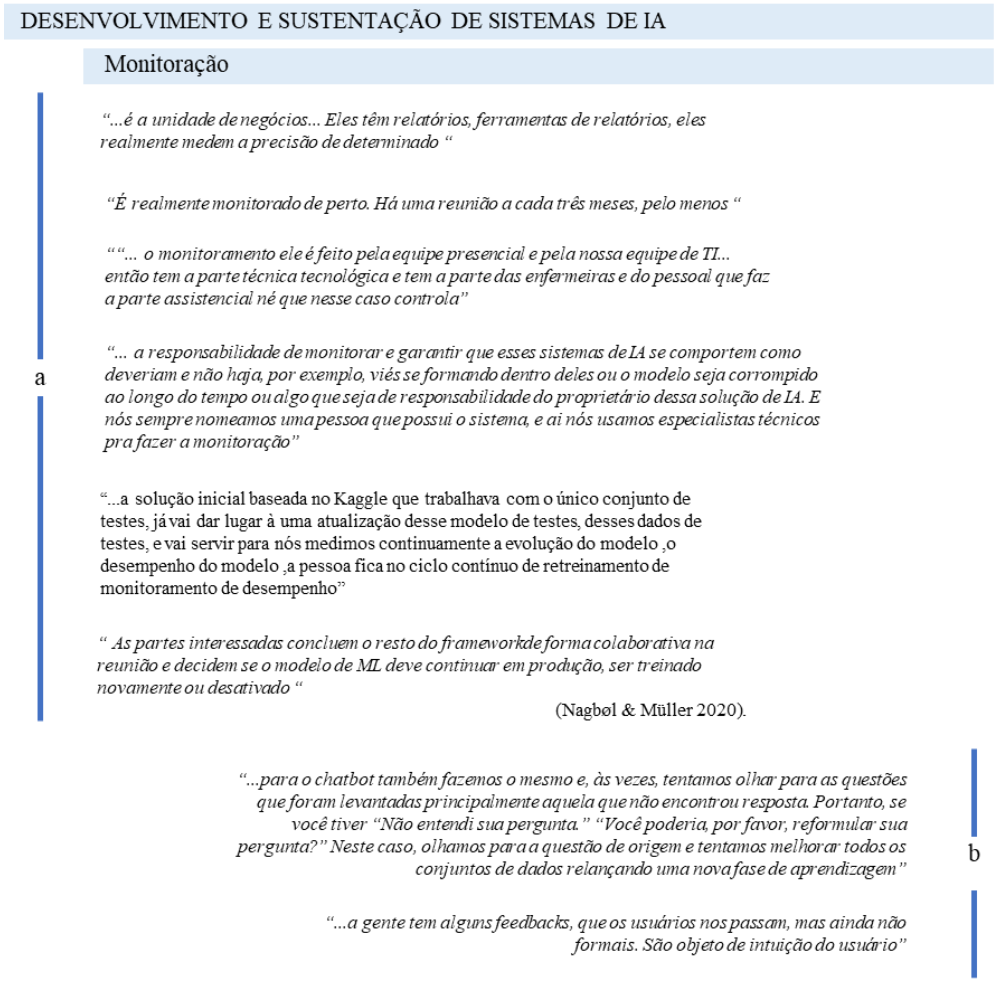

Figura 12: Trechos de entrevistas e documentos sobre monitoração

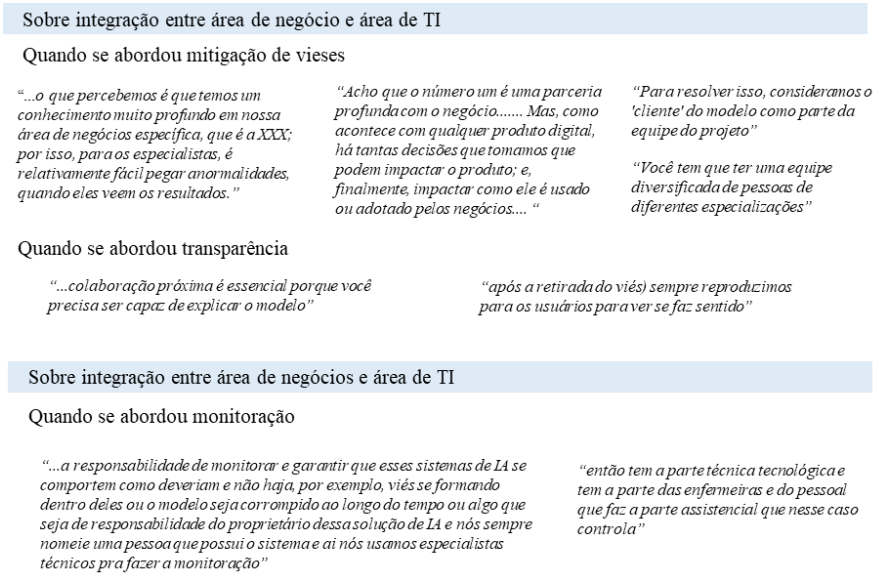

Figura 13: Trechos de entrevistas e documentos sobre integração sobre área de TI e área de negócio – vieses, transparência e monitoração

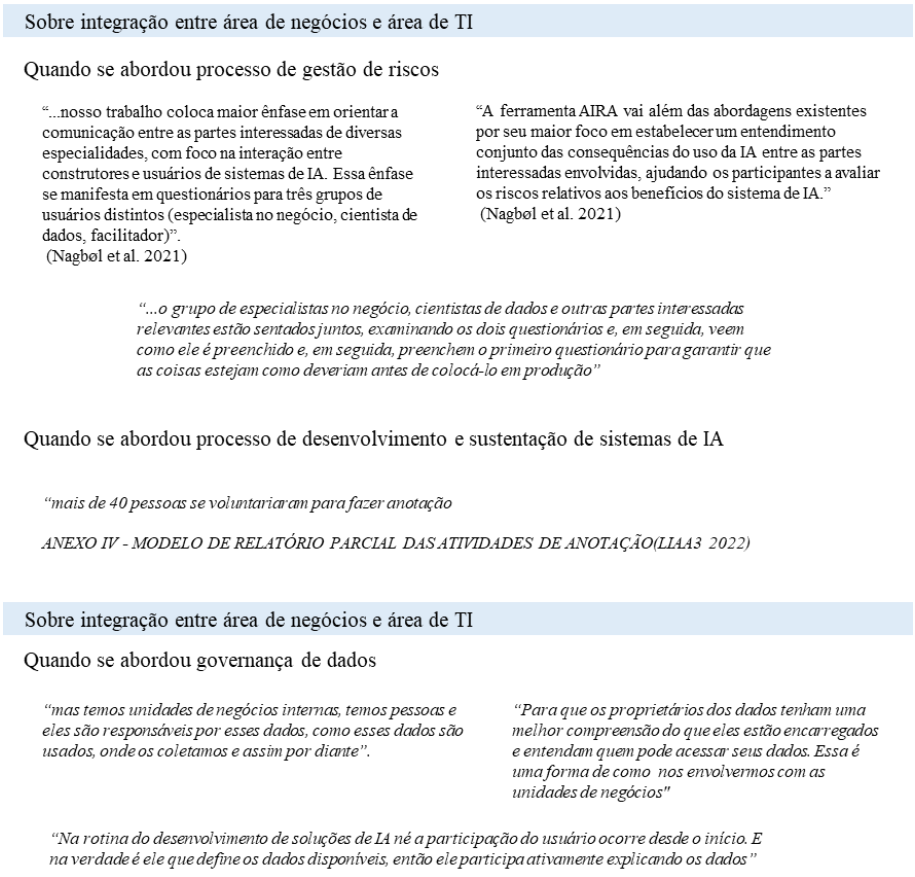

Figura 14: Trechos de entrevistas e documentos sobre integração sobre área de TI e área de negócio – riscos, desenvolvimento e governança de dados

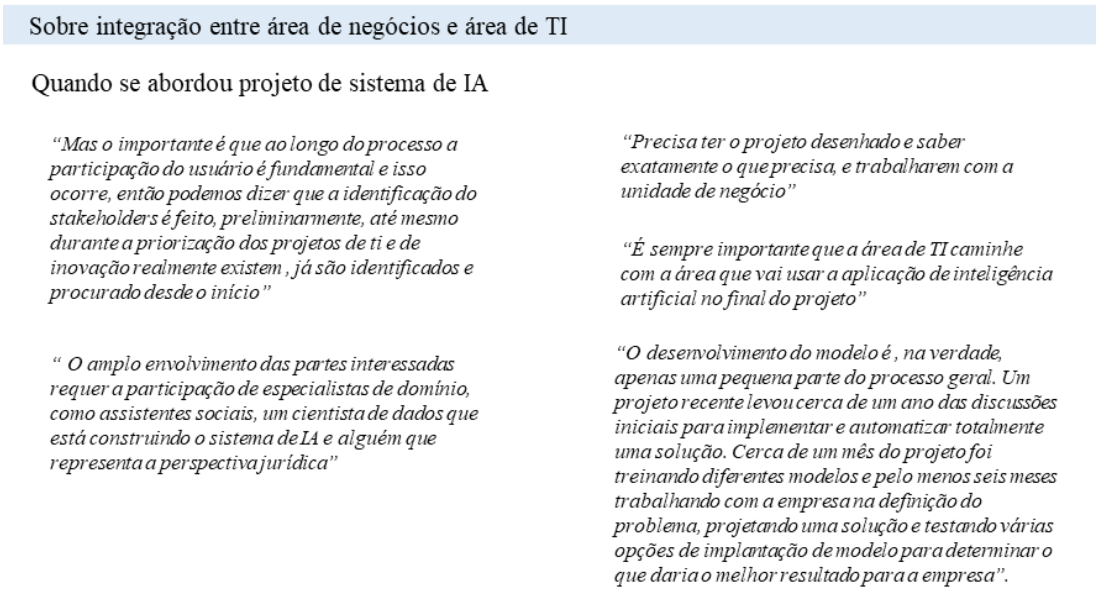

Figura 15: Trechos de entrevistas e documentos sobre integração sobre área de TI e área de negócio – projeto de sistema de IA