RCMOS – Revista Científica Multidisciplinar O Saber. ISSN: 2675-9128. São Paulo - SP.
Ano VII, v.1 2026 | submissão: 31/07/2026 | aceito: 03/08/2026 | publicação: 06/08/2026

Data Discovery as a strategic pillar for artificial intelligence governance programs: a historical-evolutionary analysis

Data Discovery como pilar estratégico para programas de governança de inteligência artificial: uma análise histórico-evolutiva

Data Discovery como pilar estratégico para programas de gobernanza de inteligencia artificial: un análisis histórico-evolutivo

Luciano da Silva Ferreira

Abstract: This article examines, from a historical-evolutionary perspective, the role of Data Discovery as a strategic pillar of Artificial Intelligence (AI) governance programs. It starts from the premise that the maturity of any algorithmic governance system depends directly on the organizational capacity to locate, classify, inventory, and understand the provenance of the data feeding machine learning models. The study traces the trajectory from traditional corporate data management, consolidated since the 1980s with the first data warehouses, to contemporary AI governance frameworks such as the NIST AI Risk Management Framework, ISO/IEC 42001:2023, and the European Artificial Intelligence Act (AI Act). A narrative literature review methodology is employed, with documentary analysis of technical standards, regulatory milestones, and indexed scientific publications, articulating concepts of data discovery, data lineage, automated cataloging, and machine learning metadata. Results indicate that Data Discovery has ceased to be a mere operational IT activity and has become a structuring requirement for compliance, explainability, and accountability in AI systems, serving as a precondition for auditability, risk mapping, and algorithmic bias mitigation. It is concluded that organizations that fail to invest in robust data discovery and cataloging capabilities face structural limitations in operationalizing ethical and regulatory AI principles, reinforcing the thesis that data governance is an inseparable foundation of algorithmic governance.

Keywords: Data Discovery. Data Governance. Artificial Intelligence Governance. Data Lineage. Regulatory Compliance.

Resumo: Este artigo examina, sob uma perspectiva histórico-evolutiva, o papel do Data Discovery como pilar estratégico dos programas de governança de Inteligência Artificial (IA). Parte-se da premissa de que a maturidade de qualquer sistema de governança algorítmica depende diretamente da capacidade organizacional de localizar, classificar, inventariar e compreender a proveniência dos dados que alimentam modelos de aprendizado de máquina. O estudo percorre a trajetória que vai da gestão tradicional de dados corporativos, consolidada a partir da década de 1980 com os primeiros data warehouses, até os frameworks contemporâneos de governança de IA, como o AI Risk Management Framework do NIST, a norma ISO/IEC 42001:2023 e o Regulamento Europeu de Inteligência Artificial (AI Act). Utiliza-se metodologia de revisão narrativa da literatura, com análise documental de normas técnicas, marcos regulatórios e publicações científicas indexadas, articulando conceitos de descoberta de dados, linhagem de dados (data lineage), catalogação automatizada e metadados de aprendizado de máquina. Os resultados indicam que o Data Discovery deixou de ser uma atividade operacional de TI para se converter em requisito estruturante de conformidade, explicabilidade e responsabilização (accountability) em sistemas de IA, funcionando como pré-condição para auditabilidade, mapeamento de riscos e mitigação de vieses algorítmicos. Conclui-se que organizações que não investem em capacidades robustas de descoberta e catalogação de dados enfrentam limitações estruturais para operacionalizar princípios éticos e regulatórios de IA, reforçando a tese de que a governança de dados constitui alicerce indissociável da governança algorítmica.

Palavras-chave: Data Discovery. Governança de Dados. Governança de Inteligência Artificial. Linhagem de Dados. Conformidade Regulatória.

Resumen: Este artículo examina, desde una perspectiva histórico-evolutiva, el papel del Data Discovery como pilar estratégico de los programas de gobernanza de Inteligencia Artificial (IA). Se parte de la premisa de que la madurez de cualquier sistema de gobernanza algorítmica depende directamente de la capacidad organizacional para localizar, clasificar, inventariar y comprender la procedencia de los datos que alimentan los modelos de aprendizaje automático. El estudio recorre la trayectoria que va desde la gestión tradicional de datos corporativos, consolidada a partir de la década de 1980 con los primeros data warehouses, hasta los marcos contemporáneos de gobernanza de IA, como el AI Risk Management Framework del NIST, la norma ISO/IEC 42001:2023 y el Reglamento Europeo de Inteligencia Artificial (AI Act). Se utiliza una metodología de revisión narrativa de la literatura, con análisis documental de normas técnicas, hitos regulatorios y publicaciones científicas indexadas, articulando conceptos de descubrimiento de datos, linaje de datos (data lineage), catalogación automatizada y metadatos de aprendizaje automático. Los resultados indican que el Data Discovery dejó de ser una actividad operativa de TI para convertirse en un requisito estructurante de cumplimiento, explicabilidad y rendición de cuentas (accountability) en sistemas de IA, funcionando como condición previa para la auditabilidad, el mapeo de riesgos y la mitigación de sesgos algorítmicos. Se concluye que las organizaciones que no invierten en capacidades robustas de descubrimiento y catalogación de datos enfrentan limitaciones estructurales para operacionalizar principios éticos y regulatorios de IA, lo que refuerza la tesis de que la gobernanza de datos constituye un fundamento indisociable de la gobernanza algorítmica.

Palabras clave: Data Discovery. Gobernanza de Datos. Gobernanza de Inteligencia Artificial. Linaje de Datos. Cumplimiento Regulatorio.

1 INTRODUCTION

The rise of Artificial Intelligence (AI) systems as critical decision-making infrastructure in sectors such as health care, finance, justice, education, and public safety has imposed on organizations a challenge once confined to technical information technology departments: knowing precisely what data exist, where they are stored, their origin, their quality, and their legal suitability for use in algorithmic models. This challenge, referred to in the technical literature as Data Discovery, has ceased to be an ancillary database-mapping activity and has assumed a central position in AI governance programs implemented by companies, governments, and multilateral organizations.

This article starts from the hypothesis that mature, auditable, and ethically responsible AI governance cannot be built without a prior robust capacity for data discovery, classification, and tracking. This hypothesis is supported by the historical finding that the first major scandals and failures involving algorithmic systems—from discriminatory bias in credit-granting systems to racially differentiated failures in facial recognition—recurrently had as their root

cause a lack of visibility into the composition, provenance, and limitations of the datasets used to train the models.

From this perspective, the general objective of this study is to analyze, through a historical-evolutionary approach, how the concept and practice of Data Discovery developed from traditional corporate data management until they became established as a strategic pillar of contemporary AI governance frameworks. The specific objectives are: a) to reconstruct the historical trajectory of data governance and its relationship with the emergence of AI governance; b) to characterize Data Discovery technically and describe its main methodologies and tools; c) to discuss how data discovery relates to regulatory requirements such as the Brazilian General Data Protection Law (LGPD), the European General Data Protection Regulation (GDPR), and the European Artificial Intelligence Act (AI Act); and d) to identify challenges and future prospects for integrating data governance and algorithmic governance.

Methodologically, the study is characterized as qualitative, exploratory, and descriptive research, conducted through a narrative literature review and documentary analysis. The materials examined included scientific articles indexed in databases such as Scopus, Web of Science, and Google Scholar; international technical standards (including the ISO/IEC 27000 family and ISO/IEC 42001:2023); institutional frameworks (the NIST AI Risk Management Framework and the OECD AI Principles); and regulatory instruments (LGPD, GDPR, and the AI Act). This made it possible to reconstruct an evolutionary timeline that connects data governance and AI governance as historically and technically interdependent phenomena.

The article is organized into five sections, in addition to this introduction. The second section presents the theoretical foundation, addressing the origins of data governance, the historical trajectory of Data Discovery, and the regulatory milestones of AI governance. The third section specifically discusses the role of Data Discovery as a strategic pillar of AI governance, detailing the data provenance chain, regulatory compliance requirements, and contemporary discovery and cataloging technologies. The fourth section discusses challenges, limitations, and future prospects. Finally, the fifth section presents the study's concluding considerations.

2 THEORETICAL FOUNDATION

2.1 Data Governance: Origins and Conceptual Evolution

Data governance, as a formal discipline, has its roots in the evolution of relational database management systems (DBMS), popularized beginning in the 1970s through Edgar F. Codd's work on the relational model. However, it was only in the 1990s, with the proliferation of corporate data warehouses and the need to integrate data from multiple transactional systems, that systematic concern with the quality, standardization, and accountability of corporate data emerged. This period corresponds to what the information management literature conventionally calls the first generation of data governance, centered primarily on technical controls for quality and referential integrity.

The second generation of data governance, situated approximately between the late 1990s and the first decade of the 2000s, was marked by the formalization of methodological frameworks such as DAMA-DMBOK (Data Management Body of Knowledge), organized by the Data Management Association International. This framework consolidated ten knowledge areas—including data architecture, modeling, quality, security, metadata, and, centrally for this study, data discovery and cataloging—establishing a common vocabulary and reference practices for professionals and organizations. During this period, data governance remained largely confined to IT departments and business intelligence areas, with no direct connection to broader ethical or regulatory concerns.

The third generation, which began around 2010, was driven by two concurrent phenomena: the explosion of big data—characterized by the attributes of volume, velocity, and variety, following Doug Laney's seminal formulation—and the enactment of personal data protection regulations, most notably the European GDPR, approved in 2016 and effective from 2018, and subsequently the Brazilian LGPD, enacted in 2018 and effective since 2020. These regulatory instruments shifted data governance from a predominantly technical concern to an issue of legal compliance, fundamental rights, and corporate accountability, requiring organizations to be able to demonstrate, at any time, what personal data they possessed, where those data were stored, for what purpose they had been collected, and how long they would be retained.

It is precisely within this context of regulatory demands that Data Discovery became established as an autonomous technical discipline, because responding to data subject requests, regulatory audits, and security incidents made reliance solely on manual data inventories—historically outdated and incomplete in large organizations—unworkable.

2.2 Data Discovery: Historical Trajectory and Technological Evolution

RCMOS – Revista Científica Multidisciplinar O
Saber. ISSN: 2675-9128. São Paulo - SP.
Ano VII, v.1 2026 | submissão: 31/07/2026 | aceito: 03/08/2026 | publicação: 06/08/2026

The term Data Discovery is used in the technical literature in two complementary senses that should be distinguished. In data analytics and business intelligence, popularized by tools such as Tableau and Qlik throughout the 2000s, Data Discovery denotes the interactive and exploratory process of identifying patterns, trends, and correlations in datasets, typically carried out by business analysts without dependence on specialized IT teams. In data governance and information security—the meaning that centrally guides this article—Data Discovery refers to the systematic and increasingly automated process of locating, identifying, classifying, and cataloging data dispersed across systems, applications, structured and unstructured databases, cloud environments, and file repositories, in order to build a reliable inventory of the existence, nature, and sensitivity of those data.

Historically, the first generation of data discovery tools, during the 1990s and 2000s, operated through scans based on regular expressions and pattern dictionaries, capable of identifying structured formats such as identification document numbers, credit card numbers, and email addresses in relational databases. This approach, still used today as a foundational layer in various commercial solutions, had significant limitations when dealing with unstructured data, such as free-text documents, images, audio, and video, which came to represent an increasing share of corporate data volume beginning in the 2010s.

The second generation of tools, consolidated between 2012 and 2018, incorporated natural language processing (NLP) and supervised machine learning techniques for semantic data classification, making it possible to identify not only syntactic patterns but also the contextual meaning of fields and documents—for example, distinguishing a nine-digit number representing a tax identification document from a similarly formatted telephone number. This technological advance coincided with the maturation of big data platforms, such as the Hadoop ecosystem and, later, data lake architectures, which aggregated massive volumes of heterogeneous data without a predefined schema (the schema-on-read approach), making automated discovery an unavoidable operational necessity.

The third and current generation of Data Discovery, under consolidation since approximately 2019, is characterized by the integration of automated metadata cataloging, end-to-end data lineage, and, more recently, specific tracking of machine learning metadata—such as training datasets, model versions, hyperparameters, and performance metrics. This generation is driven both by the adoption of distributed cloud data architectures and by the need, imposed by AI governance frameworks, to reconstruct in an auditable manner the entire provenance chain of a data item, from its origin to its use in automated algorithmic decisions.

RCMOS – Revista Científica Multidisciplinar O
Saber. ISSN: 2675-9128. São Paulo - SP.
Ano VII, v.1 2026 | submissão: 31/07/2026 | aceito: 03/08/2026 | publicação: 06/08/2026

2.3 Artificial Intelligence Governance: Regulatory and Institutional Milestones

Artificial Intelligence governance, understood as the set of structures, policies, processes, and controls designed to ensure the responsible, safe, and ethical development and use of AI systems, is a considerably more recent phenomenon than data governance and became institutionally established mainly from 2019 onward. That year, the Organisation for Economic Co-operation and Development (OECD) published its AI Principles, adopted by dozens of member and non-member countries, setting guidelines for transparency, robustness, safety, and accountability that would influence virtually all subsequent national frameworks.

In 2021, the European Commission presented its proposal for the European Artificial Intelligence Regulation, known as the AI Act, formally approved in 2024 and entering into force in stages throughout 2025 and 2026. The AI Act adopts a risk-based approach, classifying AI systems into unacceptable-risk, high-risk, limited-risk, and minimal-risk categories. For high-risk systems, it imposes explicit obligations regarding technical documentation, governance of training, validation, and test data, and traceability (automatic logging), demonstrating a direct normative connection between data governance and algorithmic governance.

In parallel, the United States National Institute of Standards and Technology (NIST) published the AI Risk Management Framework (AI RMF) in 2023. Structured around four core functions—Govern, Map, Measure, and Manage—it explicitly requires organizations to be able to map the provenance, quality, and representativeness of data used in AI systems as a precondition for algorithmic risk management. During the same period, the International Organization for Standardization (ISO), together with the International Electrotechnical Commission (IEC), published ISO/IEC 42001:2023, the first certifiable AI management system standard, which devotes specific requirements to the quality and provenance of data used throughout the life cycle of AI systems.

In Brazil, the regulatory debate on AI advanced through Bill No. 2338/2023, under consideration by the National Congress, which proposes a legal framework for the use of Artificial Intelligence inspired largely by the risk-based approach of the European AI Act, as well as through the work of the National Data Protection Authority (ANPD), which already regulates, under the LGPD, the automated processing of personal data for decision-making purposes, including the right to review automated decisions.

The convergence of these regulatory frameworks reveals a consistent historical pattern: no AI governance framework, whether regulatory or technical and voluntary in nature,

dispenses with robust mechanisms for understanding the data underlying algorithmic systems. It is precisely this convergence that supports the central thesis of this article, developed in the following section.

3 DATA DISCOVERY AS A STRATEGIC PILLAR OF ARTIFICIAL INTELLIGENCE GOVERNANCE

3.1 The Provenance Chain: From Raw Data to Algorithmic Decision

Every Artificial Intelligence system, regardless of its technical architecture, results from a chain of transformations that begins with the collection or generation of raw data and culminates in an inference, prediction, or automated decision. This chain—which the technical literature calls the data provenance chain—typically involves the stages of collection, storage, integration, cleaning, labeling, training-sample selection, feature engineering, model training, validation, deployment, and continuous monitoring. At each of these stages, apparently neutral technical decisions can introduce distortions, biases, or losses of quality that propagate and intensify along the chain.

In this context, Data Discovery acts as an instrumentation layer that makes it possible to reconstruct this chain in an auditable manner. Without data discovery and cataloging mechanisms, an organization is unable to answer, accurately and promptly, questions that are central to AI governance, such as: which datasets fed a given model; whether those datasets contain sensitive personal data; whether training samples have adequate demographic representativeness; whether consent or another legal basis existed for using particular data; and whether a specific data item can be deleted or corrected at a data subject's request without compromising the integrity of the already trained model.

This inability to respond is not merely theoretical. Case studies widely discussed in the AI ethics literature—such as the COMPAS criminal recidivism risk-assessment system used in United States courts and commercial facial-recognition systems whose performance is significantly poorer for darker skin tones—share the fact that, at the time of development, their developers lacked sufficient visibility into the demographic composition and representativeness of the training data used. This visibility gap is, in essence, a Data Discovery failure: the absence of a reliable inventory that could have identified beforehand the sample imbalance that would later manifest as algorithmic discrimination in production.

3.2 Data Discovery and Regulatory Compliance

The relationship between Data Discovery and regulatory compliance can be observed in three central normative areas. First, the Brazilian LGPD and the European GDPR require organizations to maintain Records of Processing Activities, which presupposes up-to-date knowledge of what personal data exist, where they are stored, and for what purpose they are processed. In corporate environments with multiple legacy systems and cloud platforms, such knowledge is feasible at scale only through automated data discovery and classification tools.

Second, the European AI Act establishes, for high-risk systems, the obligation that training, validation, and test datasets be relevant, representative, as error-free as possible, and complete, taking into account the specific characteristics of the intended use context. Compliance with this obligation depends directly on data discovery and profiling mechanisms capable of measuring, quantitatively and in a documented manner, the completeness, representativeness, and quality of training data before and during the model life cycle.

Third, the NIST AI RMF and ISO/IEC 42001:2023 converge in requiring organizations to maintain detailed technical documentation on the origin, processing, and limitations of data used in AI systems, often materialized in artifacts such as datasheets for datasets—a structured dataset documentation model proposed in academic literature and progressively incorporated into corporate practices—and model cards, which document the performance and limitations of machine learning models across different population subgroups. Producing these artifacts is, in practice, impossible without mature Data Discovery processes, since it requires precise knowledge of the provenance, composition, and statistical characteristics of the underlying data.

3.3 Contemporary Data Discovery Technologies and Methods Applied to AI

Contemporary Data Discovery technology solutions focused on AI governance typically combine four core technical capabilities. The first is automated data cataloging, which consists of systematically scanning structured and unstructured data sources to build a centralized metadata repository, including schemas, data types, volumes, accountable owners, and sensitivity classifications. Contemporary cataloging platforms incorporate machine learning algorithms to infer semantic relationships automatically among data fields dispersed across different systems, reducing dependence on historically incomplete manual documentation.

The second capability is sensitive data classification, which applies pattern-recognition techniques, regular expressions, semantic dictionaries, and, more recently, language models to identify categories of personal and sensitive data according to specific regulatory taxonomies —such as health data, racial or ethnic origin, sexual orientation, and biometric data—categories that receive differentiated legal treatment under the LGPD, GDPR, and AI Act.

The third capability is data lineage, which consists of automatically tracking the flow of data from its origin to its consumption points, including all intermediate transformations. Contemporary lineage tools operate by instrumenting data pipelines (parsing SQL code, data-engineering scripts, and workflow orchestrators), making it possible to reconstruct automatically and visually the path followed by a specific data item. This capability is essential for incident investigations, regulatory audits, and root-cause analyses of algorithmic bias.

The fourth capability, more recently consolidated and specific to the AI context, is machine learning metadata tracking, implemented through MLOps platforms that record, in a versioned manner, which datasets were used in each training experiment, which transformations were applied, which hyperparameters were used, and which performance metrics—including algorithmic fairness metrics disaggregated by population subgroup—were obtained. This capability extends the traditional scope of Data Discovery, originally limited to data stored in repositories, to the AI model development life cycle itself, in a technical convergence that recent literature calls data-and-model governance.

3.4 Evolutionary Timeline: From Data Governance to Artificial Intelligence Governance

The historical synthesis developed throughout this section makes it possible to propose an evolutionary periodization connecting data governance, Data Discovery, and AI governance, organized into five successive and cumulative phases. The first phase, situated between the 1970s and 1980s, corresponds to the consolidation of relational database management systems and the emergence of rudimentary data quality controls, without systematic concern for discovery or cataloging.

The second phase, from 1990 to 2005, corresponds to the formalization of data governance as a management discipline, with the consolidation of data warehouses, frameworks such as DAMA-DMBOK, and the first discovery tools based on syntactic patterns, still limited to structured data and predominantly analytical business-intelligence purposes.

The third phase, from 2005 to 2016, is marked by the explosion of big data, the adoption of data lake architectures, and the incorporation of natural language processing and machine learning into discovery tools, expanding their ability to operate on unstructured data at scale.

The fourth phase, from 2016 to 2023, corresponds to the regulatory consolidation of personal data protection through the approval of the GDPR and LGPD, which transformed Data Discovery from an analytical convenience into a legal compliance obligation. At the same time, the OECD AI Principles of 2019 initiated the institutionalization of AI governance as an autonomous field.

The fifth and current phase, which began in 2023 and is still consolidating, is characterized by the explicit convergence of data governance and AI governance, materialized in ISO/IEC 42001:2023, the NIST AI RMF, and the European AI Act. All require advanced discovery, cataloging, lineage, and metadata-tracking capabilities as a technical precondition for regulatory compliance and algorithmic risk management. This periodization shows that Data Discovery did not emerge as a response to AI governance; rather, it was an already established discipline that was reconfigured and became strategically central in view of the specific needs imposed by contemporary algorithmic systems.

4 DISCUSSION: CHALLENGES, LIMITATIONS, AND FUTURE PROSPECTS

Despite the technological and regulatory advances described, the integration of Data Discovery and AI governance faces significant structural challenges. The first concerns the scale and heterogeneity of contemporary data environments, characterized by the coexistence of legacy systems, cloud platforms from multiple providers, software-as-a-service (SaaS) applications, and third-party data sources. This makes the construction of a truly complete and up-to-date inventory an asymptotic goal, rarely achieved in full even by the most mature organizations.

The second challenge concerns the technical limitations of automated discovery tools themselves when faced with unstructured multimodal data—such as combinations of text, image, audio, and video used to train contemporary foundation models—for which semantic classification methods still exhibit higher error rates than those observed in traditional structured data. This introduces residual uncertainty even in technically advanced governance programs.

The third challenge is organizational and cultural, and perhaps the most persistent: data governance and AI governance frequently reside in separate organizational structures—the

former in information technology and data protection areas, the latter in ethics committees, legal departments, or data science units. This creates communication gaps, duplication of effort, and, in extreme cases, data inventories and AI model inventories that do not communicate with one another, compromising the end-to-end traceability required by contemporary regulatory frameworks.

The fourth challenge concerns the tension between the granularity required for regulatory auditability and the computational and financial costs of maintaining data lineage and machine learning metadata at production scale, especially in organizations operating hundreds or thousands of models simultaneously, an increasingly common scenario with the proliferation of applications based on large language models (LLMs) and generative AI architectures.

As for future prospects, there is a trend toward technical convergence between data governance platforms and MLOps platforms, with the emergence of integrated solutions that treat data and models as artifacts in the same unified provenance chain—an approach recent literature calls unified data and AI governance. There is also growing interest in discovery and classification techniques based on large language models, which promise greater semantic accuracy in identifying sensitive data in unstructured contexts, although they introduce new challenges related to the explainability and reliability of the discovery tools themselves. Finally, it is reasonable to anticipate that future regulatory frameworks in Brazil and other jurisdictions will increasingly specify minimum technical requirements for data discovery, cataloging, and lineage as a condition of compliance for high-risk AI systems, definitively consolidating Data Discovery as critical—not ancillary—infrastructure for algorithmic governance.

5 CONCLUDING CONSIDERATIONS

This article sought to demonstrate, through a historical-evolutionary analysis, that Data Discovery is an inseparable strategic pillar of contemporary Artificial Intelligence governance programs. Reconstructing the trajectory from the consolidation of relational database management systems in the 1970s and 1980s to the regulatory convergence embodied in the European AI Act, the NIST AI Risk Management Framework, and ISO/IEC 42001:2023 reveals a consistent historical pattern: the capacity to locate, classify, and track data provenance has always technically preceded the capacity to govern systems built from those data.

The main contribution of this study lies in systematizing a five-phase evolutionary periodization that connects concepts previously treated in relatively fragmented ways in the

literature—data governance, Data Discovery, and AI governance—and demonstrates their technical and regulatory interdependence. Based on this periodization, it is argued that organizations that do not invest in mature data discovery, cataloging, and lineage capabilities face structural, not merely operational, limitations in putting ethical principles into practice and meeting AI governance regulatory requirements, because compliance with representativeness, quality, explainability, and auditability requirements for algorithmic systems presupposes prior and reliable knowledge of the underlying data.

A limitation of this study is its predominantly theoretical and documentary character, without direct empirical validation through in-depth organizational case studies or field research. As an agenda for future research, empirical studies are suggested to measure, in organizations of different sectors and sizes, the correlation between the level of Data Discovery maturity and effective indicators of compliance and risk mitigation in AI governance programs, as well as in-depth investigation of unified governance models for data and machine learning models, a still-emerging field with high potential for academic and practical development in the coming years.

REFERENCES

BRAZIL. Law No. 13,709 of August 14, 2018. General Personal Data Protection Law (LGPD). Brasília: Official Gazette of the Union, 2018.

BRAZIL. Bill No. 2338 of 2023. Provides for the use of Artificial Intelligence. Brasília: Federal Senate, 2023.

DAMA INTERNATIONAL. DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. Denville: Technics Publications, 2017.

EUROPEAN COMMISSION. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Brussels: Official Journal of the European Union, 2024.

EUROPEAN PARLIAMENT; COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2016/679 of April 27, 2016 (General Data Protection Regulation—GDPR). Brussels: Official Journal of the European Union, 2016.

FLORIDI, L.; COWLS, J. A unified framework of five principles for AI in society. Harvard Data Science Review, Cambridge, v. 1, n. 1, 2019.

GEBRU, T. et al. Datasheets for datasets. Communications of the ACM, New York, v. 64, n. 12, p. 86-92, 2021.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 27001:2022—Information security, cybersecurity and privacy protection—Information security management systems—Requirements. Geneva: ISO, 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO/IEC 42001:2023—Information technology—Artificial intelligence—Management system. Geneva: ISO, 2023.

JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of AI ethics guidelines. Nature Machine Intelligence, London, v. 1, n. 9, p. 389-399, 2019.

KITCHIN, R. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. London: Sage Publications, 2014.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity and Variety. Stamford: META Group Research Note, 2001.

MITCHELL, M. et al. Model cards for model reporting. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2019, Atlanta. Proceedings [...]. New York: ACM, 2019. p. 220-229.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. AI Risk Management Framework (AI RMF 1.0). Gaithersburg: NIST, 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. Paris: OECD, 2019.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A Modern Approach. 4th ed. Rio de Janeiro: GEN LTC, 2022.

ZUBOFF, S. The Age of Surveillance Capitalism. Rio de Janeiro: Intrínseca, 2021.