Media and Communication
2025 • Volume 13 • Article 9419
https://doi.org/10.17645/mac.9419

cogitatio

ARTICLE

Open Access Journal

AI Transparency: A Conceptual, Normative, and Practical Frame Analysis

Sónia Pedro Sebastião [1] and David Ferreira-Mendes Dias [2]

[1] Centro de Administração e Políticas Públicas, Instituto Superior de Ciências Sociais e Políticas, Universidade de Lisboa, Portugal
[2] Instituto Superior de Ciências Sociais e Políticas, Universidade de Lisboa, Portugal

Correspondence: Sónia Pedro Sebastião (ssebastiao@iscsp.ulisboa.pt)

Submitted: 15 October 2024 **Accepted:** 17 December 2024 **Published:** 3 April 2025

Issue: This article is part of the issue "AI, Media, and People: The Changing Landscape of User Experiences and Behaviors" edited by Jeong-Nam Kim (University of Oklahoma) and Jaemin Jung (Korea Advanced Institute of Science and Technology), fully open access at https://doi.org/10.17645/mac.i475

Abstract

This study aims to dissect the normative discourse about artificial intelligence (AI) transparency using frame analysis. By employing a predominantly deductive, qualitative, and interpretative approach, the research leverages a qualitative frame analysis informed by a literature review on AI ethics and transparency. The study examines various AI ethical frameworks and regulations—China's Next Generation Artificial Intelligence Development Plan, the OECD's Recommendation of the Council on Artificial Intelligence, the White House's Blueprint for an AI Bill of Rights, and the EU's Artificial Intelligence Act—to understand how transparency is framed, transparency's objects, the defined accountability, and the responsible entities for ensuring transparency in the production of AI information. The study highlights transparency as a core ethical principle for trustworthy AI, emphasising its importance in final outputs and throughout AI development and deployment stages for fostering public trust. The findings indicate variability in language, priorities, and approaches to transparency across different frameworks, influenced by their socio-political, economic, and cultural contexts. Despite encouraging transparency as an ethical principle, the study notes a need for concrete guidance for its practical implementation across different AI applications. This gap underscores the need for critical examination and improvement in governance to enhance transparency and accountability in AI development and deployment. The innovative methodological approach, combining qualitative frame analysis with a theory-driven codebook, offers a novel template for investigating key concepts and issues in AI ethics and governance.

Keywords

accountability; artificial intelligence; ethical frameworks; regulation; transparency

1. Introduction

Pioneers of artificial intelligence (AI) sustained that a machine could simulate any aspect of learning or intelligence if precisely described (Lungarella et al., 2007, p. 2). AI then made a name for itself in specialised transport systems and industrial and commercial sectors.

The definition of AI is challenging due to the complexity of human intelligence. AI has different evolutionary stages and "can be classified into analytical, human-inspired and humanized AI depending on its cognitive, emotional, and social competencies" (Kaplan & Haenlein, 2020, p. 39).

Questions about AI ethics have become increasingly important (e.g., Kaplan & Haenlein, 2020; Stahl et al., 2021). Complex ethical issues surround current and near-future AI systems, especially regarding AI's social and personal impact on humans (Liao, 2020). The dangers associated with AI relate to the gap between public understanding of science and the pace of policymakers' challenges, potentially leading to apathy, lack of responsibility and accountability, moral panic, and inadequate legislation. It is, therefore, essential to analyse what is being done to explain and foster an understanding of these systems in normative terms.

Researching the ethics of generative AI is significant due to the need to align AI development with human values, the increasing inequalities, accountability and transparency issues, and AI's transformative potential (e.g., Cath et al., 2018; Gil de Zúñiga et al., 2023; Roberts et al., 2021). There is a need to ensure AI development aligns with human values and supports social good. With the evolution of generative AI systems, it will be critical to study their societal impacts and develop ethical frameworks to guide their development and deployment in beneficial ways.

The logic of transparency is linked to the attribution of responsibility, suggesting that understanding how a system works enables identifying responsible parties for malfunctions or malpractice (Ananny & Crawford, 2018). When an AI system causes harm, responsibility is assigned to different entities involved in the AI's lifecycle, including the company, the developer team, and the AI system itself (Sullivan & Wamba, 2022).

Transparency, accountability, and explainability are paramount in AI systems. Transparency is among the quintessential principles in the global ethical frameworks for AI (Jobin et al., 2019). The opacity and the extensive scale of generative models pose a significant challenge in elucidating their internal reasoning processes. Thus, facilitating a transparent rationale for the outputs generated by these models to the stakeholders affected emerges as a critical ethical concern warranting rigorous examination. Nonetheless, the discourse extends to interrogate the entities to whom AI should exhibit transparency and whether, in scenarios marked by technical constraints, the quest for transparency should compromise the performance of more transparent systems.

Using frame analysis, we aim to understand the normative discourse on AI transparency. This approach elucidates the implicit assumptions, priorities, and normative orientations encapsulated within ethical guidelines and regulations. It discloses the foundational frames, thereby revealing transparency in conceptual, normative, and practical dimensions. This analytical scrutiny may contribute towards a more refined and productive governance of AI technologies. A critical frame identification and examination are imperative for advancing transparency and accountability in developing and deploying AI technologies.

It allows for capturing multiple perspectives, uncovering underlying assumptions, identifying dominant and marginalised frames, enabling comparative analysis, and enhancing policy relevance.

This article is organised into three sections. The theoretical section discusses the concepts of AI and transparency. Based on the contributions of several authors, a theory-driven framing model is built. The third section presents the methodological approach, followed by the empirical study, discussion, and conclusions.

2. AI and Transparency

The evolution of AI systems has seen several setbacks and disappointments. As demonstrated by Gil de Zúñiga et al. (2023), there have been various definitions of the concept, some more machine learning-centric, others focusing on functions, cognitive simulation, and the creation of autonomous agents. Those definitions tend to be narrow in scope, broad, and vague, with human-centric bias and overemphasising autonomy. Despite providing valuable perspectives, those definitions' weaknesses undermine their comprehensiveness and applicability. Therefore, the authors propose a comprehensive definition of AI as "the tangible real-world capability of non-human machines or artificial entities to perform, task solve, communicate, interact, and act logically as it occurs with biological humans" (Gil de Zúñiga et al., 2023, p. 4).

Other AI definitions present a system's ability to correctly interpret external data, learn from it, and use the knowledge to achieve specific goals and tasks through flexible adaptation (Kaplan & Haenlein, 2019, p. 17), providing a competitive advantage to their holders. Both definitions stress the dimensions of performance and autonomy.

The field of AI study is cross-disciplinary and includes linguistics, cognitive sciences, neurosciences, robotics, engineering, computer science, social sciences, and humanities (Frankish & Ramsey, 2014). The advancement of technology has facilitated the shift from systems that imitate human intelligence and cognition to systems that generate content using generative AI. In the 21st century, its widespread adoption in personal technologies, multimedia content creation, and the evolution of generative pre-trained transformers and deep learning led to the AI systems' growing popularity and prevalence.

The opaque nature of deep learning raises concerns about interpretability, explainability, and trust. According to Liao (2020), deep learning is susceptible to adversarial attacks and errors, highlighting the importance of trust, interpretability, and explainability in fields like medicine and law, where human lives can be at stake.

Trustworthy AI requires transparency, including a broader socio-legal and computer-scientific perspective (Larsson & Heintz, 2020). Transparency is a concept that originated during the Enlightenment and involves the use of observation and knowledge to exercise social control (Hood, 2006). Therefore, it is a pervasive concept in political sciences studies, public and corporate governance, and communication studies. It is possible to identify different contexts in which transparency has been applied, such as in organisational and societal affairs, as a public value embraced by society to counter corruption, as a tool of good governance, and as a means of creating accountability, efficiency, and effectiveness (Larsson & Heintz, 2020).

As AI systems become embedded in more public systems' decision-making, Kemper and Kolkman (2019) advocate for more transparency in developing, implementing, and using algorithms in organisations.

According to Jobin et al. (2019), transparency refers to the clarity and openness concerning how AI systems operate, make decisions, and affect users and stakeholders. Transparency relates to how users and stakeholders can explain and understand AI systems and their decisions. Therefore, transparency is a conceptual metaphor associated with knowing and understanding (Larsson & Heintz, 2020).

AI systems' lack of transparency and accountability is a significant concern. Ananny and Crawford (2018) define transparency as seeing inside a system and understanding its mechanisms and decision-making processes. They note that transparency can be at the level of platform design and algorithmic mechanisms or, more deeply, at the level of a software system's logic.

Transparency requires disclosing information or revealing the interests of the issuer and holder of information (disclaimer). It also involves recognising and valuing transparency as an essential aspect of social control. The beholder must acknowledge and value transparency (Kemper & Kolkman, 2019).

Several typologies of transparency have been identified by considering categories such as types of information, objects of transparency, and accountability (see Table 1).

The ideal of transparency may not be the most suitable for AI ethics (e.g., Ananny & Crawford, 2018; Liao, 2020). While complete AI transparency may be ideal, it is only sometimes practical due to the complexity and potential trade-offs with other principles (Jobin et al., 2019; Liao, 2020). Ananny and Crawford (2018) argue that transparency is an ongoing process of scrutiny and adjustment, requiring more than just revealing the inner workings of AI systems. Ferrari et al. (2023) propose three structural conditions for effective AI governance: industrial observability, public inspectability, and technical modifiability. These conditions represent different levels at which AI systems must be transparent and accountable for effective oversight and regulation.

Table 1. Typologies of transparency.

Category	Type of Transparency	Description
Type of information	Fuzzy	The information provided does not reveal how institutions behave; the information is disclosed nominally or is unreliable.
	Clear	Reliable information is provided, for example, about institutional performance, responsibilities, and funds use.
Objects of transparency	Event	Event transparency focuses on disclosing specific data points, results, or impacts of a system's operations.
	Process	Process transparency aims to make visible the underlying logic, steps, and governance frameworks that determine how a system functions.
Accountability	"Soft" accountability	Organisations must answer for their actions when transparency is present.
	"Hard" accountability	Transparency brings the power to sanction organisations and demand compensation for the harm they cause.

Source: Adapted from Ananny and Crawford (2018).

However, Ananny and Crawford (2018) highlight the limitations of relying solely on transparency for accountability and understanding in AI, such as disconnection from power, professional boundary work, epistemological challenges, and the distributed nature of the actors involved. Even with information on how the system works, users may need help understanding algorithms or data usage, leading to a lack of trust and difficulty in holding the system accountable (Buiten, 2019).

The main challenges in regulating AI transparency are the complexity of the concept and the difficulty of providing technically feasible explanations helpful in specific legal contexts (Buiten, 2019). Despite these obstacles, various regulations prioritise transparency as one of their principles.

After conducting extensive bibliographic searches in the Scopus, EBSCO, and Web of Science databases using the keywords "fram*," "transparency," and "artificial intelligence" in titles, author keywords, and abstracts in English, French, and Portuguese, we were unable to find any prior studies or framing models related to AI transparency. Given this gap in the literature and the theoretical background presented previously, we propose a new framework for AI transparency. This framework includes key aspects such as understandability and explainability, accountability and governance, disclosure and communication, documentation and access to information, and ethical and legal compliance (see Table 2).

Table 2. Frames of transparency.

Frame	Description	Authors
Understandability and explainability	There is a need for AI systems to provide clear and understandable explanations of their decisions and processes. Ensuring that AI systems' functioning and decision-making processes are clear and understandable to users and stakeholders.	Ananny and Crawford (2018), Buiten (2019), Larsson and Heintz (2020)
Accountability and governance	Identifying and holding responsible parties accountable for the deployment and impacts of AI systems, ensuring transparent governance.	Larsson (2020), Sullivan and Wamba (2022)
User awareness and communication	Safeguarding users are aware when they interact with AI systems and understand the role of AI in decision-making processes. This includes effective communication about AI capabilities and limitations.	Ferrari et al. (2023), Kaplan and Haenlein (2019)
Documentation and access to information	Providing detailed documentation about AI systems' design, development, and functioning is essential for transparency. This includes making relevant information accessible to various stakeholders.	Corrêa et al. (2023), Ferrari et al. (2023), Kemper and Kolkman (2019), Larsson (2020)
Ethical and legal compliance	Transparency is framed as a means to ensure that AI systems comply with ethical standards and legal requirements. This includes adhering to principles of fairness, non-discrimination, data privacy, and human rights.	Corrêa et al. (2023), Kaplan and Haenlein (2020)

Framing is a process employed by message creators to organise and interpret information, making it one of the most widely applied theories in communication studies (Lock et al., 2020). It involves selecting a particular point of view to highlight specific message characteristics. Entman (1993) states that framing information significantly influences how people understand and react to issues. In the context of AI transparency, frame analysis helps elucidate the underlying assumptions, priorities, and normative directions embedded within these guidelines.

By utilising frame analysis, we aim to understand the normative discourse surrounding AI transparency. Transparency, accountability, and explainability are paramount in AI systems and need to be part of global ethical frameworks for AI (Jobin et al., 2019). But, how is transparency framed in ethical frameworks? (RQ1)

Therefore, the typologies presented in Table 1 raise research questions such as what are the objects of transparency (RQ2) and what kind of accountability is defined (RQ3)?

Finally, in light of the role of AI systems in algorithmic accountability and AI governance (e.g., Ferrari et al., 2023), another research question arises: Who is identified as responsible for ensuring transparency in the use of AI in producing and disseminating information? (RQ4)

The next section outlines the methodological approach, keeping in mind the research questions at hand.

3. Methodology

A predominantly deductive method of a qualitative and interpretative nature is used. The deductive approach is justified by theory-driven research (Bryman, 2016). A qualitative frame analysis based on Entman's (1993) value of framing is performed, using frames inferred from the literature review about AI ethics and transparency. The frame analysis of AI ethical guidelines provides a crucial lens for understanding how information about AI transparency is organised and can be interpreted.

This study focused on documents issued by transnational organisations (the OECD and the EU) and state organisations (China and the US), selected for their scope, recency, and relevance across different regulatory contexts (e.g., Corrêa et al., 2023). These frameworks represent the world's largest economic powers and most influential policy-setting organisations (Lee, 2018). The US, China, and the EU, along with the OECD, effectively shape AI governance for a significant portion of global AI development and deployment (Larsson, 2020).

The OECD Recommendation of the Council on Artificial Intelligence (RCAI) was launched in 2019 and is the first intergovernmental standard on AI (OECD, 2019). The EU Artificial Intelligence Act (AI Act; European Parliament, 2024) is the first transnational AI regulation. It was approved on 21st May 2024. Once adopted, it will be a binding legal act that must be applied across all EU member states. The AI Act aims to promote trustworthy and human-centred AI and establish a relationship with existing laws such as the General Data Protection Regulation and product safety, consumer protection, and labour law. As a regulation, it goes beyond a policy document and has the force of law.

Apart from these two binding documents, the analysis includes the Next Generation Artificial Intelligence Development Plan (AIDP) from China's State Council (2017) and the Blueprint for an AI Bill of Rights (BAIBR) from the White House Office of Science and Technology Policy (2022).

The AIDP outlines China's strategic goals, including becoming the world leader in AI by 2030, creating a significant AI industry, and using AI to drive economic development, social governance, and defence capabilities. It serves as a guiding document for China's national AI development, considering economic, political, cultural, and ethical factors.

The BAIBR contributes to the US regulatory and ethical AI development and deployment landscape. However, it only provides a set of non-binding principles and practices that aim to guide the design, development, and deployment of AI systems in a way that respects human rights and promotes public trust. It emphasises safe and effective design, protection against algorithmic discrimination, robust data privacy, transparency through notice and explanation, human alternatives, and availability of oversight. Table 3 organises the documents analysed, classifying them by year, issuer, nature of the document, and language.

Using the theory-driven frames described in Table 2, we aim to understand the normative discourse on AI transparency. Diverse actors have produced this discourse, including transnational institutions, states, research institutions, companies, NGOs, and professional associations (Corrêa et al., 2023; Jobin et al., 2019).

MAXQDA was selected for this study due to its advanced capabilities in qualitative data analysis, including automated lexical searches, hierarchical coding systems, and compatibility with various file formats. Its user-friendly interface and widespread academic adoption (Lewins & Silver, 2007; Woolf & Silver, 2018) further ensured its suitability for managing the extensive corpus of policy documents analysed in this study.

Given the normative nature and length of the corpus, we employed MAXQDA to automate the text search for the identified keywords. Our search included a lexical search for the keywords "transparency," "accountability," and "responsibility," incorporating lemmatisation to cover variations of these terms (e.g., transparent, accountable, responsible). The choice of search words is based on our literature review and the

Table 3. Description of the corpus.

Document Title	Issuer	Year	Number of Pages (Without Appendix/Annex)	Nature of the Document	Language
Next Generation AIDP	China's State Council	2017	28	Plan	English full translation provided by Stanford University
RCAI	OECD	2019	11	Policy	English
BAIBR	White House Office of Science and Technology Policy	2022	52	Guidelines	English
AI Act	European Parliament	2024	376	Regulation	English

realisation that the logic of transparency is associated with accountability, i.e., the attribution of responsibility (Ananny & Crawford, 2018; Jobin et al., 2019; Sullivan & Wamba, 2022). Additionally, we included the preceding and following sentences in the highlighted text segments.

These options allowed us to identify meaningful text segments, which we categorised according to the frames outlined in the codebook. The codebook was developed based on the frames presented in Table 2, along with the transparency objects and types of accountability outlined in Table 1. While the "fuzzy vs. clear" dimension was initially considered as part of the analytical framework, it was excluded after preliminary analysis for methodological and practical reasons. The dimension's subjective nature and overlap with other categories, such as understandability and explainability, posed challenges in ensuring consistent coding. Additionally, the lack of clear differentiation in the reviewed documents further justified its exclusion. This decision was made to maintain analytical rigour and focus on dimensions more directly aligned with the study's objectives. For example, the RCAI, AIDP, and BAIBR do not clearly assign responsibility for transparency in the production and dissemination of AI-enabled information to specific actors. In contrast, the AI Act is a legal document that provides clear guidelines on transparency, accountability, and the responsibilities of AI agents. However, it uses legal terminology and jargon that may not be easily understood by all AI users. To ensure the objectivity of our analysis and reach a consensus in coding, we decided to exclude this dimension.

It is important to note that all documents were reviewed beforehand to ensure that the coders were familiar with the texts. The lexical search enabled us to mark and categorise significant texts for each frame. The coding process helped us identify the key frames listed in Table 4, along with examples that we incorporated into the text. All authors agreed upon the selection of examples. The rigorous validation of coding decisions not only enhanced the reliability of the analysis but also ensured that the identified frames accurately reflected the normative and practical dimensions of AI transparency as presented in the analysed documents.

4. Results

Except for the AIDP, the analysed documents address issues of fairness, non-discrimination, data privacy, and human rights, but not necessarily in the context of transparency. The results associated with the research questions are presented below.

Considering RQ1—how is transparency framed in the AI ethical frameworks and regulations?—we note that the AIDP emphasises the importance of transparency in AI development but does not specify explainability requirements. The plan does not explicitly address documentation and access to information nor user awareness and communication. It serves as a directive setting the overall direction and priorities for AI development in China, with the expectation that various state and non-state actors will work towards these goals under the central government's guidance. The government is, therefore, responsible for establishing "a traceability and accountability system, and clarify the main body of AI and related rights, obligations, and responsibilities" (China's State Council, 2017, p. 25).

The RCAI defines AI as "a general-purpose technology that has the potential to: improve the welfare and well-being of people, contribute to positive sustainable global economic activity, increase innovation and productivity, and help respond to key global challenges" (OECD, 2019, p. 3). The document emphasises

transparency as a core value for responsible AI development. It includes: (a) user awareness and communication, which encourages communication with stakeholders about AI capabilities and limitations; (b) accountability and governance, where organisations and individuals responsible for AI systems should be identifiable and accountable, promoting transparency in governance and oversight (it does not specify governance structures); and (c) understandability and explainability, by ensuring that AI systems are transparent and understandable to users, stakeholders, and regulators, as illustrated in the excerpt:

> To enable those affected by an AI system *to understand the outcome*, and, (iv.) to enable those *adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information* on the factors, and the logic that served as the basis for the prediction, recommendation or decision. (OECD, 2019, p. 8, emphasis by authors)

The BAIBR underlines the need for transparency in AI systems to protect individual rights and promote trust. It focuses on automated systems, revealing them to users and explaining how they work. Although the document mentions other frames, the understandability and explainability frame is more emphasised in the context of transparency and accountability.

Some examples of segmented text regarding understandability and explainability, to ensure individuals are aware when an AI system is being used and provide explanations about how decisions are made, as illustrated in the excerpt: "An automated system should provide demonstrably clear, timely, understandable, and accessible notice of use, and explanations as to how and why a decision was made or an action was taken by the system" (White House Office of Science and Technology Policy, 2022, p. 43). Regarding documentation and access to information, by encouraging organisations to publicly disclose information about the use and impact of AI systems:

> Provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organisation responsible for the system, and explanations of outcomes that are clear, timely, and accessible. (White House Office of Science and Technology Policy, 2022, p. 6)

Concerning user awareness and communication, which emphasises the need for public update reporting, as illustrated in the excerpts:

> Audits and impact assessments to help identify potential algorithmic discrimination and provide transparency to the public in the mitigation of such biases. (White House Office of Science and Technology Policy, 2022, p. 24)

> The American public should be protected via built-in privacy protections, data minimization, use and collection limitations, and transparency. (White House Office of Science and Technology Policy, 2022, p. 33)

And lastly, concerning accountability and governance: "Entities responsible for the development or use of automated systems should lay out clear governance structures and procedures" (White House Office of Science and Technology Policy, 2022, p. 19).

The AI Act strongly emphasises transparency, particularly for high-risk AI systems. All frames are used to approach transparency, as illustrated in the following excerpts. Regarding understandability and explainability, where High-risk AI systems must be transparent and provide clear information to deployers: "High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to deployers" (European Parliament, 2024, p. 194). Concerning accountability and governance, it establishes governance through national supervisory authorities and conformity assessments in several articles, sections, and annexes. As for user awareness and communication, this regulation mandates user information and transparency measures for high-risk AI systems, ensuring users know the AI's role in decision-making processes:

> Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system. (European Parliament, 2024, p. 256)

> Deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. (European Parliament, 2024, p. 258)

Regarding documentation and access to Information, it requires technical documentation and record-keeping for high-risk AI systems:

> Providers shall have a choice of integrating, as appropriate, the necessary testing and reporting processes, information and documentation they provide with regard to their product into documentation and procedures that already exist and are required under the Union harmonisation legislation listed in Section A of Annex I. (European Parliament, 2024, p. 185)

Lastly, for ethical and legal compliance, it requires providers of AI systems to ensure that their systems are transparent to users, including providing information on the purpose and intended use of the AI system and the logic, significance, and potential impact of the AI system's decisions:

> Providers of high-risk AI systems shall put a quality management system in place that ensures compliance with this Regulation. That system shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions. (European Parliament, 2024, p. 202)

Based on the text excerpts provided, Table 4 summarises key transparency principles, accountability measures, and implementation challenges of the analysed documents.

Table 4. Distribution of transparency frames across AI frameworks.

Document	Key Transparency Principles	Accountability Measures	Implementation Challenges
AIDP	Traceability and accountability	Regulatory oversight	Enforcement, cultural differences
RCAI	Understandability and explainability to foster clarity	Audits, reporting	Complexity, technical limits
BAIBR	Understandability and explainability, accountability	Legal obligations	Inter-agency coordination
AI Act	Understandability and explainability, user awareness and communication, documentation and access to information as basis for risk assessments and safety	Compliance checks	Harmonisation across EU

Table 5 presents information related to RQ2 and RQ3. It identifies the analysed documents and provides excerpts illustrating the codes of objects of AI transparency and types of accountability (Ananny & Crawford, 2018). Transparency is categorised into two main types: event transparency and process transparency. Event transparency focuses on disclosing specific data points, results, or impacts of a system's operations. Process transparency aims to clarify the underlying logic, steps, and governance frameworks that determine how a system functions. On the one hand, "soft accountability" refers to voluntary or normative mechanisms based on recommendations, codes of conduct, or non-binding guidelines. These mechanisms encourage the adoption of responsible practices but rely on the voluntary adherence of those involved. Examples include internal audits, public reports, and organisational ethical commitments (Ananny & Crawford, 2018). On the other hand, "hard accountability" involves formal and binding mechanisms, such as legal sanctions, financial compensation, or regulatory obligations. These mechanisms require compliance and can impose penalties on organisations or individuals who violate established norms (Ferrari et al., 2023).

Table 5. Objects of AI transparency and kind of accountability.

Document	RQ2 (Event/Process)	RQ3 (Hard/Soft Accountability)
AIDP	The document does not explicitly mention transparency as a focus area.	The plan does not clearly define specific responsibilities.
RCAI	Event and Process "AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: i. to *foster a general understanding of AI systems*, ... iii. *to enable those affected by an AI system to understand the outcome*, and, iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the *logic that served as the basis for the prediction, recommendation or decision*" (OECD, 2019, p. 8, emphasis by authors).	Soft "AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour right" (OECD, 2019, p. 7).

Table 5. (Cont.) Objects of AI transparency and kind of accountability.

Document	RQ2 (Event/Process)	RQ3 (Hard/Soft Accountability)
BAIBR	Event "Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the *overall system functioning* and the role automation plays, notice that such systems are in use, the individual or organisation responsible for the system, and explanations of *outcomes* that are clear, timely, and accessible" (White House Office of Science and Technology Policy, 2022, p. 6, emphasis by authors).	Soft "Responsibility should rest high enough in the organisation that decisions about resources, mitigation, incident response, and potential rollback can be made promptly, with sufficient weight given to risk mitigation objectives against competing concerns. Those holding this responsibility should be made aware of any use cases with the potential for meaningful impact on people's rights, opportunities, or access as determined based on risk identification procedures" (White House Office of Science and Technology Policy, 2022, p. 19).
AI Act	Event Article 13—Transparency and provision of information to deployers (European Parliament, 2024, pp. 194–196).	Soft "Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards" (European Parliament, 2024, p. 257). Hard Article 50—Transparency obligations for providers and deployers of certain AI systems (European Parliament, 2024, pp. 256–259).

Finally, the identification of those responsible for ensuring transparency in the use of AI in producing and disseminating information (RQ4) is not addressed in a consistent manner in all documents (Figure 1).

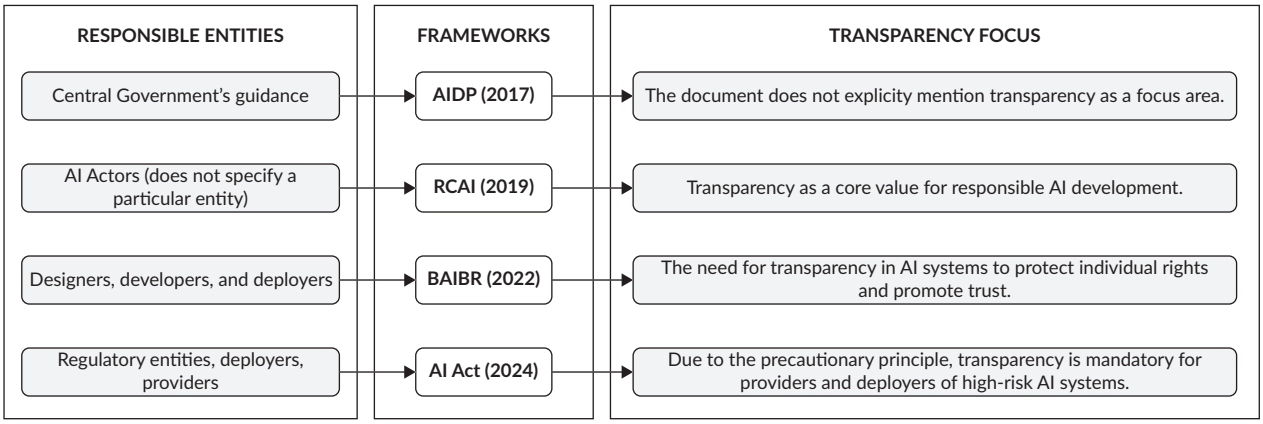

Figure 1. Responsible entities and transparency focus by AI framework.

The RCAI emphasises transparency and responsible disclosure around AI systems. However, it does not specify a particular entity responsible for ensuring this transparency. The same happens with the AIDP, but as a state plan, the implied responsibility likely falls on government entities overseeing AI development and deployment.

The BAIBR asserts the right to transparency in AI systems and calls for clear, understandable explanations. It suggests this is a shared responsibility of those designing, developing, and deploying AI systems. As a US government document, federal agencies are impliedly responsible for upholding these principles.

The AI Act is precautionary and places transparency obligations on providers and deployers of high-risk AI systems. Providers must ensure their systems are transparent and give clear information to users. Deployers have obligations related to monitoring, reporting, and facilitating oversight. Member states, through their national supervisory authorities, are responsible for enforcing these transparency requirements.

5. Discussion

Governments worldwide have begun to develop regulations to govern the use of AI. These regulations aim to ensure that AI is used ethically and responsibly and does not infringe on people's rights. Ethical challenges and principles are at the forefront of ongoing discussions about the governance and regulation of AI, advocating for a multidisciplinary, systemic, proactive, and anticipatory approach to policymaking (e.g., Corrêa et al., 2023; Jobin et al., 2019; Larsson & Heintz, 2020). There is no single approach to AI transparency that suits all contexts. Various documents highlight different aspects, including understandability, accountability, ethical compliance, and risk management.

Notwithstanding AI's potential, if it is used without a moral agenda, it can have harmful consequences (e.g., Bostrom & Yudkowsky, 2014). It seems, however, that ethical frameworks and a self-regulating moral agenda are not enough to contain the social and political impacts of AI (Suleyman & Bhaskar, 2023). For example, aside from efforts to regulate the use of AI and define principles for the governance of AI, China's approach to these issues may differ from Western perspectives due to its unique socio-political context and governance. Roberts et al. (2021) suggest that the Chinese government's interests might outweigh ethical considerations, particularly in surveillance and social governance. This tension could shape the global development and implementation of ethical norms due to state-centric governance models that hinder more decentralised, multi-stakeholder approaches.

The analysis of ethical frameworks and regulations involving AI highlights a greater focus, on one hand, on the events, i.e., on disclosing specific data points, results, or impacts of an AI system's operations (inputs, outputs, and outcomes) rather than in the system's functioning logic (RQ2). On the other hand, it highlights a greater focus on soft accountability with recommendations and prescriptions for agents/players (RQ3).

Hard accountability mechanisms lack in major AI policy frameworks from the OECD, China, and the US. These initiatives seem to fall more on the "soft accountability" end of the spectrum (Ananny & Crawford, 2018). These frameworks aim to bring transparency to AI development in the hope that it will pressure organisations to behave responsibly and be able to justify their actions. However, they do not include "hard accountability" measures that allow for formal sanctions or compensation when violations occur. They depend more on self-regulation and public pressure to encourage adherence.

Though transparency and explainability have become a dominant topic of concern for AI systems since 2018 (Corrêa et al., 2023), we agree with Ferrari et al. (2023) regarding the lack of clarity in AI transparency obligations. There are currently no specific technical details about how such modifications can be enforced in policy practice. The three structural conditions for effective AI governance (industrial observability, public inspectability, and technical modifiability) are also missing, compromising effective oversight and regulation.

This is an important limitation in the accountability paradigms. Ananny and Crawford's (2018) argument implies that achieving meaningful accountability likely requires going beyond transparency alone to include "harder" mechanisms with teeth.

There are also limitations stemming from the lack of clear responsibility for producing and distributing information about the AI system and its events (RQ4). While the understandability and explainability and documentation and access to information frames (RQ1) are acknowledged, they do not specify the conditions for producing and accessing information. Even though the RCAI, the AIDP, and the BAIBR underscore the significance of AI transparency, none of them clearly assign responsibility for transparency in AI-enabled information production and dissemination to specific actors (RQ4). Only the AI Act places transparency obligations on providers and deployers of high-risk AI systems, revealing a stronger emphasis on legal compliance compared to BAIBR, reflecting differences in regulatory approaches.

An integrated approach that combines elements from multiple frames may provide a more comprehensive solution to AI transparency. Efforts to have an integrated approach should be present in policymaking. Policymakers should take into account multiple frames to address the multifaceted nature of AI transparency, such as ensuring AI systems are understandable, holding developers accountable, adhering to ethical standards, and managing risks effectively (Bostrom & Yudkowsky, 2014; Jobin et al., 2019), since an effective AI governance requires a special balance between regulation and flexibility to technological advancements.

An international framework could be a potential solution to address the current limitations of existing frameworks, which primarily focus on "soft accountability" without enforcement mechanisms. There is a lack of clear technical specifications for implementation, an absence of structural conditions for effective oversight, and an inconsistent assignment of responsibility for transparency obligations.

The proposed framework needs to consider the different approaches among regions, as minimising governance models can impede multi-stakeholder participation. Its governing body could be composed of multi-stakeholders representing national governments, the tech industry, academic institutions, civil society organisations, and international standards bodies. Key components of this framework must include mandatory technical standards for AI transparency, clear accountability mechanisms with enforcement powers, dispute resolution procedures, and regular review and update processes. Ethical considerations and a human-centric approach should take precedence over commercial interests.

However, implementing such a framework presents significant challenges. Geopolitical tensions and competing national interests may impede international cooperation, while enforcement across jurisdictions requires complex diplomatic and legal mechanisms. Moreover, the framework must balance the protection of intellectual property rights with transparency requirements, particularly as AI technologies rapidly evolve and market dynamics shift.

AI functions as a sociotechnical system; its context of data creation and interpretation is shaped by humans, and the culture surrounding AI technologies is fundamentally human (Airoldi, 2022). This sociotechnical perspective underscores why governance frameworks must extend beyond technical specifications to encompass social, cultural, and ethical dimensions. Therefore, concerns about AI transparency are intrinsically linked to human values and social ethics, necessitating a governance approach that recognises both the technical and social complexities of AI systems.

6. Conclusion

Choosing a qualitative frame analysis, we were able to present a comprehensive and systematic approach to examine the complexity of AI transparency's multifaceted nature regarding its ethics, policymaking, and governance. The various documents discussing ethical principles for AI have different scopes and priorities. For example, China's AIDP centres on economic competitiveness, while the other documents focus on fundamental rights; the AI Act provides detailed regulations, while the US Blueprint focuses more on high-level principles. This lack of alignment could limit transparency and imply vested economic and political interests.

Some frameworks, such as the RCAI and BAIBR, are non-binding, potentially limiting their impact on driving transparent practices compared to the enforceable regulations in the AI Act. Despite these differences, the fundamental frames highlighted in the analysed AI ethical guidelines and regulations reveal some of the conceptual, normative, and practical dimensions of transparency.

Conceptually, transparency is a core ethical principle for trustworthy AI. It enables explainability and understanding of how AI systems make decisions. The documents also highlight the need for transparency at various stages of AI development and deployment, not just in the final outputs, stressing the need for an overall understanding of the AI system's functioning and logic beyond the prediction, recommendation, or decision.

Normatively, transparency is a fundamental right for individuals impacted by AI systems. Ensuring AI systems and their developers can be held accountable is critical to fostering public trust. Except for the AIDP, the documents also link transparency to normative principles of fairness and equity and prevent discriminatory impacts of AI.

However, in practical terms, the documents lack specificity on transparency requirements. While transparency is encouraged as an ethical principle, there is limited concrete guidance on what transparency entails in practice for different AI applications. Even with ethical principles established, ensuring meaningful transparency will require robust implementation, oversight, and enforcement mechanisms, which may face practical hurdles.

In summary, differences in priorities, legal obligations, specificity, and implementation across the world can hinder the consistent achievement of AI transparency goals without further alignment and strengthening of approaches. Developing a coherent, flexible, dynamic, and context-aware ethical international framework may help keep AI technology on a responsible path. Continuous learning, collaboration, and adaptation will be crucial.

This study's findings point to several critical areas that warrant further investigation in the field of AI transparency and governance. Research should examine how organisations operationalise transparency requirements across different AI applications and contexts, focusing on successful implementation strategies and practical challenges.

Scholarly attention should focus on establishing clear lines of responsibility and accountability in AI development and deployment, particularly in complex multi-stakeholder environments where responsibilities span multiple actors and jurisdictions. The field would also benefit from research exploring how different cultural, social, and political contexts influence transparency expectations, moving beyond Western-centric approaches towards more culturally sensitive and globally applicable governance frameworks.

Further empirical research is needed to assess the effectiveness and impact of AI ethical guidelines and regulations. Such studies should evaluate how different regulatory approaches influence organisational behaviour, innovation processes, and compliance mechanisms. This includes examining the implementation challenges of binding versus non-binding frameworks and their relative success in promoting transparent and responsible AI practices.

Longitudinal studies evaluating the effectiveness of transparency mechanisms in promoting responsible AI development and maintaining public trust are essential to provide empirical evidence of successful approaches across different contexts. Additionally, research should examine how individual and organisational choices influence transparency outcomes, acknowledging that human decision-making remains central to AI development and investigating how organisational culture and institutional frameworks shape transparency practices.

This article's scope is limited due to the abundance of AI ethical guidelines issued by research institutes, companies, and NGOs. However, the study includes input from various regions to reduce Western bias. The findings underscore that responsible AI fundamentally depends on responsible human actors, as humans create the technology, program the applications, select the information, and determine its use.

The path forward requires a delicate balance between establishing robust transparency frameworks and maintaining flexibility for technological advancement. Future success in AI governance will depend on continuous learning, international collaboration, and adaptive approaches that recognise both the technical and human dimensions of AI systems.

Acknowledgments

This article used AI to translate original Portuguese sentences into English using DeepL. AI was also employed in specific aspects, namely language clarity and checking using Grammarly Business. The authors double-checked all AI-assisted texts to correct selected wordings, grammatical adjustments, and paraphrases. The authors, safeguarding human oversight and expertise in the final output, performed the overall composition of the article manually. The final manuscript was proofread by a proficient/native English speaker.

Funding

This work is supported by Portuguese national funds through FCT—Fundação para a Ciência e a Tecnologia, under project UIDB/00713/2020.

Conflict of Interests

The authors declare no conflict of interests.

References

Airoldi, M. (2022). *Machine habitus: Toward a sociology of algorithms*. Polity.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 316–334). Cambridge University Press.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Buiten, M. (2019). Towards intelligent regulation of artificial intelligence. *European Journal of Risk Regulation*, *10*(1), 41–59. https://doi.org/10.1017/err.2019.8

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, *24*(2), 505–528. http://doi.org/10.1007/s11948-017-9901-7

China's State Council. (2017). *A Next Generation Artificial Intelligence Development Plan*. Stanford Cyber Policy Center. https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Barbosa, C., Massmann, D., Manbrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023), *Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance*. arXiv. https://doi.org/10.48550/arXiv.2206.11922

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, *43*(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

European Parliament. (2024). *Artificial Intelligence Act*. https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf

Ferrari, F., van Dijck, J., & van den Bosch, A. (2023). Observe, inspect, modify: Three conditions for generative AI governance. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448231214811

Frankish, K., & Ramsey, W. M. (2014). Introduction. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 1–11). Cambridge University Press.

Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2023). A scholarly definition of artificial intelligence (AI): Advancing AI as a conceptual framework in communication research. *Political Communication*, *41*(2), 317–334. https://doi.org/10.1080/10584609.2023.2290497

Hood, C. (2006). Transparency in historical perspective. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (pp. 2–23). Oxford University Press. https://doi.org/10.5871/bacad/9780197263839.003.0001

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kaplan, A., & Haenlein, M. (2019). Siri, Siri in my hand, who is the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, *63*(1), 37–50. https://doi.org/10.1016/j.bushor.2019.09.003

Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081–2096. https://doi.org/10.1080/1369118X.2018.1477967

Larsson, S. (2020). On the governance of artificial intelligence through ethics guidelines. *Asian Journal of Law and Society*, *7*(3), 437–451. https://doi.org/10.1017/als.2020.19

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, *9*(2), https://doi.org/10.14763/2020.2.1469

Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. HarperCollins Publishers.

Lewins, A., & Silver, C. (2007). *Using software in qualitative research: A step-by-step guide*. Sage.

Liao, S. M. (2020). A short introduction to the ethics of artificial intelligence. In S. M. Liao (Ed.), *The ethics of artificial intelligence* (pp. 1–42). Oxford University Press.

Lock, I., Wonneberger, A., Verhoeven, P., & Hellsten, I. (2020). Back to the roots? The applications of communication science theories in strategic communication research. *International Journal of Strategic Communication*, *14*(1), 1–24. https://doi.org/10.1080/1553118X.2019.1666398

Lungarella, M., Iida, F., Bongard, J. C., & Pfeifer, R. (2007). AI in the 21st century—With historical reflections. In M. Lungarella, F. Iida, J. C. Bongard, & R. Pfeifer (Eds.), *50 years of artificial intelligence. Essays dedicated to the 50th anniversary of artificial intelligence* (pp. 1–8). Springer.

OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. *AI & Society*, *36*, 59–77. https://doi.org/10.1007/s00146-020-00992-2

Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., Laulhé Shaelou, S., Patel, A., Ryan, M., & Wright, D. (2021). Governing the ethics of artificial intelligence for human flourishing—Beyond principles for machine learning. *Journal of Business Research*, *124*, 374–388. https://doi.org/10.1016/j.jbusres.2020.11.030

Suleyman, M., & Bhaskar, M. (2023). *The coming wave: Technology, power, and the twenty-first century's greatest dilemma*. Crown.

Sullivan, Y. W., & Wamba, S. F. (2022). Moral judgments in the age of artificial intelligence. *Journal of Business Ethics*, *178*, 917–943. https://doi.org/10.1007/s10551-022-05053-w

White House Office of Science and Technology Policy. (2022). *The Blueprint for an AI Bill of Rights: Making automated systems work for the American people*. https://www.whitehouse.gov/ostp/ai-bill-of-rights

Woolf, N. H., & Silver, C. (2018). *Qualitative analysis using MAXQDA. The five-level QDA method*. Routledge.

About the Authors

Sónia Pedro Sebastião, PhD in social sciences in the speciality of communication sciences from the Instituto Superior de Ciências Sociais e Políticas, Universidade de Lisboa (Portugal), is a full professor in communication sciences and the director of the Center of Public Administration and Public Policies (CAPP, ISCSP/FCT). Her research interests focus on strategic communication, ethics, citizenship, and cultural studies.

David Ferreira-Mendes Dias, PhD in social sciences in the speciality of communication sciences from the Instituto Superior de Ciências Sociais e Políticas, Universidade de Lisboa (Portugal), is an invited auxiliary professor in communication sciences and a TV professional at RTP (Portuguese public service broadcaster). His research interests focus on digitalization, social media, platformization, and cultural studies.