Conceptualizing Algorithmic Transparency as a Legal Concept in Comparative Perspective

Lucas Costa dos Anjos

June 2026

TECHGLOBAL INSTITUTE

Aknowledgements

Lucas Costa dos Anjos would like to acknowledge the support received from the Tech Global Institute and of the European University of Social Sciences (CIVICA) in funding and facilitating the research and the development of this paper. The author would also like to disclose that the views expressed are solely those of the author and do not necessarily reflect the opinions or positions of the Tech Global Institute, CIVICA, or any other entities related to the author. The author made use of TurboScribe, DeepL and NotebookLM to assist with the drafting of this paper, especially for purposes of transcribing recordings, of translating transcripts and official documents from Portuguese into English, and of better suiting the vocabulary needs of the paper, from February to June 2025.

TECHGLOBAL INSTITUTE

Table of Contents

Summary

This research examines algorithmic transparency as a regulatory principle increasingly deployed to address accountability and oversight challenges in artificial intelligence systems. Through comparative legal analysis of major international and regional frameworks (including the EU's GDPR, AI Act, DSA, and DMA, the Council of Europe Framework Convention, OECD guidelines, UNESCO recommendations, and IEEE technical standards) combined with qualitative interviews with twelve multistakeholder experts from Brazil and Europe, the study reveals fundamental tensions between transparency's democratic promise and its practical implementation. The analysis demonstrates that while European frameworks establish sophisticated procedural transparency requirements, global application remains uneven. Technology companies implement differential transparency standards based on regulatory stringency and market significance rather than uniform principles, with Global South jurisdictions receiving substantially less disclosure despite operating under nominally similar systems. The research distinguishes between procedural transparency (disclosure of governance frameworks and processes) and material transparency (access to technical specifications and source code), finding that both face significant barriers including technical complexity, intellectual property concerns, and regulatory capacity asymmetries. Expert interviews establish consensus that transparency functions as an instrumental mechanism enabling accountability rather than a terminal objective, yet effectiveness depends critically on meaningful information presentation calibrated to recipient capacities. Current frameworks often prioritize formal compliance over substantive oversight, creating "performative transparency" that legitimizes existing power structures without challenging concentrated technological decision-making authority. The study concludes that algorithmic transparency, while necessary for democratic governance, operates within structural constraints limiting its effectiveness in addressing global power imbalances. Meaningful reform requires integrating transparency within broader frameworks of technological justice that address underlying capacity asymmetries between regulatory authorities and technology corporations, particularly affecting Global South jurisdictions.

1. Introduction

The proliferation of artificial intelligence (AI) systems across public and private spheres has fundamentally altered the landscape of legal accountability and democratic oversight. As automated decision-making increasingly mediates critical aspects of social life, from content moderation and financial services to criminal justice and healthcare, the traditional mechanisms of legal transparency face unprecedented challenges. There is an **inherent opacity** in complex computational systems, which threatens established principles of due process, while the transnational nature of AI technologies complicates jurisdictional approaches to regulatory enforcement.

Algorithmic transparency has emerged as a central organizing principle in contemporary technology regulation in the European Union (EU), featuring prominently in major legislative frameworks from the *General Data Protection Regulation* to the *Artificial Intelligence Act*.

Nonetheless, despite its widespread adoption across diverse legal and regulatory contexts, the concept remains contested in both its definition, interpretation, and implementation.

This conceptual ambiguity reveals significant practical challenges, including technical limitations of complex machine learning systems, competing commercial interests, and variations in regulatory capacity across different jurisdictions.

Whether this ambiguity reflects intentional policy design or unintentional regulatory shortcoming merits examination. Intentionalist interpretations suggest definitional flexibility constitutes deliberate adaptation to technology's evolving nature, avoiding obsolescence while enabling contextual application across diverse algorithmic systems that resist uniform specification. Transparency provisions emerge from protracted multi-stakeholder negotiations balancing industry resistance against civil society demands, with vagueness representing necessary compromise among actors with divergent interests and asymmetric power.

Ambiguity may represent the best attainable outcome given competing constraints: legislative windows coupling problem recognition, political will, and available solutions yield imperfect compromise as the maximally achievable consensus rather than optimal precision.

Conversely, unintentionalist critiques identify ambiguity as regulatory failure reflecting policymaking shortcomings, such as insufficient technical expertise, compressed timelines, fragmented institutional responsibility, which are evidenced by post-enactment confusion among implementing authorities and divergent enforcement interpretations.

The GDPR's contested "right to explanation" exemplifies unintended ambiguity, with ongoing scholarly debate regarding whether Articles 13-15 and 22 mandate meaningful algorithmic explanations or merely procedural information. The distinction proves consequential: intentional ambiguity assumes institutional capacity for iterative refinement through implementation experience, while unintentional vagueness suggests transparency's promise exceeds policymakers' operationalization capacity.

The reality likely combines both—some ambiguity serves legitimate flexibility while other aspects reflect genuine limitations—with distinguishing between these requiring granular analysis of specific provisions' drafting histories beyond this study's scope.

This research examines algorithmic transparency through a comparative legal lens, investigating how different regulatory frameworks conceptualize and operationalize this principle, with particular attention to the experiences of both Global North (European, in particular) and Global South (Brazil, in particular) jurisdictions.

The analysis employs a mixed-methods approach, combining doctrinal examination of major international and regional instruments with qualitative interviews conducted with multistakeholder experts from academia, civil society, government, and industry sectors.

The paper is structured to provide a comprehensive examination of transparency as both legal concept and regulatory practice. Following this introduction, section 2 outlines the methodological approach, emphasizing the comparative and law-in-context methods employed to analyze regulatory frameworks and stakeholder perspectives.

Section 3 undertakes a conceptual analysis, examining how transparency is defined across different legal instruments, exploring the diversity of interpretations and applications, and distinguishing between procedural and material approaches to transparency requirements. Section 4 analyzes the purposes and perceived benefits of algorithmic transparency as identified through expert interviews, revealing transparency's instrumental role in facilitating accountability, oversight, and democratic participation. Section 5 critically examines the challenges and limitations facing transparency initiatives, including technical constraints, commercial resistance, regulatory fragmentation, and underlying power asymmetries that shape implementation patterns.

.

This research reveals a fundamental tension between transparency's democratic promise and its practical limitations. While European frameworks have established sophisticated procedural requirements for algorithmic disclosure, their global application remains uneven, with tech behemoths implementing differential transparency standards based on regulatory capacity rather than uniform principles.

This pattern reflects not merely technical constraints, but strategic compliance decisions that privilege markets with stronger enforcement mechanisms. Algorithmic transparency, while necessary for democratic governance, operates within structural constraints that limit its effectiveness in addressing global power imbalances in technology governance. Current frameworks often prioritize formal compliance over substantive accountability, creating a form of "performative transparency" that may legitimize existing arrangements rather than challenging them.

.

Despite these limitations, the research identifies pathways for enhancing transparency's democratic potential. Effective reform requires moving beyond purely procedural approaches toward comprehensive frameworks that integrate transparency with broader technological sovereignty initiatives. This includes developing domestic technical capacity, establishing regional regulatory cooperation mechanisms, and creating enforcement systems proportionate to algorithmic systems' societal impact.

The paper concludes that transparency's promise for democratic technology governance can only be realized through structural reforms that address underlying capacity asymmetries between regulatory authorities and technological actors. Rather than abandoning transparency as a regulatory tool, the analysis suggests that meaningful algorithmic governance requires embedding transparency within broader frameworks of technological justice that prioritize substantive accountability over formal compliance.

This investigation aims to contribute to emerging scholarship on technology regulation by providing empirical evidence. of how transparency operates in practice across different jurisdictions, while offering a critical assessment of its limitations and potential for democratic technology governance.

The research ultimately argues that transparency, properly understood and implemented, remains essential for accountability in algorithmic systems, but only when accompanied by structural reforms that address the fundamental power imbalances shaping contemporary technology governance.
.

2. Methodology

This study employs a mixed-methods approach, integrating a comparative legal analysis with qualitative interviews, to examine the intricacies of algorithmic transparency as a concept within the context of AI regulation. The research is specifically focused on how transparency is conceptualized, implemented, and perceived across different regulatory frameworks and by various stakeholders. The research is particularly interested in identifying potential divergences and convergences between different jurisdictions and perspectives.

In this study, various approaches were adopted to evaluate algorithmic transparency as a regulatory principle and its operational implications within AI systems. The methodology combined legal and doctrinal analysis, expert interviews, and an extensive literature review, each providing complementary inputs into the conceptualization, application, and limitations of algorithmic transparency in AI regulations. This allows for a comprehensive examination of transparency's regulatory dimensions and practical challenges.

This research seeks to answer questions about the role, and practical challenges of implementing algorithmic transparency within AI regulatory frameworks.

(i) First, it investigates how transparency is conceptualized and operationalized across legal, technological, and regulatory contexts, with a particular focus on European frameworks.

(ii) The research then explores the extent to which transparency requirements effectively support interoperability across diverse regulatory regimes, particularly in light of AI's transnational nature, and whether transparency can indeed serve as a harmonizing principle in AI governance.

(iii) The study also interrogates the functional limitations of transparency, asking whether transparency-based regulations sufficiently equip stakeholders —including regulators, developers, and users—to address the inherent opacity of advanced AI systems.

To answer these questions, I first adopt a comparative law perspective using a combination of functional and law-in-context methods. The functional method merges theoretical examination with black letter analysis of how legal the institutions chosen tackle the conceptual complexities of algorithmic transparency. By studying the functional equivalence between the regulatory approaches to transparency, it is possible to identify "similarity in difference."

My study builds upon the existing scholarship and proposes a comprehensive exploration of the complex nature of transparency as a legal principle, particularly focusing on technology regulation.

The primary methodological approach involves a legal analysis of major European and prominent international regulatory documents, including the Council of Europe Framework Convention on AI, the EU's General Data Protection Regulation (GDPR), the Digital Markets Act (DMA), the Digital Services Act (DSA), in addition to various OECD guidelines and reports.

Each document is examined for its specific references to transparency, exploring how this principle is framed in relation to other regulatory objectives by association, such as accountability, user protection, and trustworthiness. This analysis involved both a qualitative assessment of how transparency was presented and a quantitative examination of the frequency and contextual applications of the term.

While case law, administrative guidance, enforcement decisions, and secondary legislation undoubtedly shape how transparency principles operate in practice (for instance, Court of Justice of the European Union rulings interpreting GDPR transparency obligations, national data protection authority guidance on algorithmic explainability, or DSA implementing regulations) systematic analysis of these interpretive materials exceeds this research's scope. Such analysis would require jurisdictional depth incompatible with this study's comparative breadth across multiple frameworks and geographic contexts.

Large language models (LLMs) were used as analytical instruments to facilitate a systematic comparison of transparency mentions and associated terms across the documents. The use of ChatGPT (developed by OpenAI) and Claude (developed by Anthropic) was employed to highlight word frequency patterns, phrase contexts, and thematic associations with transparency, including identifiability, explainability, and accountability.

The second phase of the research entails the conduction of qualitative, semi-structured interviews with a diverse group of multistakeholder experts. These experts were purposefully selected from various sectors, including academia, civil society, government, and industry.

A fundamental element of the sampling strategy entailed the inclusion of participants from both the Global South (specifically, Brazil) and the Global North.

This approach was undertaken to investigate the potential geographic disparities in comprehension and experience concerning algorithmic transparency.

The interviews, of which there have been 12 thus far, have been executed in accordance with a semi-structured format, with the interviews guided by a set of predetermined questions. This approach permitted the exploration of emergent themes while ensuring the coverage of fundamental topics. The participants were informed about the research purpose and their right to anonymity. All interviews were recorded with the explicit consent of the participants. The collected data was subsequently transcribed using an AI tool (TurboScribe).

The interviews conducted in Portuguese were subsequently translated into English for the purpose of analysis through DeepL, another AI tool specialized in translation. The analysis of the interview data focuses on the identification of key themes, shared understandings, points of divergence, and specific examples related to algorithmic transparency. This encompasses an examination of the manner in which experts from disparate fields and regions delineate and interpret the concept of transparency.

The following aspects were considered in order to achieve a comprehensive understanding of the subject: firstly, the perspectives on the purposes and perceived benefits of the system (such as enabling accountability, oversight, and user autonomy); secondly, the challenges, limitations, and criticisms that have been raised. The interviews also illuminate the manner in which transparency is conceptualized in relation to other concepts, such as explainability, and its interaction with values such as privacy, security, and fairness. The integration of the findings from the comparative legal analysis and the qualitative interviews is intended to provide a nuanced understanding of algorithmic transparency as a complex and multi-faceted regulatory concept.

The use of diverse perspectives, as evidenced by the inclusion of voices from the Global South, is imperative for elucidating the practical challenges and varied interpretations of transparency in a globalized technological landscape, wherein AI systems frequently function across disparate legal and cultural contexts.

3. Defining and Conceptualizing Algorithmic Transparency

The Oxford English Dictionary defines "algorithm" as "a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer." While this definition may be technically accurate, it obscures the profound shift that algorithms represent in contemporary governance. Black's Law Dictionary's recent editions have begun incorporating technology-specific terminology, yet notably lack a standalone entry for "algorithm," a telling omission reflecting law's lag behind technological development.

"Transparency," by contrast, enjoys extensive treatment in both lexicons. The Oxford English Dictionary offers the quality of being "easy to perceive or detect" and, more pertinently, "openness; lack of hidden agendas or conditions, accompanied by availability of full information required for collaboration, cooperation, and collective decision making."

This second definition approximates the aspirational understanding deployed in regulatory discourse. Black's Law Dictionary defines transparency as "the quality or state of being transparent," with "transparent" denoting something "easily seen through or detected; readily understood; characterized by visibility or accessibility of information, especially concerning business practices." The legal dictionary's emphasis on visibility and accessibility (rather than mere disclosure) introduces a normative dimension absent from purely descriptive definitions. Legal transparency implies not only information provision but its meaningful accessibility to relevant parties.

When these terms combine into "algorithmic transparency," however, neither general nor legal dictionaries offer authoritative definitions. This lexicographical void reflects the concept's novelty and contested nature across disciplines, permitting divergent interpretations across regulatory contexts.

The literal understanding of algorithmic transparency suggests straightforward disclosure: revelation of code, data inputs, decision rules, and outputs. Yet regulatory frameworks rarely adopt such literalism. Instead, they construct transparency as a relational concept, calibrated to different audiences and purposes.

A software engineer might understand transparency as access to source code; a data subject might expect explanation of why an automated decision affected them; a competition regulator might prioritize transparency regarding market effects over technical implementation. This divergence reflects deeper tensions about transparency's proper scope and function.

On one hand, law instrumentalizes transparency, treating it as means toward accountability, fairness, contestability, and informed consent rather than an end itself. Literal transparency (complete algorithmic disclosure) may satisfy technical openness without advancing these legal objectives. On the other hand, legal frameworks may mandate measures providing limited insight into actual operations while serving regulatory purposes.

Legal conceptions also incorporate procedural dimensions absent from literal definitions. Transparency requirements often mandate specific formats, timing, and accessibility standards. Transparency reporting obligations may prescribe machine-readable formats and standardized categories, which are procedural specifications transforming transparency from general principle into enforceable obligation with verifiable compliance markers.

Furthermore, legal interpretations necessarily engage transparency's limits in ways literal definitions need not. Trade secrets, privacy rights, security concerns, and competitive dynamics all constrain comprehensive disclosure.

This distinction between transparency as complete disclosure and transparency as contextualized, purposeful information provision runs through subsequent analysis. Understanding this divergence proves essential for assessing transparency's practical limitations and democratic potential.

The subsequent analysis requires clarification regarding which frameworks merit inclusion and why. This research adopts an inclusive approach, recognizing that "algorithmic transparency" as a composite term remains absent from many foundational regulatory instruments despite their substantive engagement with the concept.

Three categories of frameworks inform this analysis.

First, certain technical standards explicitly employ "algorithmic transparency" or closely related formulations. The IEEE Standard 7001-2021, for instance, addresses "transparency of autonomous systems" with technical specificity.

Such frameworks provide the most direct engagement with the concept, offering operational definitions and measurable criteria.

Second, and more commonly, regulatory instruments address "transparency" within the specific context of AI, automated decision-making, or algorithmic systems without deploying the compound term itself. The EU AI Act exemplifies this approach, establishing extensive transparency obligations for AI systems, particularly high-risk systems, without consistently using "algorithmic transparency" as such.

Article 13's requirements for transparency and provision of information to deployers, alongside Article 50's transparency obligations for general-purpose AI systems, substantively address algorithmic transparency despite terminological variance. Similarly, the GDPR's provisions on automated decision-making and profiling establish transparency requirements applicable to algorithmic systems, even though the regulation predates widespread adoption of "algorithmic transparency" as regulatory vocabulary.

Third, certain frameworks address transparency in sufficiently general terms that their relevance to algorithmic systems emerges through interpretation and application rather than explicit textual reference.

The Council of Europe Framework Convention, for example, mandates "adequate transparency and oversight requirements tailored to the specific contexts and risks" for AI system lifecycles, employing transparency as a general principle adaptable across technological contexts. The OECD recommendations similarly frame transparency as a cross-cutting principle applicable to AI systems alongside other governance objectives.

This inclusive approach reflects practical realities of technology regulation. Jurisdictions and institutions adopt varied terminological conventions, influenced by drafting traditions, temporal factors, and institutional contexts. A framework's failure to employ "algorithmic transparency" as a term of art does not diminish its substantive contribution to conceptualizing and operationalizing the underlying principle.

The selection criteria prioritize frameworks that:
(i) establish binding or influential norms regarding information disclosure, explainability, or comprehensibility of automated systems;
(ii) represent significant jurisdictions or international bodies with norm-setting authority; and
(iii) demonstrate practical implementation through enforcement mechanisms or compliance frameworks.

This explains the inclusion of instruments ranging from the legally binding GDPR and AI Act to soft law instruments like UNESCO's Recommendation on the Ethics of Artificial Intelligence and OECD guidelines, which shape regulatory discourse despite lacking direct enforcement mechanisms. The subsequent comparative analysis therefore tracks how different instruments conceptualize transparency requirements for algorithmic systems, recognizing that semantic differences often mask substantial convergence, or, conversely, that terminological uniformity may obscure divergent operationalizations.

This methodological choice acknowledges that "algorithmic transparency" functions as an analytical category imposed by researchers and commentators rather than a universally adopted legal term.

The concept's utility lies in its capacity to organize disparate regulatory approaches around common concerns: opacity of automated decision-making, information asymmetries between system deployers and affected parties, and democratic accountability of increasingly autonomous technologies.

3.1 Legal and Regulatory Frameworks

Regulation of algorithmic transparency as a principle in AI governance reveals a legal approach that has become indispensable yet challenging to implement, from an interoperability perspective. An *ex ante* strategy of technology regulation has flourished in the EU over the last decade, and arguably influenced the international and foreign adoption of similar legal designs, reflecting what Bradford terms the *Brussels Effect*. Thinking of this intrinsic regulatory dynamics, in addition to an assumed initial hypothesis of mutual influence, this research elected the following legal documents (comprising, amongst others, regulations, recommendations, treaties, and interpretative studies) as a base-layer for comparison:

OECD	Document analyzed	Date enacted
European Union	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)	2016
European Union	Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services	2019
UNESCO	Recommendation on the Ethics of Artificial Intelligence	2021
IEEE	IEEE Standard for Transparency of Autonomous Systems (IEEE Std 7001-2021)	2021
European Union	Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)	2022
European Union	Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act)	2022
OECD	Common Guideposts to Promote Interoperability in AI Risk Management	2023
OECD	AI, Data Governance, and Privacy: Synergies and Areas of International Co-operation	2024
OECD	Recommendation of the Council on Artificial Intelligence	2024
OECD	Explanatory Memorandum on the Updated OECD Definition of an AI System	2024
European Union	Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828	2024
Council of Europe	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law	2024

Through the documents analyzed, from the Council of Europe's international human rights- and democracy-oriented framework to the OECD's focus on international regulatory harmonization and interoperability, the theme of transparency emerges as a fundamental component, albeit interpreted differently depending on the normative premises and objectives of each framework. The differences in these interpretations reflect a divergence in policy objectives from human rights and democracy to competition, consumer rights, and data protection.

While interoperability and harmonization constitute significant challenges in operationalizing algorithmic transparency across jurisdictions, they represent only one dimension of a multifaceted implementation landscape. The analysis of legal and regulatory frameworks in this section focuses primarily on how different instruments conceptualize transparency, revealing terminological variations and normative emphases that complicate cross-jurisdictional coordination.

However, subsequent sections demonstrate that effective transparency implementation confronts challenges extending well beyond regulatory alignment. Technical limitations inherent in complex machine learning architectures, competing commercial interests in proprietary system protection, asymmetrical enforcement capacities between Global North and Global South regulators, and fundamental tensions between transparency and other values such as privacy and security all constrain transparency's practical realization. This section therefore examines how frameworks define and structure transparency obligations, while sections 4 and 5 address the broader ecosystem of implementation challenges (including but not limited to interoperability) that determine whether formal transparency requirements translate into meaningful accountability mechanisms.

The emphasis on interoperability in this comparative analysis reflects the transnational nature of AI systems and the reality that major technology platforms operate across multiple regulatory regimes simultaneously. Divergent transparency standards create compliance complexity and enable strategic forum-shopping, whereby companies implement more rigorous disclosure practices in jurisdictions with stronger enforcement while maintaining opacity elsewhere. Yet even perfect regulatory harmonization would not resolve the underlying challenges that technology companies cite when resisting transparency mandates: the purported inscrutability of neural networks, legitimate security concerns regarding adversarial exploitation, and trade secret protections that shield competitive advantages.

Moreover, as expert interviews revealed, transparency's democratic potential depends not merely on regulatory design but on recipient capacity, both individual digital literacy and institutional technical expertise, to convert disclosed information into actionable oversight. The following examination of specific frameworks thus provides necessary groundwork for understanding how transparency is legally constructed, recognizing that this construction represents only the initial step in a complex implementation process shaped by technical, commercial, political, and epistemic factors analyzed throughout subsequent sections.

The Organisation for Economic Co-operation and Development

First, the privacy-oriented frameworks and guidelines of OECD tend to integrate transparency with a primary focus on data protection and user autonomy. Under OECD's broader privacy framework, transparency is seen as a precondition for ethical data governance, often intertwined with concepts of accountability, non-discrimination, and informed consent. The OECD Recommendation on Artificial Intelligence, for example, particularly emphasizes the significance of transparency as a cornerstone in ensuring explainability within AI systems. It mandates that AI actors provide comprehensive information regarding AI operations, potential risks, and inherent limitations, with the objective of fostering public trust and facilitating effective oversight.

> *1.3. Transparency and explainability: 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, including their capabilities and limitations, ii. to make stakeholders* ***aware of their interactions with*** *AI systems, including in the workplace, iii. where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and, iv. to provide information that enable those adversely affected by an AI system to challenge its output.*

As an underlying policy objective implicit in the document, this transparency is often linked with overarching principles of awareness, understanding, **meaningful information,** and stakeholder empowerment, to enable comprehension of the rationale behind AI-driven decisions and actions, particularly in instances where they may impact fundamental rights and safety.

The OECD's *Explanatory Memorandum on the Updated OECD Definition of an AI System* and the *Common Guideposts to Promote Interoperability in AI Risk Management* also explore transparency as a central aspect of both risk assessment and regulatory alignment. The OECD's emphasis on transparency within risk management frameworks suggests a more pragmatic view, one that seeks to harmonize risk-related disclosures across jurisdictions to alleviate compliance burdens. The idea behind positioning transparency as integral to responsible AI governance, also aligns with OECD objectives of free data flows, for example, where an AI ecosystem's transparency functions also as a facilitator of interoperability among disparate frameworks.

Transparency serves risk assessment by enabling systematic evaluation of AI systems' potential harms through disclosure of training data, model architecture, and testing protocols. Simultaneously, transparency facilitates regulatory alignment by establishing comparable disclosure standards that allow different jurisdictions to recognize and validate each other's oversight mechanisms.

These functions constitute complementary aspects of a unified governance objective: standardized transparency requirements enable both consistent risk evaluation within jurisdictions and mutual recognition across them.

The OECD's emphasis on transparency within risk management frameworks thus reflects a pragmatic strategy whereby harmonized risk-related disclosures reduce compliance burdens for multinational operators while maintaining regulatory effectiveness. Transparency operates as a common regulatory language, translating diverse institutional approaches into comparable information formats that permit cross-jurisdictional coordination. This positioning aligns transparency with broader OECD objectives including free data flows, where an AI ecosystem's transparency mechanisms facilitate interoperability among disparate frameworks by providing a foundational layer of shared information that different regulatory regimes can interpret according to their specific normative priorities.

A report by OECD on AI, data governance, and privacy also emphasizes the importance of transparency as a fundamental principle to ensure the responsible and ethical deployment of AI. In it, transparency is often conceptualized as a "responsible disclosure", which enables individuals to comprehend the utilization of AI, particularly in applications such as chatbots and other automated decision-making tools.

This disclosure is consistent with data protection principles, guaranteeing that individuals are informed about the data utilized by AI systems. The report also suggest some good practices related to data protection that can be incorporated into AI governance in order to increase transparency:

> *Initiatives and frameworks that have been suggested for enhancing transparency in the AI context can also follow similar methods. One such example is the use of "model cards" to report essential information about the characteristics of machine learning models. These cards can encompass a comprehensive range of metrics, evaluating bias, fairness, and inclusivity aspects (Margaret Mitchell, 2019), alongside providing insights into the provenance of the data, details of the statistical distribution of various factors in the training data sets, and other details on the data sets used in the creation of the model data's origin, statistical distribution of pertinent factors within the training datasets, and additional details pertinent to the datasets utilised in constructing the model. To enhance transparency in AI systems, organisations can establish dedicated organisational roles and functions, develop new policies and procedures, or revise existing ones. Additionally, documenting each stage of the design and deployment process of AI systems can facilitate the provision of meaningful explanations to affected individuals (ICO, 2020).*

The report indicates that there is a common objective between the AI and privacy policy communities to provide transparency that fosters user awareness, thereby supporting individual agency and informed consent. However, it also identifies challenges specific to AI, such as the "black box" problem in complex systems, which complicates the ability to provide transparency in AI outputs.

The United Nations Educational, Scientific and Cultural Organization

UNESCO's Recommendation on the Ethics of Artificial Intelligence frames transparency primarily through a human rights lens, positioning it as an ethical prerequisite for accountability rather than merely a technical or regulatory requirement. This rights-centered approach distinguishes UNESCO's framework from more technocratic formulations, emphasizing transparency's role in enabling contestation, facilitating democratic oversight, and preventing human rights violations.

The recommendation explicitly links transparency to explainability and accountability, arguing that transparent AI processes enable public scrutiny particularly when automated decisions affect safety, fairness, or fundamental rights:

> *37. Transparency is necessary for relevant national and international liability regimes to work effectively. A lack of transparency could also undermine the possibility of effectively challenging decisions based on outcomes produced by AI systems and may thereby infringe the right to a fair trial and effective remedy, and limits the areas in which these systems can be legally used.*

These principles find concrete manifestation in several high-profile cases where opacity undermined accountability and produced discriminatory outcomes. The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithmic risk assessment tool, widely deployed in the criminal justice systems in the United States, exemplifies transparency's absence hindering fair trial rights.

Courts have struggled to evaluate COMPAS-generated risk scores due to the proprietary algorithm's opacity, with defendants unable to meaningfully challenge assessments that influenced sentencing decisions. ProPublica's investigation revealed racial bias in COMPAS predictions, yet the **algorithm's black-box nature prevented effective judicial scrutiny or remediation**.

Similarly, Obermeyer and colleagues documented how a widely used healthcare algorithm in the United States systematically discriminated against Black patients, assigning them lower risk scores than equally ill white patients.

The algorithm's **lack of transparency obscured its reliance on healthcare costs as a proxy for health needs**, which is a metric that reflected existing inequities in access rather than objective medical necessity.

Only external research exposing the algorithm's logic enabled identification and correction of this bias, demonstrating how opacity can perpetuate discrimination invisible to both deployers and affected populations.

Australia's Robodebt scandal represents perhaps the most comprehensive illustration of automated systems' capacity for systemic harm absent transparency safeguards. The government deployed an income-averaging algorithm to identify allegedly overpaid welfare recipients, generating debt notices demanding repayment.

The system's opacity **prevented recipients** from understanding calculation methodologies or effectively contesting assessments.

Ultimately, the Royal Commission into the Robodebt Scheme determined the scheme unlawfully raised debts against hundreds of thousands of people, with devastating financial and psychological consequences including suicides linked to the program. It found that lack of transparency (both to recipients and within government itself) enabled the scheme's continuation despite its unlawfulness, demonstrating how opacity facilitates administrative illegality and undermines rule of law.

These cases substantiate UNESCO's assertion that transparency constitutes a prerequisite for effective liability regimes and meaningful contestation. Where algorithmic systems remain opaque, affected individuals **cannot mount** informed challenges, regulators **cannot conduct** meaningful oversight, and discriminatory or unlawful operations **persist undetected**. The examples also reveal transparency's limitations: even after exposure, remediation depends on political will and institutional capacity. Nonetheless, they demonstrate that opacity guarantees injustice, whereas transparency at minimum creates conditions for accountability.

UNESCO's recommendation ventures further, suggesting proceduralized transparency mechanisms that approach material disclosure, including potential access to source code and datasets, measures that have historically faced substantial resistance from developers citing intellectual property and data protection concerns. This resistance finds precedent across multiple domains:

> *39. Specific to the AI system, transparency can enable people to understand how each stage of an AI system is put in place, appropriate to the context and sensitivity of the AI system. It may also include insight into factors that affect a specific prediction or decision, and whether or not appropriate assurances (such as safety or fairness measures) are in place. In cases of serious threats of adverse human rights impacts, transparency may also require the sharing of code or datasets.*

Additionally, UNESCO's recommendation stresses that transparency should be balanced with privacy and security, recommending context-appropriate transparency measures to foster trust and accountability in AI deployment.

This formulation reflects a deliberate choice to avoid prescriptive technical specifications, instead affording developers discretion in designing and implementing transparency mechanisms suited to particular deployment contexts.

Such flexibility acknowledges the heterogeneity of AI applications and the impossibility of uniform technical solutions across disparate use cases, but it simultaneously renders transparency conceptually and operationally fluid, which is a standard subject to variable interpretation.

This interpretive latitude creates space for what might be characterized as strategic compliance, wherein organizations satisfy formal transparency obligations through minimalist disclosure that meets literal requirements **without enabling meaningful oversight.** The phenomenon finds parallel in GDPR privacy notice practices, where lengthy, legally compliant disclosures nonetheless fail to inform users meaningfully about data processing practices.

Research demonstrates that organizations exploit discretion regarding "clear and plain language" requirements to produce notices that technically comply while remaining functionally opaque to average users. Context-appropriate transparency risks similar outcomes: developers may invoke contextual specificity and security concerns to justify limited disclosure, asserting compliance with flexibility-oriented standards while resisting substantive transparency that might enable external scrutiny.

The **challenge intensifies** across jurisdictions with varying regulatory capacity. Context-appropriate transparency may provide more robust disclosure in jurisdictions with sophisticated regulators capable of demanding clarification and rejecting inadequate implementations, while the same standard permits minimal compliance in jurisdictions lacking technical expertise or enforcement resources. This regulatory arbitrage undermines transparency's harmonizing potential, creating tiered accountability regimes wherein the same AI system operates with different transparency levels depending on deployment location.

UNESCO's framework thus embodies a hybrid approach: it combines prescriptive principles (transparency as necessary for human rights protection and liability regimes) with discretionary implementation (context-appropriate measures balancing competing values).

This hybridity reflects transparency's dual nature as both legal standard and governance objective.

As a legal standard, transparency requires sufficient specificity to enable compliance assessment and enforcement, which are characteristics that favor prescriptive rules with clear boundaries.

As a governance objective, transparency must adapt to technological evolution, diverse application contexts, and legitimate competing interests, characteristics favoring discretionary standards and contextual judgment.

The tension between these dimensions proves difficult to resolve. Overly prescriptive standards risk obsolescence as technologies evolve and may impose requirements ill-suited to particular contexts, while purely discretionary standards sacrifice predictability and enable strategic non-compliance. UNESCO's approach seems to prioritize adaptability, accepting reduced legal certainty as the price of contextual responsiveness. This choice aligns with soft law's characteristic flexibility but raises questions about enforceability and consistency absent accompanying mechanisms to constrain discretion or harmonize interpretation.

The practical effect is that transparency becomes an **evolving and inherently uncertain standard**, subject to iterative negotiation between regulators, industry, civil society, and affected communities. This fluidity may reflect transparency's necessary character in rapidly changing technological contexts, but it also fundamentally limits transparency's capacity to function as a stable legal protection.

The practical effect is that transparency becomes an **evolving and inherently uncertain standard**, subject to iterative negotiation between regulators, industry, civil society, and affected communities. This fluidity may reflect transparency's necessary character in rapidly changing technological contexts, but it also fundamentally limits transparency's capacity to function as a stable legal protection. Where standards remain indeterminate, rights become aspirational rather than enforceable, and accountability depends more on regulatory vigilance and corporate goodwill than on clear legal obligation.

For practitioners and policymakers, this suggests that transparency frameworks require complementary mechanisms, such as detailed implementing regulations, standardized reporting templates, independent auditing requirements, or judicial interpretation establishing concrete parameters, in order to translate flexible principles into meaningful constraints on algorithmic opacity.

Council of Europe

The Council of Europe's *Framework Convention on Artificial Intelligence, Human Rights, Democracy, and Rule of Law* (CAI) represents the first legally binding international treaty specifically addressing AI governance.

Once ratified by member states, the CAI will oblige parties to enact domestic legislation or adopt regulatory measures embodying its core principles, creating binding obligations enforceable through international law mechanisms.

This legal architecture positions the CAI as potentially transformative for transparency standardization across diverse jurisdictions, though its effectiveness depends critically on how states translate treaty obligations into implementable national frameworks.

The CAI articulates transparency as a fundamental safeguard for democratic values, dedicating Article 8 to transparency and oversight requirements:

> *Article 8 – Transparency and oversight: Each Party shall adopt or maintain measures to ensure that adequate transparency and oversight requirements tailored to the specific contexts and risks are in place in respect of activities within the lifecycle of artificial intelligence systems, including with regard to the identification of content generated by artificial intelligence systems.*

From a legal conceptualization perspective, Article 8 embodies significant interpretive flexibility that may facilitate either harmonization or divergence.

The requirement for "adequate" transparency "tailored to the specific contexts and risks" avoids prescriptive specifications, instead establishing a principles-based standard permitting contextual adaptation. This reflects international treaty conventions accommodating diverse legal systems and institutional capacities.

However, this flexibility creates substantial divergence risks. States may interpret "adequate" transparency vastly differently based on domestic priorities, regulatory capacity, and industry influence. The AI-generated content identification requirement illustrates these tensions: the principle commands agreement, but operationalization varies depending on whether states mandate technical watermarking, metadata disclosure, interface indicators, or voluntary labeling.

A state might satisfy Article 8 through minimal disclosure technically complying with treaty language while failing to enable meaningful accountability.

This architecture has profound implications for transparency's legal conceptualization. For states with sophisticated AI frameworks, particularly EU members implementing the AI Act, the convention functions as a floor, with domestic law exceeding minimal treaty obligations.

For states lacking comprehensive AI legislation, the convention provides a template, potentially accelerating regulatory convergence around core principles even as implementation details diverge.

Transparency here is framed within a broader commitment to uphold human dignity and the rule of law, requiring, for instance, that content generated by AI systems be explicitly identifiable. Or, for instance, in the case of an AI system used for automated welfare eligibility assessments, transparency across the system's lifecycle may plausibly require that the datasets used for training be documented and auditable; that design and testing processes maintain clear records of model decisions. This demand for traceability aims to ensure that AI systems function in a manner that respects individual rights, thus reinforcing public trust in digital systems that intersect with everyday human experiences.

This interpretation of transparency is, however, not without its complications; identification of AI generated content introduces a level of operational complexity, as it compels AI developers and deployers to make algorithmic outputs labeled and comprehensible by individuals, without compromising system integrity or function.

Beyond interpretive flexibility, the CAI's transparency obligations raise unaddressed practical challenges. **Provenance traceability** (documenting training data origins, model iterations, and decision pathways) **presents formidable hurdles in contemporary AI development** involving distributed teams, third-party datasets, and open-source components.

Maintaining auditable records requires sophisticated infrastructure and cross-functional coordination between data scientists, engineers, and compliance teams. For complex models incorporating federated learning or foundation model fine-tuning, establishing clear provenance becomes exponentially difficult as contributing factors span multiple organizations and jurisdictions.

Model explainability presents related challenges: transparency mandates demand explanations of system outputs, but technical capacity to generate meaningful explanations remains limited for state-of-the-art deep learning architectures.

Post hoc explainability techniques provide approximations rather than complete accounts, creating a compliance paradox where regulations demand explainability that the technical state of the art may not fully deliver.

The CAI's silence on implementation support or proportionality principles leaves these burdens unmitigated, potentially undermining transparency's democratic objectives by making compliance itself a barrier to responsible AI development.

As of yet, the CAI has not entered into force. Adopted in May 2024, the CAI was open for signatures in September. The framework convention represents the first legally binding international treaty specifically addressing AI governance. Opened for signature in September 2024, it extends participation beyond members of the Council of Europe to include non-member states and the EU itself, which reflects its global normative ambitions.

However, signature indicates political commitment rather than legal obligation: the framework convention requires ratification through domestic constitutional processes and enters force following ratification by five states, including three members of the Council of Europe. As a binding treaty rather than soft law, it creates enforceable obligations for ratifying parties, though it lacks EU law's supranational enforcement mechanisms and depends primarily on diplomatic pressure for compliance.

Numerous Council of Europe members have signed, but the ratification progress is limited, and major AI-developing nations outside Europe (including the United States, China, and India) have not committed to ratification.

Institute of Electrical and Electronics Engineers

On the technical community sphere, the 2021 edition of the IEEE Standard 7001-2021, entitled "*IEEE Standard for Transparency of Autonomous Systems,*" offers a birds' eye view on transparency standards for the industry.

It employs a framework-based approach to transparency, delineating and defining quantifiable, testable levels of transparency for autonomous systems (which are defined as "a system that has the capacity to make decisions itself in response to some input data or stimulus with a varying degree of human oversight or intervention depending on the system's level of autonomy").

This standard highlights the necessity for transparency to encompass not only the transfer of information from autonomous systems to stakeholders who have a **right, share, claim or interest** in a system but also the assurance that this information is presented in a manner that is readily comprehensible to those stakeholders.

The standard categorizes stakeholders into groups with varying transparency requirements, such as non-expert users, domain experts, and auditors, which is normatively a sound approach to transparency, because it recognizes transparency conceptually is also relational. Each of these groups requires different levels of detail to achieve transparency appropriate for their roles. Conceptualizing transparency in a practical way, the standard states that:

> *Transparency refers to a transfer of information from an autonomous system or its designers to a stakeholder that is truthful; contains information relevant to the causes of some action, decision, or behavior; and is presented at a level of abstraction and in a form meaningful to the stakeholder. Transparency should be mindful of the stakeholders' likely perception and comprehension, and should avoid disclosing information in a manner that, while technically true, is framed in a way that leads to misapprehension.*

This definition highlights transparency's essential elements for **IEEE, relating** it to **truthfulness, relevance,** and stakeholder comprehension, reinforcing its relational nature. However, these ancillary constituent elements are not conceived or defined in other regulations, guidelines and interpretive documents analyzed.

Beyond its relational nature, this also indicates that transparency, similar to other legally significant terms, is not **a static or fixed construct** but one that admits a **multiplicity of interpretations and applications** depending on differing functional needs, technical architectures, and risk vectors of the system.

A facial recognition system deployed in law enforcement contexts would require fundamentally different transparency mechanisms than a content recommendation algorithm on a social media platform, yet both operate under generic "transparency" mandates that provide limited guidance regarding appropriate disclosure depth, target audiences, or verification methods.

This contextual variability creates significant interpretive uncertainty regarding transparency's exact scope and legal effect in algorithmic contexts.

Regulators, deployers, and affected parties may hold dramatically different understandings of what transparency requires for particular systems, with no authoritative mechanism to resolve disputes absent litigation or regulatory enforcement actions.

The absence of standardized transparency definitions across frameworks thus enables strategic ambiguity wherein organizations claim compliance through minimalist disclosure while critics demand comprehensive access, with neither position definitively supported or refuted by existing legal standards.

The 7001-2021 framework regards transparency as a critical factor in accountability, trust, and safety, particularly in instances where autonomous systems operate in conjunction with humans or make decisions with significant consequences. In order to achieve this, the standard defines specific "transparency levels," which range from basic operational transparency to more complex forms that might include real-time explanations, interactive user functionalities, and transparency for auditing and validation purposes.

The standard defines progressive transparency levels for different stakeholder needs, from "no transparency" at **Level 0** to "continuous explanation of behavior" at **Level 5**, where the system provides adaptive explanations based on user interaction history.

For example, a Level 3 transparency requirement for non-expert users includes "user-initiated functionality that produces a brief and immediate explanation of the system's most recent activity" in plain language, such as a robot explaining, **"I stopped because I am programmed not to bump into you."** By contrast, Level 5 transparency would provide continuous, adaptive explanations calibrated to the user's demonstrated expertise and past interactions.

The same robot operating at Level 5 might continuously communicate: **"Proximity sensor detected object at 0.5 meters, initiating deceleration protocol. Based on your previous requests for detailed information, this followed decision tree path A2 prioritizing collision avoidance over task completion. Object identified as human using thermal signature. Would you like to adjust the collision threshold for future encounters, or review the sensor data log?"**

This illustrates how Level 5 transparency transitions from reactive, simple explanations to proactive, contextual, and technically detailed disclosure that adapts to user sophistication and enables parametric adjustments, which is a level of transparency functionality rarely implemented in deployed systems due to complexity and resource requirements.

For validation and certification stakeholders, as a parallel to audits strategies present in the DSA, the IEEE standard requires the system to supply documentation, specifications, models, and test results, escalating to full source code access at higher transparency levels. This approach enables agencies to verify a system's compliance and reliability effectively, which can already be a benchmark for future EU harmonization approaches. Furthermore, the standard emphasizes that for transparency to be truly effective, it must be embedded in the system's design phase (transparency by design), rather than being retrofitted into a system post-design. This is a recommendation also adopted for privacy within the GDPR (privacy by design).

In alignment with the broader regulatory landscape examined in this research, IEEE Standard 7001-2021 reflects a technically grounded, stakeholder-specific approach to transparency. Its objective seems to be to ensure that autonomous systems remain comprehensible, accountable, and ethically aligned with human values across diverse application contexts.

European Union

Concerning the right to access information from the data subject's standpoint, this right relating to processing data "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" has been established in Article 12 of the GDPR. This provision is linked to the purpose of reducing knowledge asymmetries in favor of individuals. Article 12 of the GDPR is also related to consumer rights, because it specifically requires transparency, intelligibility, and accessibility of form and language, since "the wording of the new regulation echoes typical consumer protection clauses." This is deemed as the first layer of transparency, according to a systematic interpretation of the GDPR. Individual information and access rights for data subjects provide a better understanding of how an automated decision is made, the logic behind the algorithm used in decision making, and the anticipated consequences of the decision. Users with more information are able to make informed choices. More information also contributes to inspiring more confidence in the products and services offered.

The GDPR's conceptualization of transparency differs significantly from frameworks examined elsewhere in this analysis. Unlike the DMA and DSA, which position transparency instrumentally as a mechanism for achieving competition and content moderation objectives, the GDPR treats transparency as both instrumental (enabling rights exercise) and intrinsic (constituting respect for data subject autonomy).

Where the AI Act emphasizes technical documentation for regulatory oversight, the GDPR prioritizes individual comprehensibility and actionable information for data subjects themselves. The Council of Europe's CAI articulates transparency as a general governance principle requiring contextual implementation, while the GDPR specifies detailed procedural obligations with enforceable individual rights. This specificity—combined with the dual privacy-consumer protection foundation—positions GDPR transparency as **more expansive** in individual empowerment objectives yet narrower in scope than AI Act requirements addressing non-personal-data algorithmic systems. The GDPR's approach thus illustrates transparency operating simultaneously across multiple normative registers (privacy, consumer protection, autonomy), contrasting with single-objective frameworks that risk reducing transparency to mere procedural compliance divorced from rights protection.

Other technology regulation frameworks, the EU's *Digital Markets Act* (DMA) and *Digital Services Act* (DSA), adopt transparency primarily as a tool for fostering fair competition and user protection. These regulations mandate transparency in terms of advertising practices, ranking algorithms, and content moderation policies, thus aiming to **mitigate the informational asymmetries** between major digital platforms and their business users or end consumers. In the DMA, transparency is often associated with a procedural means of achieving fair competition among selected market players:

> *Article 6.5 - The gatekeeper shall not treat more favourably, in ranking and related indexing and crawling, services and products offered by the gatekeeper itself than similar services or products of a third party. The gatekeeper shall apply transparent, fair and non-discriminatory conditions to such ranking.*

This focus is reinforced by Recital 45 of the same regulation, which approaches the conditions under which gatekeepers provide online advertising services to business users, including both advertisers and publishers, as often being non-transparent and opaque. This opacity is partly linked to the practices of a few platforms, but is also due to the sheer complexity of modern day programmatic advertising and potentially locks consumers into arrangements with specific platforms.

Furthermore, the reduced competition and transparency can lead to higher overall costs for online advertising services, which are then likely to impact the prices that end-users pay for everyday goods and services that rely on online advertising.

To address this issue, the referred Recital, alongside Recital 58 of the DMA, advises gatekeepers to provide advertisers and publishers, as well as third parties authorized by advertisers and publishers, upon request, with cost-free access to information on the price breakdown for each type of online advertising service within the advertising value chain. The guidance also mentions measuring tools and data, including aggregated and non-aggregated data, as core elements of transparency. It appears the intended purpose of these suggested transparency measures is to allow for independent verification of digital services.

A secondary purpose articulated in the DMA is both related to contestability of profiling practices and to leveling the playing field among players that may not have access to the same databases:

> *Recital 72 - [,,,] The data protection and privacy interests of end users are relevant to any assessment of potential negative effects of the observed practice of gatekeepers to collect and accumulate large amounts of data from end users. Ensuring an adequate level of transparency of profiling practices employed by gatekeepers, including, but not limited to, profiling within the meaning of Article 4, point (4), of Regulation (EU) 2016/679, facilitates contestability of core platform services. Transparency puts external pressure on gatekeepers not to make deep consumer profiling the industry standard, given that potential entrants or start-ups cannot access data to the same extent and depth, and at a similar scale. Enhanced transparency should allow other undertakings providing core platform services to differentiate themselves better through the use of superior privacy guarantees.*

This regulatory focus on transparency seems to **seek to redress imbalances of power in digital market**s by promoting a level playing field, ensuring that users are not **only informed** but **empowered to make choices** based on a clear understanding of algorithmic influences.

The occurrences of transparency in the DSA reflect a different perspective on transparency, one that is much more focused and detailed with regard to how this concept ought to be proceduralized and operationalized. Providers of intermediary services and online platforms find specific chapters in Articles 15 and 24, respectively, that describe these steps.

Common measures include, but are not limited to: yearly transparency reports regarding content moderation that must be publicly available, in machine-readable format, in easily comprehensible language; disputes submitted to dispute settlement

its outcomes, and the median time needed for completing the dispute settlement procedures, as well as the share of disputes where the provider of the online platform implemented the decisions of the body; number of suspensions imposed against users; average monthly active users in the EU; among other provisions.

An even greater emphasis on transparency is given to platforms that use recommender systems, which must inform, in their terms and conditions, "in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters", according to Article 27. The provision also includes an obligation to divulge "the criteria which are most significant in determining the information suggested to the recipient of the service" and "the reasons for the relative importance of those parameters".

In addition to the transparency information already required, since this is also an asymmetric regulation, for the so-called very large online platforms (VLOPs) or very large online search engines (VLOSEs) it is necessary to **publicize** the number of people employed in content moderation, their linguistic expertise and qualifications, indicators of accuracy, risk assessment **reports, mitigation measures, audit reports, and the average monthly users of the service** in each Member State of the EU. This, in turn, is creating an entire secondary service industry related to these compliance measures. In its Recitals, the DSA also provides interpretation input on how to frame transparency principles within its implementation, relating it to specific ex ante due diligence obligations:

> *Recital 40. In order to achieve the objectives of this Regulation, and in particular to improve the functioning of the internal market and ensure a safe and transparent online environment, it is necessary to establish a clear, effective, predictable and balanced set of harmonised due diligence obligations for providers of intermediary services.*

As can be observed, the legal design supporting these measures is arguably comprehensive and meticulous, seemingly tailored to align with a compliance checklist. This precision, however, creates risks alongside benefits.

The granular specification of transparency obligations may facilitate enforcement through clear benchmarks, yet simultaneously invites tick-box compliance wherein corporations prioritize technical adherence over substantive transparency objectives.

Platforms may satisfy each enumerated requirement (publishing transparency reports in machine-readable formats, disclosing specified metrics, providing requisite documentation) while structuring these disclosures to minimize practical utility for oversight. Compliance becomes performative: extensive documentation demonstrates regulatory conformity without enabling meaningful public scrutiny or informed user decision-making.

Moreover, while the DSA's framework is comprehensive from a regulatory drafting perspective, this does not necessarily enhance transparency for end-users who may find these measures effectively opaque despite formal accessibility. The sheer volume and technical complexity of mandated disclosures (content moderation statistics, dispute resolution metrics, algorithmic ranking parameters, risk assessment reports) can overwhelm rather than inform intended audiences.

Individual users navigating multi-page transparency reports filled with statistical tables and technical terminology face substantial barriers to extracting actionable information, while researchers and civil society organizations require significant resources to systematically analyze and interpret platform disclosures across multiple jurisdictions and reporting periods.

The overarching transparency principle evident in other regulations manifests in tangible forms and directives within the DSA, yet implementation must grapple with the practical realities of enforcing meaningful transparency. The technical complexity inherent in ranking algorithms and advertising models presents significant barriers to achieving transparency that is simultaneously accessible to diverse users, verifiable by regulators, and implementable by platforms without excessive compliance burden, which are tensions the DSA's detailed requirements make visible without fully resolving.

The more recent EU AI Act's transparency requirements, particularly as outlined in Article 13, **mandate** that high-risk AI systems must be designed and developed with a level of transparency **sufficient to enable those who deploy them to interpret outputs meaningfully and to utilize the system appropriately.** This stipulation entails that the transparency provided must align with the regulatory obligations of providers and deployers, ensuring clarity and accessibility of essential information.

> *Article 13.2. 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.*

Article 13 emphasizes the necessity for high-risk AI systems to be accompanied by comprehensive instructions that include accurate, concise, and readily understandable information pertinent to the effective utilization by the deployer, which is coherent to an asymmetric *ex ante* legal design. In particular, the instructions focus on deployers' interpretability (or understandability) and address several key areas, including the system's intended purpose, operational accuracy, robustness, cybersecurity measures, and any potential scenarios that could affect the system's expected performance.

For example, the documentation must make clear the system's constraints, foreseeing any potential hazards to health, safety, or fundamental rights, particularly in instances of misuse or when specific groups of individuals are involved. Furthermore, the AI Act stipulates that providers must include technical specifications and metrics related to input data quality, data set origins, and testing protocols, particularly in relation to the AI system's intended functions.

The AI Act also requires providers to disclose information that enables deployers to interpret the outputs of the AI system.

This encompasses guidance on human oversight mechanisms, which emphasize technical measures to assist deployers in monitoring and responding to the system's performance. Furthermore, providers have an obligation to inform deployers of any pre-determined system alterations, essential maintenance procedures, and the computational resources required for optimal functionality.

Additionally, the AI Act incorporates provisions for logging capabilities, stipulating that deployers have the requisite tools to collect, store, and interpret logs, which record data that capture a sequence of events, system states, or operational activities within the AI system over time. These logs contain information such as timestamps, user actions, system outputs, error reports, and changes in configuration or performance, which are essential elements for having an auditable record of how systems operate (accountability and compliance).

A considerable effort will be made over the next years towards harmonization of standards related to AI.

The European Commission has announced steps to establish harmonized standards for the AI Act in order to address areas such as transparency, risk management, and

technical requirements to ensure that AI systems are safe, trustworthy, and respect fundamental rights.

Harmonized standards for the AI Act, **provided** they are published in the Official Journal of the EU, **will grant a legal presumption of conformity to AI systems** developed in accordance with them.

The initiative is calling on European standardization bodies to develop guidelines that support compliance, with a focus on protecting both consumers and industry stakeholders. The EU has also recently appointed a group of experts to help guide compliance with the AI Act, as reported by Reuters. This multistakeholder team, which includes representatives from industry, academia, and civil society, will work alongside the European Commission to develop practical guidance and standards to AI systems.

In the case of the EU AI Act, standards ought to explicitly address and prioritize the potential risks that AI could pose to the health, safety, and fundamental rights of individuals. However, current international standardization efforts tend to prioritize the protection of organizational objectives in the context of AI. There are significant differences between managing risks to organizational objectives and addressing the potential risks of AI systems to individuals.

It is therefore essential that standards supporting the implementation of the AI Act prioritize the latter.

Previously analyzed in this piece, the IEEE Standard 7001-2021 introduces a granular approach towards transparency that accounts for the diverse needs of stakeholders, such as non-expert users, domain experts, and incident investigators. These levels range from basic operational transparency to detailed data logging and real-time explanations, designed to allow stakeholders to understand, assess, and potentially intervene in the system's operations.

The IEEE 7001-2021 presents a functional and stakeholder-oriented approach that prioritizes user-specific transparency needs. In contrast, the AI Act embeds transparency within a regulatory compliance framework aimed at safeguarding societal interests.

While IEEE's Standard 7001-2021 primarily frames transparency in terms of stakeholder-specific information access and testable transparency levels for autonomous systems, the EU AI Act incorporates transparency as a legal compliance requirement, especially for high-risk AI systems, with a focus on public health, safety and fundamental rights.

The AI Act mandates that high-risk AI systems provide transparency information that includes operational characteristics, potential risks, and limitations. Transparency within the AI Act is thus embedded as a compliance-driven obligation, encouraging harmonized standards to foster trust and predictability across the EU. This compliance orientation promotes transparency as an instrument for market uniformity and user protection, with an explicit mandate to facilitate oversight and a specific alignment with fundamental rights.

It has been reported that the EU AI Act standardization process is facing hurdles trying to harmonize the vertical industries that the regulation covers horizontally, which is a legal design problem of the regulation. There is an intense debate between the technology sector and industries covered by product safety legislation on how prescriptive the standards should be and how this would impact compliance costs.

3.2. Diverse Definitions and Interpretations

Transparency in technology regulation should not be restricted to just a matter of principle, but also be accompanied by a relational process of implementation. One needs to put forth a systematic and interoperable approach towards algorithmic transparency, beyond the contours of just the GDPR, the EU AI Act, or the DSA. A few questions could guide this approach: ***What information would be valuable for users and data subjects to understand the decision-making criteria of AI systems? What is their average knowledge on the topic?***

An ideal transparency regime must prioritize the informational needs and comprehension capacities of those subject to algorithmic systems (whether individual end-users or business users) rather than focusing solely on technical disclosure of algorithmic mechanics. Effective transparency centers on recipients' ability to understand, evaluate, and act upon information rather than mere provision of algorithmic details. This user-centric approach requires transparency mechanisms tailored to specific audiences: explanations enabling individuals to comprehend how decisions affecting them were reached, operational parameters allowing business users to optimize platform engagement and contest unfair treatment, and technical documentation permitting regulatory verification of compliance. The critical distinction lies between transparency designed to satisfy formal disclosure obligations and transparency calibrated to empower its intended recipients with actionable understanding.

This analysis has attempted to perform a broad-reaching comparative overview, one that encompasses matters related to data protection, competition, risk assessment, standardization and fundamental rights. An approach focused solely on the GDPR and the systematic interpretation of its provisions, for example, though well-intentioned, would be very limited.

Current regulatory approaches suffer from significant fragmentation, with transparency requirements dispersed across domain-specific instruments, such as data protection (GDPR), competition (DMA), content moderation (DSA) and AI systems (AI Act), which operate largely in silos despite addressing overlapping technological systems and corporate actors.

This fragmentation creates coordination challenges, inconsistent transparency

standards across regulatory domains, and opportunities for strategic compliance in which platforms satisfy narrow sectoral obligations while evading holistic accountability.

A recommendation algorithm, for instance, implicates data protection (personal data processing), competition (self-preferencing), consumer protection (manipulative practices), and fundamental rights (freedom of expression), yet transparency requirements **vary substantially** across these frameworks with minimal coordination mechanisms ensuring comprehensive oversight.

The siloed approach reflects institutional path dependencies and regulatory specialization, but proves increasingly inadequate for governing multifaceted algorithmic systems whose harms transcend traditional regulatory boundaries. Effective transparency governance would require integrated frameworks that coordinate across data protection authorities, competition regulators, consumer protection agencies, and sectoral supervisors, enabling comprehensive assessment of algorithmic systems' cumulative impacts rather than fragmented evaluation through narrow regulatory lenses, which is an institutional transformation that current frameworks gesture toward without fully achieving.

Civil society, academia, policymakers, and regulators, in general, ought to remember that "what we do and don't know about the social (as opposed to the natural) world is not inherent in its nature, but is itself a function of social constructs." That is to say that the amount of obligations we propose for companies, the standards that we set for algorithms, and the regulations that we impose on governments are all societal choices, expressed by law.

Thus, the workings of market players can be subjected to a transition from black boxes to transparency and accountability, in the same way that transparency is expected from political actors and democratic institutions exercising their mandates in society. By means of some of the instruments analyzed throughout this thesis, the EU has certainly taken the most steps in this direction.

However, whether these developments move in the right direction remains contested: while European frameworks establish important precedents for algorithmic accountability, they risk entrenching procedural transparency that legitimizes existing power structures rather than challenging them, particularly as implementation disparities create stratified transparency regimes wherein Global South jurisdictions lack enforcement capacity to compel meaningful disclosure

from the same platforms operating with relative opacity beyond Europe's regulatory reach.

Significant gaps persist in translating formal transparency requirements into substantive accountability mechanisms capable of addressing systemic algorithmic harms and redistributing technological decision-making authority from concentrated corporate control toward democratic governance.

The aspect of improving social standards through the enforcement of laws is at the core of regulating technologies. If one proceeds on the basis of the hypotheses that algorithmic technologies are supposed to contribute to the economic, social, and educational development of our society, then these products of human ingenuity ought to act in service of individual users, and not the other way around. To that end, as Dignum observed, "AI reasoning should be able to take into account societal values, moral and ethical considerations; weigh the respective priorities of values held by different stakeholders in various multicultural contexts; explain its reasoning; and guarantee transparency."

An additional challenge that the conceptualization of transparency has to face is that there is no actual agreement on what transparency is.

Indeed, some works provide transparency as a set of meaningful information about a system, others a decision tree, others a record of logs, and so on. Moreover, transparency can be case-based, contextual, contrastive, counterfactual, scientific, simulation-based, statistical, and trace-based, among others.

For example, a relational concept of transparency may associate datasets with the algorithmic models and the public to which appropriate explanations must be provided. Comprehensive transparency is necessarily complex, involving a generalization of principles and, at the same time, a particularization of parameters, which is not an easy task to achieve.

Copyright law, for example, must reconcile universal principles of authorship protection and public interest with highly contextual determinations of originality thresholds, fair use applicability, and derivative work status.

Similarly, tort law translates broad normative principles, such as duty of care, reasonableness and foreseeability, into fact-specific judgments varying dramatically across circumstances, relationships, and technological contexts.

Both domains demonstrate that coherent legal frameworks emerge through sustained evolution rather than comprehensive initial codification. Copyright and tort doctrines have developed over centuries through continuous adaptation to shifting social norms, technological change, and accumulated judicial interpretation, which are processes that remain ongoing today as courts grapple with digital technologies, AI, and novel social practices.

Algorithmic transparency faces analogous developmental challenges. Just as copyright law required generations to establish workable doctrines balancing creator incentives against public access, and tort law continuously refines reasonableness standards for emerging risks, transparency frameworks require sustained contestation, refinement, and adaptation over time.

The current regulatory moment represents early stages of this evolution, with frameworks like the GDPR, AI Act, and DSA establishing foundational principles that will inevitably require substantial elaboration through enforcement practice, judicial interpretation, and iterative legislative revision as technological capabilities advance and societal expectations shift.

Expecting comprehensive, stable transparency frameworks to emerge immediately reflects unrealistic expectations inconsistent with legal development patterns across other complex domains. The **conceptual unsettledness** characterizing contemporary transparency discourse may thus represent not regulatory failure but rather the natural condition of emergent legal frameworks addressing novel challenges whose full dimensions remain contested and incompletely understood.

Any approach to transparency has to be focused on the user, that is, focused on its target audience and context. Relational transparency takes into account its audience and provides useful information accordingly.

However, what is useful to one user may not be useful for another. Just like consumer protection takes into account information asymmetries in order to assess the right amount of information necessary to inform transactions, the transparency of algorithmic decisions must also be adequate for its target audience, the best possible version of an "average user" of the platform. As Powell explains, "the level, quality and target of explanation became a significant issue of governance, because there is no standard format for explanation that would apply to all algorithmic or AI systems."

Relational transparency also employs an appropriate vocabulary, one that will be palatable and intelligible for the user in order to get the message across. For example, since accurate vocabulary may be excessively technical and even cause greater confusion, especially for end-users, plain language and even a visual explanation, which does not provide comprehensive details but is easier to understand by those who are not educated in the area, may be more suitable to them.

When it comes to business users, greater levels of detail, technical vocabulary, and more intricate parameters of the inner-functioning of the platform may not only be more suitable, but also more useful and a necessity.

> *All computer programs (thereby artificial agents) can provide execution traces, which show what statements are being executed as the program runs. Those traces can be analysed by a human expert to understand how an agent, being automated or autonomous, made a given decision. For example, a logic-based system can provide a complete formal proof showing how a given decision will allow a given goal to be reached. While this approach can be useful, such traces are difficult to use for non-expert humans. Thus, it might be preferable to rely on another kind of information, called interpretations. Interpretations are descriptions of an agent's operations "that can be understood by a human, either through introspection or through a produced explanation." Unlike traces, interpretations are tailored to be readable and understandable not only by experts but also by users.*

Regulatory authorities and compliance auditors require yet another distinct transparency mode, one that prioritizes technical verifiability, comprehensive documentation, and systematic evidence enabling independent validation of compliance claims.

Where end-users need comprehensible explanations and business users require operational parameters, regulators demand access to underlying system architecture, training data characteristics, performance metrics disaggregated across relevant demographic categories, validation protocols, and audit trails documenting decision-making processes.

This regulatory transparency must facilitate expert technical assessment rather than broad accessibility, potentially including source code review, algorithmic specifications, and statistical analysis of system outputs, which are materials incomprehensible to average users but essential for regulatory verification that systems operate as claimed and comply with legal obligations.

The EU AI Act's provisions for market surveillance authority access to technical documentation and source code exemplify this regulatory transparency mode, creating confidential disclosure obligations distinct from public-facing transparency requirements.

Effective transparency frameworks must therefore accommodate **multiple simultaneous transparency regimes calibrated to different audiences**: simplified explanations for affected individuals, operational details for commercial users, and comprehensive technical access for regulatory oversight, each serving distinct accountability objectives within a layered transparency architecture.

Therefore, online platforms on which users and business interests combine to support the goal of transparency should ideally provide at least two forms of information: one that is simplified, employing simple vocabulary, providing information intelligible to non-specialists, and including only the details that are needed for the user to assert their individual rights on the platform; and another providing the right amount of technical information to describe the main parameters used for automated decisions, clarify opaqueness regarding competition standards, and provide details on operations of the utmost importance to business users, such as the ranking of results.

Just like accessible and intelligible privacy notices are required by the GDPR, other required transparency standards can be layered with the aim of achieving qualified and personalized transparency: "artificial intelligence and superhuman information processing capabilities could redefine the optimal complexity of legal rules and refine, for example, the content of disclosures to a hitherto unachievable level of granularity."

Such techniques ought to be employed to provide information at different stages of a decision-making process, which also **expands** the transparency journey of algorithmic models beyond just their final results.

3.3. Procedural vs. Material Transparency

Procedural algorithmic transparency encompasses disclosure of governance frameworks, design methodologies, training protocols, and operational parameters guiding algorithmic development and deployment, focusing on processes and structures surrounding algorithmic systems (governance-level processes) rather than their internal mechanics. It documents the "how" and "why" of algorithmic governance through policies, procedures, and decision-making frameworks. In contrast, material algorithmic transparency involves direct access to underlying technical architecture, including source code, algorithmic specifications, training data, and computational models themselves, the actual technical artifacts and implementation details (technical-level details).

The COMPAS risk assessment algorithm illustrates this distinction. Material transparency would entail access to COMPAS' source code, the specific variables it considers, their weighting, and the mathematical formulas calculating recidivism risk scores. In *State v. Loomis*, the defendant sought precisely this access to challenge his sentencing, but Northpointe refused disclosure citing trade secrets, a refusal courts upheld despite COMPAS' use in liberty-depriving decisions.

Procedural transparency, by contrast, would encompass how corrections departments validated the algorithm before adoption, what oversight bodies approved its deployment, whether users received training on its limitations, and what accountability mechanisms existed for erroneous predictions. The COMPAS controversy exemplifies dual transparency failures: material opacity prevented technical scrutiny that might have revealed bias, while procedural opacity enabled deployment without robust validation.

This distinction is (supposed to be) critical in regulatory contexts. Procedural transparency may satisfy accountability requirements through comprehensible explanations of algorithmic purpose and oversight mechanisms, whereas material transparency, though offering deeper technical insight, often presents practical limitations.

Raw algorithmic code is not useful to non-technical stakeholders, including average citizens and regulatory personnel lacking specialized competencies.

The complexity and opacity of source code render such material **effectively meaningless** absent interpretation and contextual explanation, precisely what procedural transparency mechanisms aim to provide.

Contemporary scholarship recognizes that transparency, whether procedural or material, constitutes a mechanism rather than a terminal objective in algorithmic governance. In fact, transparency functions as an instrumental foundation enabling accountability and oversight rather than standalone resolution to governance challenges. This instrumental understanding shapes how frameworks emphasize procedural versus material approaches based on intended audiences and regulatory objectives.

Technological complexity presents the most frequently cited impediment to material transparency implementation. Advanced AI systems' inherent intricacies **create substantial technical barriers to both interpretability and granular disclosure**. However, scholarly opinion diverges regarding whether these limitations constitute legitimate constraints or convenient justifications for avoiding transparency obligations. Commercial considerations, particularly proprietary information and trade secret protection, represent additional barriers affecting both procedural and material transparency, though material disclosure faces more direct resistance given competitive advantages embedded in specific technical implementations.

This divergence explains varying regulatory emphases. Many frameworks explicitly prioritize procedural transparency as more pragmatically effective, citing material transparency's complexity and intellectual property concerns. On one hand, the GDPR's requirement for "meaningful information about the logic involved" exemplifies procedural emphasis, avoiding material disclosure mandates while requiring comprehensible process explanations.

On the other hand, the EU AI Act's provisions for regulatory access to source code represent conditional material transparency, limited to authorized authorities under confidentiality protections (rather than public disclosure).

This hybrid approach acknowledges that procedural transparency better serves public accountability, while material transparency enables expert regulatory oversight: different audiences requiring different transparency modes. A critical distinction emerges between formal transparency compliance and meaningful implementation across both procedural and material approaches.

Mere information availability **does not ensure effective transparency** where information remains incomprehensible or useless by intended recipients.

A "performative" procedural transparency involves disclosure mandates generating compliance exercises without substantive accountability, such as lengthy privacy policies or algorithmic impact assessments that technically satisfy requirements while failing to inform users meaningfully. Similarly, material transparency mandates requiring code repositories or technical documentation may produce voluminous materials that satisfy formal obligations without enabling practical scrutiny by resource-constrained regulators.

G**eopolitical and economic factors significantly influence which type of transparency corporations prioritize across jurisdictions.** Where regulatory capacity for technical audit exists (primarily sophisticated EU authorities), companies may accept material transparency under confidentiality constraints while minimizing public-facing procedural disclosures.

In jurisdictions lacking such capacity, the same companies may offer procedural transparency that cannot be verified against actual system operations. This different approach highlights power asymmetries wherein transparency implementation reflects regulatory stringency rather than uniform principles, with material transparency reserved for jurisdictions capable of utilizing such access and procedural transparency serving as lower-cost compliance elsewhere.

The procedural-material distinction intersects with transparency *versus* explainability frameworks, though these constitute analytically distinct concepts. Transparency encompasses information provision or access; explainability refers to comprehensibility and intelligibility of such information.

Material transparency provides access to technical artifacts, but does not guarantee explainability: source code may be transparent yet unexplainable even to technical experts for complex neural networks. Procedural transparency aims explicitly at explainability, translating technical operations into comprehensible narratives, but may sacrifice technical accuracy for accessibility.

This creates competing scholarly emphases. Those prioritizing external stakeholder empowerment (users, civil society, affected individuals) generally advocate for more robust procedural transparency, enabling contestation and informed consent, accepting that material details may remain inaccessible.

This creates competing scholarly emphases. Those prioritizing external stakeholder empowerment (users, civil society, affected individuals) generally advocate for more robust procedural transparency, enabling contestation and informed consent, accepting that material details may remain inaccessible.

Technical governance scholars emphasizing regulatory oversight prioritize material transparency permitting expert audit, arguing that procedural explanations risk oversimplification or misrepresentation. Both perspectives acknowledge that neither procedural nor material transparency alone constitutes a definitive solution, but rather that effective governance requires calibrating transparency modes to specific accountability objectives and recipient capacities.

It is my contention that **underlying power dynamics present fundamental issues that neither procedural nor material transparency fully addresses**. Information asymmetries between platforms and users persist despite disclosure mandates: *procedural transparency may inform users of profiling practices without enabling meaningful choice alternatives, while material transparency accessible only to regulators does not empower affected individuals directly.*

This suggests the inevitable conclusion that transparency's limitations are a redistributive mechanism: it can illuminate power imbalances without necessarily correcting them. Transparency initiatives thus connect to broader struggles over digital governance structures. Disclosure serves as a necessary but insufficient condition for democratic accountability, requiring complementary mechanisms including regulatory enforcement, competitive market structures, and individual rights frameworks to translate information into actual constraint on algorithmic power.

4. Purposes and Perceived Benefits of Algorithmic Transparency

The following analysis stems from qualitative data of semi-structured interviews conducted with twelve multistakeholder experts purposefully selected from academia, civil society, government, and industry sectors across both Global North and Global South jurisdictions, with particular emphasis on Brazilian and European perspectives. The interview methodology employed **predetermined questions** (see Annex I), **but explored emergent themes**, which facilitated an analysis of *how diverse stakeholders conceptualize and experience algorithmic transparency within their respective institutional and geographic contexts*. The empirical findings reveal that **algorithmic transparency is subject to substantial limitations and systemic challenges that constrain its efficacy as a regulatory mechanism**.

Expert interviews establish universal consensus that algorithmic transparency constitutes an instrumental mechanism rather than a terminal objective serving as a foundational enabler for broader governance goals rather than providing standalone solutions to complex algorithmic challenges. This understanding positions transparency as a necessary component within comprehensive regulatory frameworks that facilitate subsequent accountability measures and oversight mechanisms.

Transparency serves as a procedural prerequisite for assessing accountability within algorithmic processes, functioning as a foundational mechanism that facilitates institutional and civic oversight of automated systems. The absence of transparency opposes liability assignment, but its presence facilitates oversight by public institutions and civic scrutiny through litigation, campaigns, and information requests, thus contributing to democratic oversight models. Additionally, regulatory authorities understand transparency as an indispensable oversight instrument, enabling algorithmic system auditing, control mechanism evaluation, and safeguard feedback provision. Without adequate transparency parameters or auditing mechanisms, regulators must rely upon unverifiable explanations rather than substantive insights into complex algorithmic operations.

However, raw or overly technical transparency provides limited value to average users, who often lack

comprehension of technical language, time for document review, or meaningful choice alternatives. Access to source code proves largely meaningless for understanding algorithmic operations even among technically knowledgeable individuals. This is because transparency benefits derive from meaningful and comprehensible explanations tailored to users' understanding capacity (that is, relational transparency) rather than mere information availability.

There is a wide perception that algorithmic transparency also contributes to broader societal objectives including trust development, democratic participation enhancement, and fairness promotion. Enhanced transparency fosters trust in digital platforms, potentially increasing user engagement, generating feelings of security and comfort, particularly in sensitive data contexts. According to Santos, the democratic implications involve concerns regarding algorithmic influence on electoral processes and power concentration within limited corporate entities, positioning transparency as a gateway for societal discourse regarding desired technological futures and fundamental values.

The Cambridge Analytica scandal exemplifies these democratic concerns. The company's opaque harvesting of Facebook user data and deployment of psychographic profiling algorithms to target voters in the 2016 U.S. presidential election and Brexit referendum operated entirely without public or regulatory visibility. Users remained unaware that their data fed sophisticated behavioral models designed to influence democratic outcomes, while regulators lacked insight into the algorithmic targeting methodologies employed.

This opacity enabled systematic manipulation of democratic processes through personalized political messaging calibrated to individual psychological vulnerabilities, precisely the scenario transparency mechanisms aim to prevent. The scandal's exposure triggered widespread calls for platform accountability and algorithmic transparency in political advertising, demonstrating how opacity facilitates democratic harms while transparency, though not preventing all manipulation, at minimum creates conditions for public scrutiny and informed debate about acceptable technological influence on electoral processes.

Transparency implementation also addresses fairness and bias concerns through connection to anti-discrimination principles and procedural fairness mechanisms.

Even though it is not perceived as a complete solution, transparency serves as a central component for addressing AI biases by enabling improved action on complementary mechanisms. As Rodrigues asserts, the absence of transparency significantly limits researchers investigating algorithmic bias, preventing objective analysis of disparate content delivery or disproportionate moderation affecting specific groups.

This limitation is exemplified by research documenting how Google Search displayed systematic bias toward Indian content over Bangladeshi news when reporting on incidents occurring in Bangladesh. Algorithmic opacity prevented investigators from determining the underlying causes of this bias, whether stemming from training data composition, ranking parameters, geolocation weighting, or other factors, which forced the analysis to rely primarily on observational patterns of disparate outputs rather than direct empirical insight into the system's decision-making logic. This demonstrates precisely how transparency's absence constrains not only affected users' ability to contest biased outcomes but also researchers' capacity to diagnose root causes and propose targeted remediation.

Research facilitation represents an additional transparency benefit, with increased focus providing enhanced information availability through transparency reports, privacy policies, and related documentation. This enables comparative analysis and improved understanding of corporate operations and system functionality, at the same time it potentially mitigates ethical concerns regarding sensitive data handling in academic research contexts.

Finally, transparency initiatives address many power asymmetries within digital environments, connecting to disputes between public and private sector interests. Private sector resistance to transparency, frequently justified through business secrecy claims, serves to avoid external scrutiny while maintaining control over interactions increasingly resembling public spaces. However, transparency alone cannot address fundamental power differentials and information asymmetries between platforms and users, though it may partially address informational imbalances without fundamentally altering underlying power structures.

Search engine algorithmic curation biases illustrate this dynamic. Research has documented systematic bias in search results, including gender stereotyping in image searches, racial bias in autocomplete suggestions, and geopolitical bias in news rankings, yet the

proprietary nature of ranking algorithms prevents comprehensive investigation of root causes or assessment of remedial measures' effectiveness.

The evidence of the interviews conducted demonstrates that algorithmic transparency serves multiple instrumental purposes including accountability facilitation, user empowerment, trust development, democratic participation enhancement, and power dynamic challenges. Nonetheless, effectiveness depends critically upon meaningful information presentation rather than mere availability.

5. Challenges, Limitations, and Criticisms

Algorithmic transparency is a concept that faces significant barriers that limit its impact and make it vulnerable to negative criticism. The main reasons for these problems are the complexity of the computational systems themselves, the commercial interests of private businesses (trade secrets), and the regulatory environment across different parts of the world which is still fragmented and creates inconsistent implementation frameworks.

Highly sophisticated learning AI technologies are the **major contributors** to the opacity of the so-called "black box" models which resist full interpretability. According to Ramiro, some stages of the algorithms are not at all comprehensible and are therefore **quite difficult impenetrable** to democratic oversight and regulatory control requirements. This phenomenon casts transparency in a new light as it becomes an internal administrative problem before the organization of the outside part of the implementation process.

Trade secrets, industrial confidentiality, commercial privacy, and intellectual property rights are the primary justifications commercial entities invoke when resisting comprehensive transparency provisions. Recommendation algorithms typically constitute core components of business models, positioning transparency demands as direct threats to proprietary information protection.

Anonymous interviewees asserted that corporations strategically deploy these justifications to avoid accountability rather than due to genuine insurmountable constraints. However, even where regulators possess legal authority to demand algorithmic disclosure or enable independent technical audits, corporations demonstrate variable compliance based on jurisdictional origin and enforcement capacity.

Regulatory requests from jurisdictions outside the United States and the EU face particularly significant pushback, with companies either refusing

disclosure entirely, providing minimal information that technically satisfies formal requirements while obscuring substantive operations, or imposing procedural delays that undermine timely oversight.

This differential treatment reflects underlying power asymmetries wherein transparency implementation correlates with regulatory market leverage rather than universal legal principles, effectively creating stratified accountability regimes where Global South regulators lack meaningful access to algorithmic systems operating within their jurisdictions despite nominal legal authority to demand such transparency.

Rodrigues and Teffé assert that security considerations present additional transparency constraints, with arguments that excessive algorithmic disclosure could create system vulnerabilities or facilitate adversarial attacks. Content moderation contexts raise concerns that transparency might harm protection and freedom of expression by enabling malicious actors to circumvent safety mechanisms.

These security arguments acknowledge legitimate risks associated with comprehensive algorithmic disclosure while potentially serving as additional justifications for maintaining opacity.

However, opacity in content moderation **can enable catastrophic harms**, as demonstrated by the UN Independent International Fact-Finding Mission on Myanmar's investigation into Facebook's role in perpetuating genocide against the Rohingya. It found that Facebook's algorithms amplified hate speech and disinformation targeting the Rohingya community while the company's opaque content moderation systems failed to adequately detect or remove incitement to violence, despite repeated warnings from civil society organizations.

Meta's refusal to provide transparency regarding algorithmic amplification mechanisms, content moderation training data for Burmese-language content, or enforcement statistics

prevented both external assessment of system adequacy and timely intervention to prevent escalating violence.

It concluded that Facebook "substantively contributed to the level of acrimony and dissension and conflict" and that the platform had become "a useful instrument for those seeking to spread hate." The case demonstrates that, while transparency regarding specific moderation techniques might theoretically enable circumvention, systematic opacity prevents identification of catastrophic moderation failures until after mass atrocities occur. This suggests that concerns about circumvention, though legitimate in certain contexts, cannot justify comprehensive opacity in content moderation systems affecting fundamental rights and human safety.

The inadequacy of content moderation opacity has been repeatedly highlighted also by Meta's own Oversight Board, which has issued systematic calls for enhanced transparency across multiple decisions. In its ruling on the Myanmar military accounts case, the Oversight Board explicitly recommended that Meta "provide more information about how its systems detect and action content and accounts that violate the Dangerous Organizations and Individuals policy," emphasizing that transparency deficits prevent meaningful assessment of whether the platform adequately addresses incitement risks.

Similarly, in cases involving hate speech in India, COVID-19 misinformation, and political speech moderation, the Oversight Board has consistently requested greater transparency regarding algorithmic amplification mechanisms, enforcement error rates, disaggregated removal statistics by region and language, and the criteria used to prioritize content for human review.

The Oversight Board's 2022 quarterly transparency report noted that Meta had implemented some recommendations, but fundamental questions about algorithmic systems remained unanswered, with the company citing technical complexity and competitive concerns as barriers to disclosure.

This pattern demonstrates that even an independent oversight body specifically created by Meta to review moderation decisions encounters opacity that prevents effective accountability, which suggests that transparency deficits reflect deliberate strategic choices rather than insurmountable technical constraints. **If Meta's own oversight mechanism cannot access sufficient information to evaluate system adequacy, external regulators and affected communities face exponentially greater barriers to meaningful scrutiny.**

The accelerated AI development environment prioritizes speed over transparency and safeguards ("move fast and break things", according to Mark Zuckerberg's motto), resulting in inadequate legal enforcement and applicability. Furthermore, legal fragmentation creates substantial complications through diverse and sometimes conflicting jurisdictional approaches, generating risks of interpretive clashes and inconsistent implementation standards. The EU's stringent approach through GPDR and AI Act mandates contrasts with less proceduralized enforcement mechanisms in other jurisdictions, creating disparate transparency obligations.

According to anonymous interviewees, global technology companies demonstrate varying transparency levels based on market relevance and local regulatory authority assertiveness, resulting in double standards that reflect underlying power asymmetries.

Corporate lobbying efforts by major technology companies utilize algorithmic opacity and secrecy to promote non-regulatory discourses, including freedom of expression arguments that resist transparency mandates.

These entities demonstrate significantly greater transparency disposition toward EU markets compared to Global South countries, reflecting asymmetrical power relationships. This dynamic manifests as **digital neocolonialism**, wherein core algorithmic decisions occur in major global centers with

with minimal influence from other regions despite local regulatory frameworks.

Transparency efforts often collide with ongoing debates over the roles and powers of the public and private sectors. Private companies tend to guard their proprietary software and keep business operations secret to prevent external scrutiny, even as they run platforms that act like public forums, shaping social interactions. Regulatory agencies often struggle with analyzing massive, complex data sets or source code, even when they have access to it.

Overly procedural approaches, such as requiring impact assessments and detailed transparency reports, can sometimes serve more to legitimize private sector practices through formal compliance rather than meaningful openness ("regulatory fatigue"). This even creates a risk of regulatory fatigue and leads to reliance on external audits, which might pose conflicts of interest, further reducing public authorities' capacity to handle complex issues effectively.

Many people lack basic digital literacy and do not fully appreciate how transparency adds value to their digital experience. This makes it harder to prioritize transparency as an important political goal. Citizens often focus on surface-level issues and overlook how algorithms influence critical areas like public safety, health, and traffic control.

Then, efforts to promote transparency often clash with other important values, such as protecting privacy, ensuring security, safeguarding intellectual property, and encouraging innovation.

Some interviewees contended that there is a prevalent perception that strict transparency requirements can slow down innovation and increase costs for companies, though there are still questions about whether these costs are too high for technologies that pose important risks.

Overall, the challenges to algorithmic transparency are **complex and multi-faceted**. They include technical difficulties, resistance from businesses, fragmented regulations, power imbalances, and conflicting values. In order to render transparency more effective, there needs to be comprehensive strategies that address these widespread issues.

7. Conclusion

This analysis exposes algorithmic transparency as a regulatory performance that may perpetuate rather than actually challenges asymmetrical structures of technology governance. Even though European instruments such as the GDPR, AI Act, DSA and DMA establish comprehensive versions of algorithmic transparency requirements, their global application remains uneven, creating distinct experiences for users across different markets.

European regulations demonstrate significant **extraterritorial reach**. The Brussels Effect enables these transparency requirements to proliferate across borders as multinational platforms implement EU-compliant systems globally to avoid maintaining parallel infrastructures. However, even as European transparency standards diffuse internationally, they r**emain subject to abstraction, fragmentation, and domestic reformulation**.

Non-EU jurisdictions adopt superficially similar transparency frameworks that nonetheless differ substantially in enforcement rigor, technical specifications, and interpretive approaches, creating regulatory convergence in form but divergence in substance. The conceptual distinction between procedural and material transparency, though analytically useful, reflects deeper challenges in operationalizing accountability mechanisms across diverse technological and institutional contexts. European frameworks emphasize detailed procedural requirements and documentation obligations, while other jurisdictions still lack equivalent enforcement mechanisms or technical capacity for meaningful oversight.

From a Global South perspective, this investigation highlights how Big Tech companies implement differential transparency standards based on regulatory stringency and market considerations. European users tend to benefit from several disclosure requirements and enforcement mechanisms, while users in emerging (and developing) markets frequently encounter the same algorithmic systems with substantially fewer transparency protections. This pattern reflects not just technical constraints, but strategic corporate compliance decisions that privilege markets with stronger regulatory frameworks.

Overall, this analysis demonstrates that algorithmic transparency, while essential for democratic governance, operates within structural constraints that limit its effectiveness in addressing global power imbalances in technology governance. The research reveals a critical disconnect between transparency's democratic promise and its practical implementation, particularly evident in the differential treatment of regulatory jurisdictions based on market significance rather than legal principle.

Algorithmic transparency could advance if there were **effective institutional capacity and enforcement mechanisms**. However, the prevailing model of procedural compliance over substantive oversight perpetuates informational asymmetries between multinational technology corporations and regulators, especially Global South regulators.

This dynamic constitutes a form of regulatory stratification through which transparency protections correlate with economic leverage rather than legal entitlement. The technical limitations narrative frequently invoked to justify algorithmic opacity reveals corporate resistance to disclosure, with compliance standards frequently adjusting to regulatory appetite and market importance.

This **selective implementation** undermines claims of universal technical constraints and exposes the **fundamentally political nature** of transparency decisions.

Current frameworks suffer from three interconnected deficiencies. First, they individualize responsibility for algorithmic harms while obscuring systemic governance failures. Transparency requirements predominantly operate through individual rights frameworks (data subject access requests under GDPR, individual explanations for automated decisions, personal consent mechanisms) that position affected individuals as primary enforcement agents.

This structure places the burden of identifying harms, requesting information, interpreting disclosures, and pursuing remedies on individual users, who typically lack the technical expertise, financial resources, or sustained attention necessary for effective accountability.

A data subject exercising their right to explanation for an algorithmic credit decision receives information about their specific case, but this individualized transparency

provides no visibility into whether the underlying model systematically discriminates against protected groups, which is a pattern only detectable through aggregate analysis. This individualization obscures systemic governance failures that enable algorithmic harms at scale.

Second, they prioritize formal disclosure over meaningful accountability mechanisms. Transparency frameworks predominantly mandate information provision (privacy policies, terms of service disclosures, transparency reports, algorithmic impact assessments) without ensuring such disclosures enable substantive oversight or behavioral change.

Platforms satisfy formal compliance by publishing voluminous documentation that meets technical legal requirements while remaining functionally inaccessible to those it ostensibly informs. Meta's transparency reports exemplify this pattern: they provide aggregate statistics on content removals, government requests, and policy enforcement that appear comprehensive while obscuring critical information about algorithmic amplification, error rates for automated moderation, or demographic disparities in enforcement. The reports generate the appearance of accountability without enabling external verification of platform claims or meaningful assessment of whether disclosed practices comply with substantive legal obligations.

Third, they operate within existing power structures rather than challenging the concentration of technological decision-making authority. Current transparency frameworks implicitly accept that algorithmic systems will continue to be developed, deployed, and controlled by a small number of multinational corporations, with transparency serving merely to make these processes somewhat more visible rather than democratizing technological governance itself.

Regulatory interventions seek to constrain corporate discretion at the margins through disclosure mandates and procedural requirements, but do not question whether concentrated private control over critical digital infrastructure serves public interest or whether alternative governance models might better align algorithmic systems with democratic value.

Moving beyond these limitations requires structural reforms that address underlying capacity asymmetries. Effective algorithmic governance ought to integrate transparency with technological sovereignty initiatives, including mandatory algorithmic impact

assessments, real-time auditing capabilities, and enhanced regulatory expertise development. International cooperation frameworks should prioritize capacity building over compliance harmonization, enabling substantive rather than performative oversight.

Transparency without power redistribution merely legitimizes existing arrangements. Meaningful regulatory reform requires developing domestic technical capacity, establishing regional regulatory cooperation agreements, and creating enforcement mechanisms proportionate to algorithmic systems' societal impact. Ultimately, algorithmic transparency's democratic potential can only be realized through integration within broader frameworks of technological justice.

Bibliography

Legal Instruments and Official Documents

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Council of Europe Treaty Series No. 225, 2024.

IEEE Standard for Transparency of Autonomous Systems (IEEE Std 7001-2021). IEEE Standards Association, 2021.

OECD. Common Guideposts to Promote Interoperability in AI Risk Management. OECD Artificial Intelligence Papers No. 5, November 2023.

OECD. AI, Data Governance, and Privacy: Synergies and Areas of International Co-operation. OECD Artificial Intelligence Papers, No. 22, June 2024.

OECD. OECD Recommendation on Artificial Intelligence. OECD Legal Instrument No. 0449, 2024.

OECD. Explanatory Memorandum on the Updated OECD Definition of an AI System. OECD Artificial Intelligence Papers No. 8, March 2024.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L 119, 4 May 2016.

Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services. Official Journal of the European Union, L 186, 11 July 2019.

Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending

Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act). Official Journal of the European Union, L 265, 12 October 2022.

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act). Official Journal of the European Union, L 277, 27 October 2022.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. Official Journal of the European Union, L, 12 July 2024.

UNESCO. Recommendation on the Ethics of Artificial Intelligence. UNESCO, 2021.
UN Human Rights Council. Report of the Independent International Fact-Finding Mission on Myanmar. UN Doc. A/HRC/39/64, September 12, 2018.

Books and Monographs

Anjos, Lucas Costa dos. Can Law Ever Be Code? Beyond Google's Algorithmic Black Box and Towards a Right to Explanation. Doctoral Thesis, 2021. Available at: https://dipot.ulb.ac.be/dspace/bitstream/2013/334146/3/DoctoralThesisLucasAnjos.pdf.

Couldry, Nick, and Ulises A. Mejias. The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford: Stanford University Press, 2019.

Ellickson, Robert C. Order Without Law: How Neighbors Settle Disputes. Cambridge, MA: Harvard University Press, 1991.

Ezrachi, Ariel, and Maurice E. Stucke. Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy. Cambridge, MA: Harvard University Press, 2016.

Gillespie, Tarleton. Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. New Haven: Yale University Press, 2018.

Graef, Inge. EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility. Alphen aan den Rijn: Kluwer Law International, 2016.

Kingdon, John W. Agendas, Alternatives, and Public Policies. 2nd ed. New York: Longman, 2003.

Kurbalija, Jovan & Valentin Katrandjiev. Multistakeholder Diplomacy: Challenges and Opportunities. Malta Geneva: DiploFoundation, 2006.

Noble, Safiya Umoja. Algorithms of Oppression: How Search Engines Reinforce Racism. New York: NYU Press, 2018.

O'Neil, Cathy. Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy. New York London: Crown Allen Lane, 2016.

Pasquale, Frank. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press, 2015.

Patry, William. How to Fix Copyright. Oxford: Oxford University Press, 2011.

Pistor, Katharina. The Code of Capital: How the Law Creates Wealth and Inequality. Princeton: Princeton University Press, 2019.

Roberts, Sarah T. Behind the Screen: Content Moderation in the Shadows of Social Media. New Haven: Yale University Press, 2019.

Tusikov, Natasha. Chokepoints: Global Private Regulation on the Internet. Oakland: University of California Press, 2016.

Journal Articles

Ananny, Mike, and Kate Crawford. ‘‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability.’’ New Media & Society 20, no. 3 (2018): 973–89. doi:10.1177/1461444816676645.

Bamberger, Kenneth A., and Deirdre K. Mulligan. "Privacy on the Books and on the Ground."

Stanford Law Review 63, no. 2 (2011): 247-316.

Bauer, Kevin and Andrej Gill. "Mirror, Mirror on the Wall: Algorithmic Assessments, Transparency, and Self-Fulfilling Prophecies." Information Systems Research 35, no. 1 (2024): 226–48. doi:10.1287/isre.2023.1217.

Bayamlıoğlu, Emre. "The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the So-called 'Right to Explanation'." Regulation & Governance (March 14, 2021): 1-23. https://doi.org/10.1111/rego.12391.

Bradford, Anu. "The Brussels Effect." Northwestern University Law Review 107, no. 1 (2012): 1-68.

Brkan, Maja and Grégory Bonnet. "Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Discretionary Spaces." European Journal of Risk Regulation 11, no. 1 (2020): 18-50 https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/legal-and-technical-feasibility-of-the-gdprs-quest-for-explanation-of-algorithmic-decisions-of-black-boxes-white-boxes-and-fata-morganas/7324CDE80A300179C170C5BA8CA7E851.

Burrell, Jenna, and Marion Fourcade. "The Society of Algorithms." Annual Review of Sociology 47 (2021): 213-237. doi:10.1146/annurev-soc-090820-020800.

Busch, Christoph. "Implementing Personalized Law: Personalized Disclosures in Consumer Law and Data Privacy Law." The University of Chicago Law Review 86, no. 2 (March 2019): 309-332. https://www.jstor.org/stable/10.2307/26590557.

Calo, Ryan. "Robotics and the Lessons of Cyberlaw." California Law Review 103, no. 3 (2015): 513-563.

Bayamlıoğlu, Emre. "The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the So-called 'Right to Explanation'." Regulation & Governance (March 14, 2021): 1-23. https://doi.org/10.1111/rego.12391. Law Journal 41, no. 1 (2018): 120-14

Chapman, Elizabeth N., and I. Glenn Cohen. "Transparency and Explainability in Clinical AI." American Journal of Bioethics 25, no. 2 (2025): 4-27. doi:10.1080/15265161.2025.2458425.

Coglianese, Cary, and David Lazer. "Management-Based Regulation: Prescribing Private Management to Achieve Public Goals." Law & Society Review 37, no. 4 (2003): 691-730.

Chau, M., M.G Rahman, and T. Debnath. "From Black Box to Clarity: Strategies for Effective AI Informed Consent in Healthcare." Artificial Intelligence in Medicine 167 (2025). doi:10.1016/j.artmed.2025.103169.

Cranor, Lorrie Faith. "Necessary But Not Sufficient: Standardized Mechanisms for Privacy Notice and Choice." Journal of Telecommunications and High Technology Law 10, no. 2 (2012): 273-307.

De Hert, Paul, and Vagelis Papakonstantinou. "The New General Data Protection Regulation: Still a Sound System for the Protection of Individuals?" Computer Law & Security Review 32, no. 2 (2016): 179-194.

Dignum, Virginia. "Ethics in Artificial Intelligence: Introduction to the Special Issue." Ethics and Information Technology 20, no. 1 (March 2018): 1-3. https://doi.org/10.1007/s10676-018-9450-z.

Ding, Weiping, Mohamed Abdel-Basset, Hossam Hawash, and Ahmed M Ali. "Explainability of Artificial Intelligence Methods, Applications and Challenges: A Comprehensive Survey." Information Sciences 615 (2022): 238–92. doi:10.1016/j.ins.2022.10.013.

Edwards, Lilian, and Michael Veale. "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For." Duke Law & Technology Review 16, no. 1 (2017): 18-84.

Elkin-Koren, Niva. "Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence." Big Data & Society 7, no. 2 (2020). doi:10.1177/2053951720932296.

Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. “Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns.” Big Data & Society 6, no. 1 (2019). doi:10.1177/2053951719860542.

Fenster, Mark. "Transparency in Search of a Theory." European Journal of Social Theory 18, no. 2 (2015): 150-167.

Fuller, Lon L. "The Forms and Limits of Adjudication." Harvard Law Review 92, no. 2 (1978): 353-409.

Gillespie, Tarleton. “The Relevance of Algorithms.” In Media Technologies. The MIT Press, 2014. doi:10.7551/mitpress/9780262525374.003.0009.

Goldberg, John C. P., and Benjamin C. Zipursky. "Torts as Wrongs." Texas Law Review 88, no. 5 (2010): 917-986.

Graef, Inge. "Differentiated Treatment in Platform-to-Business Relations: EU Competition Law and Economic Dependence." Yearbook of European Law 38 (2019): 448-499. doi:10.1093/yel/yez004.

Grimmelmann, James. "Regulation by Software." Yale Law Journal 114, no. 7 (2005): 1719-1758.

Gstrein, Oskar J., and Andrej Zwitter. "Extraterritorial Application of the GDPR: Promoting European Values or Power?" Internet Policy Review 10, no. 3 (2021). doi:10.14763/2021.3.1570.

Helberger, Natali. "On the Democratic Role of News Recommenders." Digital Journalism 7, no. 8 (2019): 993-1012. doi:10.1080/21670811.2019.1623700.

Kaminski, Margot E. "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability." Southern California Law Review 92, no. 6 (2019): 1529-1616.

Kaminski, Margot E., and Jennifer M. Urban. "The Right to Contest AI." Columbia Law Review 121, no. 7 (2021): 1957-2048.

Kerber, Wolfgang, and Heike Schweitzer. "Interoperability in the Digital Economy." Journal of Intellectual Property, Information Technology and E-Commerce Law 8, no. 1 (2017): 39-58.

Kesselheim, Aaron S., and Jerry Avorn. "Using Patent Law to Improve Access to Generic Drugs in the United States." Journal of Law, Medicine & Ethics 37, no. 2 (2009): 192-205.

Kharitonova, Yu. S. "Legal Means of Providing the Principle of Transparency of Artificial Intelligence." Journal of Digital Technologies and Law 1, no. 2 (2023): 337–58. doi:10.21202/jdtl.2023.14.

Kwet, Michael. "Digital Colonialism: US Empire and the New Imperialism in the Global South." Race & Class 60, no. 4 (2019): 3-26. doi:10.1177/0306396818823172.

Lehmann, Cedric A., Christiane B Haubitz, Andreas Fügener, and Ulrich W Thonemann. "The Risk of Algorithm Transparency: How Algorithm Complexity Drives the Effects on the Use of Advice." Production and Operations Management 31, no. 9 (2022): 3419–34. doi:10.1111/poms.13770.

Lipton, Zachary C. "The Mythos of Model Interpretability." Queue 16, no. 3 (2018): 31-57. doi:10.1145/3236386.3241340.

Lynskey, Orla. "Regulating 'Platform Power.'" LSE Law, Society and Economy Working Papers 1/2017.

Mahler, Tobias, and Michael Kluth. "Building Trust in Data Sharing: Key Elements of Trustworthy Data Spaces." European Data Protection Law Review 8, no. 4 (2022): 524-540.

Malgieri, Gianclaudio, and Giovanni Comandé. "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation."

International Data Privacy Law 7, no. 4 (2017): 243-265.

Malgieri, Gianclaudio. "Automated Decision-Making in the EU Member States: The Right to Explanation and Other 'Suitable Safeguards' in the National Legislations." Computer Law & Security Review 35, no. 5 (October 2019): 1-28. https://doi.org/10.1016/j.clsr.2019.05.002.

Marda, Vidushi, and Shivangi Narayan. "Data in New Delhi's Predictive Policing System." In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021): 317-324.

Martin, Kirsten. "Ethical Implications and Accountability of Algorithms." Journal of Business Ethics 160, no. 4 (December 2019): 835-850. https://doi.org/10.1007/s10551-018-3921-3.

McDonald, Aleecia M., and Lorrie Faith Cranor. "The Cost of Reading Privacy Policies." I/S: A Journal of Law and Policy for the Information Society 4, no. 3 (2008): 543-568.
Mejias, Ulises A., and Nick Couldry. "Datafication." Internet Policy Review 8, no. 4 (2019). doi:10.14763/2019.4.1428.

Morano-Foadi, Sonia and Stelios Andreadakis. "Reflections on the Architecture of the EU after the Treaty of Lisbon: The European Judicial Approach to Fundamental Rights." European Law Journal 17, no. 5 (September 2011): 595-610. https://doi.org/10.1111/j.1468-0386.2011.00568.x.

Nieuwenhuizen, Esther. "Algorithm Registers: A Box-Ticking Exercise or Meaningful Tool for Transparency?" Information Polity 29, no. 4 (2024): 415–33. doi:10.1177/15701255241297107.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." Science 366, no. 6464 (2019): 447-453. doi:10.1126/science.aax2342.

Omri Ben-Shahar and Carl E. Schneider, "The Failure of Mandated Disclosure," University of Pennsylvania Law Review 159, no. 3 (2011): 647-749. https://www.jstor.org/stable/41149884.

Powell, Alison B. "Explanations as Governance? Investigating Practices of Explanation in Algorithmic System Design." European Journal of Communication 36, no. 4 (August 2021): 355-374. https://doi.org/10.1177/02673231211028376.

Pozen, David E. "Transparency's Ideological Drift." Yale Law Journal 128, no. 1 (2018): 100-165.

Rahman, K. Sabeel. "The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept." Cardozo Law Review 39, no. 5 (2018): 1621-1690.

Reed, Chris. "Taking Sides on Technology Neutrality." SCRIPTed 4, no. 3 (2007): 263-284.

Samuelson, Pamela. "The Copyright Principles Project: Directions for Reform." Berkeley Technology Law Journal 25, no. 3 (2010): 1175-1245.

Schmiegelow, Mateus Maia Ramos, Sergio Luiz Stevan Jr., Hugo Valadares Siqueira, and Paulo J. L. Adeodato. "Explainable Machine Learning for Crime Prediction: A Systematic Review." Applied Soft Computing 170 (2025): 112656. https://doi.org/10.1016/j.asoc.2025.112656.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." In Proceedings of the Conference on Fairness, Accountability, and Transparency (2019): 59-68.

Selbst, Andrew D. "An Institutional View of Algorithmic Impact Assessments." Harvard Journal of Law & Technology 35, no. 1 (2021): 117-186.

Shaffer, Gregory. "Globalization and Social Protection: The Impact of EU and International Rules in the Ratcheting Up of U.S. Privacy Standards." Yale Journal of International Law 25, no. 1 (2000): 1-88.

Solove, Daniel J. "Privacy Self-Management and the Consent Dilemma." Harvard Law Review 126, no. 7 (2013): 1880-1903.

Sweeney, Latanya. "Discrimination in Online Ad Delivery." Communications of the ACM 56, no. 5 (2013): 44-54.

Trites, Allison. "Black Box Ethics: How Algorithmic Decision-Making Is Changing How We View Society and People: Advocating for the Right for Explanation and the Right to Be Forgotten in Canada." Global Media Journal: Canadian Edition 11, no. 2 (2019): 18-30. http://gmj-canadianedition.ca/wp-content/uploads/2020/06/03_Trites-Volume-11-issue-2-Final.pdf

Ursic, Helena. "The Failure of Control Rights in the Big Data Era: Does a Holistic Approach Offer a Solution?" In Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?, edited by Mor Bakhoum et al., 73-96. MPI Studies on Intellectual Property and Competition Law, vol. 28. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018. https://doi.org/10.1007/978-3-662-57646-5.

Veale, Michael, Reuben Binns, and Lilian Edwards. "Algorithms That Remember: Model Inversion Attacks and Data Protection Law." Philosophical Transactions of the Royal Society A 376, no. 2133 (2018). doi:10.1098/rsta.2018.0083.

Voss, W. Gregory. "Cross-Border Data Flows, the GDPR, and Data Governance." Pacific Rim Law & Policy Journal 29, no. 3 (2020): 485–531.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." International Data Privacy Law 7, no. 2 (2017): 76-99.

Wachter, Sandra, Brent Mittelstadt, and Chris Russell. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR." Harvard Journal of Law & Technology 31, no. 2 (2018): 841-887.

Watson, David S. and Luciano Floridi. "The Explanation Game: A Formal Framework for Interpretable Machine Learning." Synthese 198 (2021): 9211-9242.

Yang, Qiang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated Machine Learning: Concept and Applications." ACM Transactions on Intelligent Systems and Technology 10, no. 2 (2019): 1-19. doi:10.1145/3298981.

Zipursky, Benjamin C. "Reasonableness In and Out of Negligence Law." University of Pennsylvania Law Review 163, no. 7 (2015): 2131-2176.

Zuiderveen Borgesius, Frederik. "Informed Consent: We Can Do Better to Defend Privacy." IEEE Security & Privacy 13, no. 2 (2015): 103-107.

Book Chapters

Anjos, Lucas Costa dos. "Rethinking Algorithmic Explainability Through the Lenses of Intellectual Property and Competition." In Digital Governance: Confronting the Challenges Posed by Artificial Intelligence. T.M.C. Asser Press The Hague, 2024 https://doi.org/10.1007/978-94-6265-639-0.

Ferreira, Juliana J. and Mateus S. Monteiro. "What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice." In Design, User Experience, and Usability: Design for Contemporary Interactive Environments, edited by Aaron Marcus and Elizabeth Rosenzweig, 56-73. Cham: Springer, 2020.

Hood, Christopher. "Transparency in Historical Perspective." In Transparency: The Key to Better Governance?, edited by Christopher Hood and David Heald, 3-23. Oxford: Oxford University Press, 2006.

Koops, Bert-Jaap. "Should ICT Regulation Be Technology-Neutral?" In Starting Points for ICT Regulation, edited by Bert-Jaap Koops, Miriam Lips, Corien Prins, and Maurice Schellekens, 77-108. The Hague: TMC Asser Press, 2006.

La Diega, Guido Noto. "Data as Digital Assets: The Case of Targeted Advertising." In Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?, edited by Mor Bakhoum et al., 489-520. MPI Studies on Intellectual Property and Competition Law, vol. 28. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018. https://doi.org/10.1007/978-3-662-57646-5.

Lughofer, Edwin. "Model Explanation and Interpretation Concepts for Stimulating Advanced Human-Machine Interaction with 'Expert-in-the-Loop'." In Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent, edited by Jianlong Zhou and Fang Chen, 177-221. Human–Computer Interaction Series. Cham: Springer, 2018.

Senden, Linda, and Sacha Prechal. "Differentiation in and through EU Law: Assessing Flexibility and Diversity in the European Union." In Research Handbook on EU Institutional Law, edited by Anthony Arnull and Damian Chalmers, 157-183. Cheltenham: Edward Elgar, 2016.

Zetzsche, Dirk A., Ross P. Buckley, Douglas W. Arner, and Janos N. Barberis. "The Automated Financial Advisor: Opportunities and Challenges of Robo-Advisory in Wealth Management." In Disruptive Technology in Banking and Finance: An International Perspective on FinTech, edited by Lynn Stout, Sergio Gramitto, and Kathryn Judge. Palgrave Macmillan, 2019.

Zheng, Robert and Kevin Greenberg. "Effective Design in Human and Machine Learning: A Cognitive Perspective." In Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent, edited by Jianlong Zhou and Fang Chen, 55-74. Human–Computer Interaction Series. Cham: Springer, 2018.

News Articles, Working Papers and Reports

Article 29 Data Protection Working Party. Guidelines on Transparency under Regulation 2016/679. WP260 rev.01, April 11, 2018.

Australasian Institute of Judicial Administration. Artificial Intelligence and Judicial Decision-Making: Report of the AIJA AI and Judicial Decision-Making Committee. Melbourne: AIJA, December 2023. https://aija.org.au/wp-content/uploads/2023/12/AIJA_AI-DecisionMakingReport_2023update.pdf.
Australian Competition and Consumer Commission. Digital Platforms Inquiry: Final Report. Canberra: ACCC, June 2019.

Bertuzzi, Luca. "Sectors Duel over Required Details in EU's AI Standardization Process." MLex, 28 October 2024. https://lnkd.in/eu2CrFBr.

Cadwalladr, Carole, and Emma Graham-Harrison. "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach." The Guardian, March 17, 2018.

Chari, Shruthi et al. "Directions for Explainable Knowledge-Enabled Systems." arXiv preprint, March 17, 2020. https://arxiv.org/pdf/2003.07523.pdf.
Commonwealth of Australia. Royal Commission into the Robodebt Scheme. Report. Canberra: Commonwealth of Australia, 2023.

Commonwealth of Australia. Royal Commission into the Robodebt Scheme. Report. Canberra: Commonwealth of Australia, 2023.

"EU Picks Experts to Steer AI Compliance Rules." Reuters, 30 September 2024. https://www.reuters.com/technology/artificial-intelligence/eu-picks-experts-steer-ai-compliance-rules-2024-09-30/.

European Commission. "Harmonised Standards for the European AI Act." AI Watch, 25 October 2024. https://ai-watch.ec.europa.eu/news/harmonised-standards-european-ai-act-2024-10-25_en.

Garrido, Josep Soler et al. Harmonised Standards for the European AI Act. Joint Research Centre (JRC), European Commission, 2024.

Geurkink, Brandi, Paddy Leerssen, Javier Sánchez-Monedero, and Luca Bellstam. "Assessing the Impact of the Discontinuation of CrowdTangle on Social Media Research." OSF Preprints, August 2024. doi: 10.31219/osf.io/npr73.

Helberger, Natali, Paddy Leerssen, Jef Ausloos, Platon Tervezis, Max van Drunen, and Catalina Goanta. "EU Election 2024: Five Recommendations on Transparency for Very Large Online Platforms." Internet Policy Review, February 26, 2024.

Haugen, Frances. Testimony before the U.S. Senate Committee on Commerce, Science, and Transportation, Subcommittee on Consumer Protection, Product Safety, and Data Security, October 5, 2021.

Information Commissioner's Office (UK). Investigation into the Use of Data Analytics in Political Campaigns: Final Report. London: ICO, November 2020.

Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Mahieu, René L. P. and Jef Ausloos. "Harnessing the Collective Potential of GDPR Access Rights: Towards an Ecology of Transparency." Internet Policy Review, July 6,

2020. https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487.

Mozur, Paul. "A Genocide Incited on Facebook, With Posts From Myanmar's Military." New York Times, October 15, 2018.

Oversight Board. Case Decision 2021-001-FB-FBR, "Shared Content in India." January 28, 2021. https://oversightboard.com/decision/FB-I4B7NLUX/.

Oversight Board. Case Decision 2021-003-FB-UA, "Alleged Spreading of COVID-19 Misinformation in Brazil." April 8, 2021. https://oversightboard.com/decision/FB-691QAMHJ/.

Oversight Board. Case Decision 2021-005-FB-UA, "Removal of Page Praising Violent Actors in Myanmar." September 29, 2021. https://oversightboard.com/decision/FB-XAK9ESPT/.

Oversight Board. Quarterly Transparency Report (Q2 2022). https://oversightboard.com/quarterly-transparency-reports/.

Shahzeb Mahmood and Sabhanaz Rashid Diya, "Whose Stories Count? How Google Search Erases Local Media in Bangladesh," Tech Global Institute. Accessed January 18, 2026, https://techglobalinstitute.com/research/whose-stories-count-how-google-search-erases-local-media-in-bangladesh/.

Von Grafenstein, Max, Peggy Valcke, and Natali Helberger. "Data Access and Research Under the Digital Services Act: An Implementation Roadmap." Weizenbaum Institute, Berlin, 2023.

UN Special Rapporteur on Freedom of Opinion and Expression, Organization for Security and Co-operation in Europe Representative on Freedom of the Media, Organization of American States Special Rapporteur on Freedom of Expression, and African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access to Information. Joint Declaration on Freedom of Expression and Elections in the Digital Age. FOM.GAL/3/17, March 2017.

Interviews

Antoine, Elise. Interview by author. London School of Economics, October 31, 2024.

Alves, Marco Antônio. Interview by author. Online via Zoom, February 27, 2025.

Mansell, Robin. Interview by author. London School of Economics, October 29, 2024.

Powell, Alison. Interview by author. London School of Economics, October 29, 2024.

Ramiro, André. Written interview by author. Online, May 14, 2025.

Rodrigues, Fernanda. Interview by author. Online via Zoom, February 25, 2025.

Rouvroy, Antoinette. "Algorithmic Governmentality and the Death of Politics." Interview by Green European Journal, March 27, 2020. https://www.greeneuropeanjournal.eu/algorithmic-governmentality-and-the-death-of-politics/.

Santos, Pedro Henrique. Interview by author. Online via Zoom, February 27, 2025.

Teffé, Chiara de. Written interview by author. Online, March 2, 2025.

Legal Cases

Case C-507/17, Google LLC v. Commission nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:772 (Sept. 24, 2019).

State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

Web Sources

Council of Europe. "Council of Europe Opens First-Ever Global Treaty on AI for Signature." Accessed October 27, 2024. https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature.

Appendix I

Interview Questions

Semi-structured interviews were conducted following the questionnaire below, of around 30 minutes each. Interviewees were first asked about confidentiality and anonymity preferences, in addition to their personal identification in terms of multistakeholder categorization (civil society, academia, government, industry).

Interpretations of transparency How do you define "algorithmic transparency," and how does your field approach or interpret this concept?
Transparency as a solution In what ways do you think transparency addresses concerns related to accountability and oversight? Are there aspects where transparency may fall short?
Transparency across jurisdictions How do you see the approach to transparency varying between jurisdictions, such as the EU and the UK? What impacts do these differences have on global tech industries?
Technological limitations What challenges do you identify in implementing transparency in complex systems, both technologically and practically?
Research and practical implications Has the focus on transparency influenced research in technology in your field? Has it brought about significant shifts in practice or innovation within the industry?
Incentives and deterrents What incentives do you see as most effective for organizations to adopt transparent practices in AI systems? Conversely, what barriers might prevent them from doing so?
Transparency vs. other values How does the emphasis on transparency intersect or conflict with other regulatory values, such as privacy, fairness, and security? How can these tensions be balanced in practice?
Future of transparency in regulation Looking forward, how do you envision the role of transparency evolving in tech regulation?
Impact on stakeholders How do you think different stakeholders—such as regulators, developers, and the public—perceive the value of transparency?
Practical examples and case studies Could you share any examples where transparency in algorithmic processes has led to measurable improvements (or failures) in oversight, accountability, or trust?

Author

Lucas Costa dos Anjos *is a 2025 Tech Policy Fellow at TGI and is currently a postdoctoral researcher at Sciences Po Law School in Paris. He holds a PhD in Law from Université libre de Bruxelles, and is also a Professor of Law at UFJF and a founder of the Institute for Research on Internet and Society (IRIS).*

www.techglobalinstitute.com