DUMLUPINAR ÜNİVERSİTESİ SOSYAL BİLİMLER DERGİSİ
DUMLUPINAR UNIVERSITY JOURNAL OF SOCIAL SCIENCES
E-ISSN: 2587-005X https://dergipark.org.tr/tr/pub/dpusbe
Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 89, 343-366; 2026.
DOI: 10.51290/dpusbe.1879148

Araştırma Makalesi / Research Article

EXPLAINABILITY BURDEN AND ACCOUNTABILITY OF ORGANIZATIONAL AI DECISIONS: A BLOCKCHAIN BASED GOVERNANCE MODEL

Arif YILDIRIM[1]

Abstract
The speed with which artificial intelligence has proliferated in organizational decision-making has intensified the accountability crisis. In finance, healthcare, and human resources, AI systems increasingly influence high-risk decision outcomes, with stakeholders demanding transparency concerning how decisions are made and accountability for failures. Yet prevailing understandings of explainability are limited by siloed responsibility structures, weak audit trails, and an under-specified "explainability burden"—the labor associated with the production, maintaining, verifying, and assuring of explanations for decisions made by AI. This conceptual article builds a framework for distributing that burden among key stakeholders (AI developers, data providers, process owners, auditors) using blockchain as an immutable governance infrastructure. Building on research in algorithmic accountability, institutional theory and distributed governance, the article proposes a framework for quantifying explanation expectations as a function of risk exposure and capacity and implementing these allocations through smart contracts as smart contracts on blockchain platforms. The model regards explainability as a measurable burden that must be strategically allocated to ensure transparency and legitimacy. I derive five testable propositions linking blockchain-based auditability to accountability outcomes, stakeholder trust, regulatory compliance and organizational learning, and provide insights for organizations dealing with the EU AI Act, GDPR Article 22 and upcoming AI governance regimes.
Keywords: Explainable AI, Blockchain, Algorithmic accountability, Distributed governance, Responsibility allocation, Institutional theory
JEL Codes: C61, O33, L86

ÖRGÜTSEL YAPAY ZEKÂ KARARLARININ AÇIKLANABİLİRLİK YÜKÜ VE HESAP VEREBİLİRLİĞİ: BLOK ZİNCİRİ TABANLI BİR YÖNETİŞİM MODELİ

Öz
Kurumsal karar alma alanında yapay zekânın (YZ) hızla yaygınlaşması, hesap verebilirlik krizini keskinleştirmiştir. Finans, sağlık ve insan kaynaklarında YZ sistemleri daha yüksek riskli karar çıktıları üzerinde etkili olurken, paydaşlar kararların nasıl verildiğine dair şeffaflık ve hatalar karşısında daha güçlü hesap verebilirlik talep etmektedir. Ancak açıklanabilirliğe ilişkin yaklaşımlar; silo temelli sorumluluk yapıları, zayıf denetim izleri ve yeterince tanımlanmamış bir "açıklanabilirlik yükü" nedeniyle sınırlı kalmaktadır. Bu yük, YZ kararlarına ilişkin açıklamaları üretme, sürdürme, doğrulama ve güvence altına alma emeğini ifade eder. Bu kavramsal-kuramsal makale, blok zincirini değiştirilemez bir yönetişim altyapısı olarak ele alarak bu yükün paydaşlar (geliştiriciler, veri sağlayıcılar, süreç sahipleri ve denetçiler) arasında nasıl dağıtılabileceğine ilişkin bir çerçeve kurar. Algoritmik hesap verebilirlik, kurumsal kuram ve dağıtık yönetişim literatürüne dayanarak; açıklama beklentilerini risk maruziyeti ve kapasiteye göre nicelleştiren ve tahsisleri akıllı sözleşmeler aracılığıyla uygulanabilir kılan biçimsel bir model önerilmektedir. Model, açıklanabilirlik faaliyetini ölçülebilir bir yük olarak görür ve şeffaflık ile meşruiyeti güvence altına almak için tahsis edilmesi gerektiğini savunur. Blok zinciriyle sağlanan denetlenebilirlik, paydaş güveni, düzenleyici uyum ve örgütsel öğrenme arasındaki ilişkiye dair test edilebilir beş önerme geliştirir; ayrıca AB Yapay Zekâ Tüzüğü (EU AI Act), GDPR Madde 22 ve yaklaşan YZ yönetişimi rejimleri için çıkarımlar sunar ve uygulamaya dönük yön vermektedir.
Anahtar Kelimeler: Açıklanabilir yapay zeka, Blok zincir, Algoritmik hesap verebilirlik, Dağıtık yönetişim, Sorumluluk tahsisi, Kurumsal teori
JEL Kodları: C61, O33, L86

[1] Assoc. Prof . Dr., Çanakkale Onsekiz Mart University, arify@comu.edu.tr, ORCID: 0000-0002-4446-4865
Başvuru Tarihi (Received): 01.02.2026 **Kabul Tarihi** (Accepted): 29.04.2026

Introduction

Organizational decision-making mediated by artificial intelligence (AI) - from credit scoring to clinical decision support, hiring and management of performance - has consequential impacts on humans and larger societal impacts (Diakopoulos, 2016, p. 57; Kroll et al., 2017, p. 636; Gerlings, 2025, p. 21). As these systems are being integrated into core operations, demands for transparency, accountability and explainability have risen, and often have outpaced current arrangements of governance (Banerjee et al., 2024, p. 426).

This pressure manifests itself in an "explainability dilemma" when it comes to organizational AI: while stakeholders increasingly demand easy-to-use, positive explanations for AI-based results, organizations struggle with technical, economic and institutional boundaries to deliver them in the necessary quality and scale (Asif et al., 2023, p. 5). High-performing machines for ML often give up interpretability to achieve a better response, reinforcing the feasibility of a "black-box" for making decisions in many of their deployments (Ribeiro et al., 2016, p. 1135; Lundberg & Lee, 2017, p. 4765; Lipton, 2018, p. 38). Responsibility for explanations also fragments across roles - developers, business owners, compliance functions and executives - creating coordination failures, accountability gaps, when there are waves of explanation arguments, incompleteness and strategic withholding (Banerjee et al., 2024, p. 430).

These are exacerbated by regulatory requirements. The EU AI Act establishes transparency and documentation requirements for high-risk systems, forcing organizations to keep comprehensive records of the source of data, model development and training processes, and rationales behind decisions taken The EU AI Act imposes transparency and documentation requirements on high-risk systems, we can see an effort in organizations to keep records and traceability of the source of data, model development and the training process, and the rationale behind decisions made (European Commission, 2021). The GDPR likewise protects safeguards and "meaningful information" for automated decision-making processes, in particular Articles 13-15 and 22, with prioritization on the logic and likely consequences associated with individuals affected (European Parliament & Council, 2016). In the United States sectoral regimes create explanation requirements in areas such as credit (Kroll et al., 2017, p. 636). Noncompliance entails legal, financial and reputational risks (European Commission, 2021; European Parliament & Council, 2016) to the organizations.

Existing approaches are only part of the solution to the dilemma. Technical explainability tools like LIME and SHAP can generate post-hoc interpretations of model outputs, but they have, of course, not answered who is required to produce explanations, who checks if explanations are valid and who takes responsibility when explanations fail (Ribeiro et al., 2016, p. 1136; Lundberg & Lee, 2017, p. 4765). Governance frameworks in practice are also frequently a technical addition and neglect explainability as a sociotechnical issue with role structure, incentives and institutional pressure (de Bruijn, et al., 2022, p. 7). Meanwhile, traditional Audit trails under centralized systems are subject to tampering or selective disclosure or made unverifiable to the outside world, making it less valuable as accountability infrastructure (Raji et al, 2020, p. 33).

Blockchain technology is one possible solution to this problem as it provides powerful means for tamper-resistant distributed records that can be verified by several parties without the need for a single trusted intermediary entity (Nakamoto, 2008, p. 1; Beck et al., 2018, p. 1021; Lumineau et al., 2021, pp. 507-508). Smart contracts can be used to write and automate compliance checks and responsibility triggers, that can possibly shift explainability from informal expectations to valid enforceable protocols (Szabo, 1997; Christidis & Devetsikiotis, 2016, p. 2296; Buterin, 2014, p. 15). Recent scholarship contains work on AI settings designed to explore decision traces and auditability of promotional fairness decisioning using blockchains (Parlak, 2025, p. 1), AI model lifecycle documentation (Butt et al., 2023, p. 1), and AI decision amplification-procedures which

would include the calculation of coverage and contexts of these tools as a way to accommodate explanatory friction (Jarsania et al., 2025, p. 332).

Yet there is a fundamental conceptual gap in that there is no cohesive theoretical framework for distributing the "explainability burden" - the responsibility for producing, maintaining, verifying and communicating explanations for AI decisions - across organizational stakeholders in realistic capacity and constraint conditions (Lipton, 2018, p. 36; Gerlings, 2025, p. 27). Explainability is not a binary characteristic of a system, and it depends on the stakeholder, context, and regulation context. While developers may need model-internal diagnostics, business owners may need rationales that make sense for their business, affected individuals need intelligible justifications for their care, and regulators want verifiable evidence of compliance (Ehsan & Riedl, 2020, p. 450; Gerlings, 2025, p. 54). There are costs that every demand imposes in terms of computation, expertise, coordination, and risk exposure and burden allocation is more of a governance issue than a modeling choice (Lipton, 2018, p. 40).

This article fills this void by suggesting a conceptual-theoretical model of allocate explainability burden in blockchain based AI governance systems. The model builds on the scholarship related to algorithmic accountability (Diakopoulos, 2016, p. 58; Kroll et al., 2017, pp. 633-634), institutional theory on structures of legitimacy and responsibility (DiMaggio and Powell, 1983, p. 147; Scott, 2014, p. 74), and distributed governance theory on other forms of coordinated authority in decentralized settings (Ostrom, 2010, pp. 641-642; Beck et al., 2018, p. 1020).

The research questions guiding the study are:

RQ1: What is the concept of “explainability burden” and how can it be theoretically specified and operationalized as a measurable organizational resource which needs to be distributed across stakeholders in AI-driven decision systems?

RQ2: What are the significant stakeholder roles and layers of responsibility in organizational AI systems, and how does their explanation responsibility vary according to their technical capacity, and their responsibility to make decisions, etc. (as well as their exposure under regulation)?

RQ3: What might be done with blockchain technology and smart contracts to implement transparent, auditable and enforceable explainability burden allocation protocols?

RQ4: When does the explainability burden allocation built upon blockchain increase organizational accountability, stakeholder trust, regulatory compliance and system legitimacy?

To answer these questions, I develop a formal model which involves stakeholder-specific expectations of explainability that is a function of risk exposure, technical sophistication and capacity, subject to participation, fairness, and budget constraints. I then discussed architecture for a blockchain-based approach by implementing the allocation in terms of smart contracts and immutable decision logs and making the verification multi-stakeholder based. Finally, I form testable propositions relating the internal relationship between blockchain-based explainability governance and the resulting outcome of accountability and stakeholder trust in conjunction with regulatory compliance and organizational learning.

1. Conceptual Framework and Literature Review

1.1. The Algorithmic Accountability Theory

Algorithmic accountability research raises the question of how to make automated systems that make decisions account to human oversight and societal values (Diakopoulos, 2016, p. 60; Kroll et al., 2017, p. 637). In organizational context, accountability extends across (i) the intelligibility of decisions techniques (i.e., can decision logic be explained or audited?) and (ii) the answerability of decisions institutions (i.e., does a person in charge have to provide justification for decisions to

an appropriate forum who can assess and sanction them?)(Bovens (2007, p. 452), Wieringa, 2020, p. 1). This literature very often makes sharply explicit the notions of transparency (ability to see into the mechanics of systems components), of interpretability (ability to interpret decision logic) and of contestability (ability to contest and appeal against system results).

Explainable AI (XAI) forms part of providing technical devices to approximate or summarize model behavior. LIME builds local surrogate models for individual predictions, and SHAP applies the logic of Shapley's value contribution to attribute feature contributions (Ribeiro et al., 2016, p. 1137; Lundberg & Lee, 2017, pp. 4765-4766). Counterfactual explanations focus on small changes in input which should change a prediction, therefore they are more relevant for the claims of actionability for high-stakes decisions (Wachter et al., 2017, pp. 76-77). These kinds of tools go a long way to amplify what can be said about model outputs, but they do not solve the problems of determining who is responsible for explaining, how well explanations are being judged, and how explanations become organizational responsibilities rather than optional entities.

In the eyes of companies, there are still some limiting factors to consider when deploying XAI: If you want to get started with an ethical, explainable-everything approaches to multinational organizations, read Fat's 70-page Grand Designs Research report here. Hygiene developer and auditor expectations calculation to develop granular diagnostics to inform managers and those affected by the development rationales in line with the needs of the domain in comprehensible terms Post-hoc explanations can also misrepresent model behavior such that truth of transparency is made irrelevant between the perceived and actual transparency, increasing the possibility of "explanation theater" instead of meaningful accountability (Rudin, 2019, p. 1). Explanations also induce security, privacy and strategic risks as they may leak sensitive information or allow for attackers to exploit the system by "gameplay" (Shokri et al., 2017, p. 3; Tramèr et al., 2016, p. 601).

Accordingly, more recent studies have approached accountability as a challenge of sociotechnical governance and not necessarily primarily technical, focusing on role clarity, lifecycle auditing, and the organizational processes that govern and determine when explanations are needed and how they are adjudged (e.g. by de Bruijn et al, 2022, p. 6; Raji et al, 2020, pp. 33-34). The "black-box" problem - particularly for more complex models - makes the tension between model performance and human understanding more acute and makes it unrealistic to assume that accountability is possible with 100% interpretability (Lipton, 2018, p. 42). The central governance issue becomes the event of organizations allocating explanation obligations across an actor and across a decision layer under actual constraints.

1.2. Institutional Theory and Organizational Responsibility

Institutional theory is a theory that describes how organizations react to pressures from outside the sector (norms) for legitimacy by adapting to regulatory requirements, number, socially accepted norms, and cultural expectations of appropriate behavior (DiMaggio & Powell, 1983, p. 151; Scott, 2014, pp. 73-74). I have in the case of AI governance coercive mandates (such as GDPR and the EU AI Act), normative expectations articulated in the form of professional and ethical discourse, and imitative learning of "best practices" against the backdrop of uncertainty (European Commission, 2021; European Parliament & Council, 2016; DiMaggio & Powell, 1983, pp. 150-152). Because the transformation of the decision-justification principles, as well as the jurisprudence built around these norms, and the procedure of contesting the goodness of the results made possible by AI, means that legitimacy-questions lie at the very center-are not extraneous to the acceptance or rejection of the appropriate outcomes of the whole AI process by the stakeholders and the assessment by the regulators that the governance provided is adequate or not.

One of the institutional challenges that exists on a recurring basis will be responsibility diffusion. Organizational AI systems have multiple actors with partly overlapping powers - model

developers, data providers, product and process owners, business units responsible for system deployment, compliance teams, and executives with the ultimate responsibility (Banerjee et al., 2024, p. 425). Distributed roles can increase specialization, but they also create spaces where no actor is in a position to give a plausible account, check its accuracy, or suffer consequences when actors question answers to stakeholders (Bovens, 2007, p. 458; Nissenbaum, 1996, p. 27). Institutional theory thus prompts a taxonomy of governance design needs: explainability mandates have to be made operational in ways which specify responsible roles, acceptable evidence and enforceable procedures rather than being aspirational principles.

1.3. Distributed Governance and Blockchain

Distributed governance characterizes systems of governance in which authority and responsibility are so interspersed among actors and oriented according to protocols, shared rules and transparent information flows (Ostrom, 2010, p. 647; Beck et al., 2018, p. 1027). Blockchain operationalizes aspects of this logic insofar as it gives us an immutable ledger which can be verified by different parties and can be updated through the consensus with respect to the ledger rather than through unilateral control (Nakamoto, 2008, p. 3; Beck et al., 2018, p. 1027). Smart contracts are an extension of such infrastructure, which provide programmable rules that execute obligations and checks on an automated basis, which may translate governance expectations into mechanisms that can be enforced (Szabo, 1997; Christidis & Devetsikiotis, 2016, p. 2296-2297; Buterin, 2014, p. 14).

In demand of AI governance settings, blockchain-based audit can solve the weaknesses of centralized logs. For example, immutability makes the retroactive changed records of decisions less easy to change; transparency allows the independent verification of different stakeholders; the only automations of smart contracts can manage the compliance triggers and escalation processes (Nakamoto, 2008, p. 5; Beck et al., 2018, pp. 1027-1028; Christidis & Devetsikiotis, 2016, p. 2296). Prior work suggests block chain aid in constructing decision traces and life cycle registries to provide connections between predictions and explanation artefacts as well as model documentation to enhance audit readiness and dispute resolution capacity (Parlak, 2025, p. 1; Butt et al., 2023, p. 3). Voting or coordination mechanisms can also be used for distributing some choices related to governance across stakeholders while leaving an auditable footprint.

At the same time, governance with blockchain has well-known limitations on its design which influences what is truly feasible to implement. Scalability tends to restrict the practicality of record explaining window data on-chain, which has fostered hybrid architectures with at least partial artefacts stored off-chain, with integrity of the well-constrained part supplied by a hash (Croman et al., 2016, p. 109). Privacy obligations may clash with the immutability of blockchains particularly due to restrictions associated with the European General Data Protection Regulations (GDRP), which results in a careful architectural approach to what is recorded and feature access.1 Smart contracts pose other risks as vulnerabilities may have unintended consequences and erode trust in enforcement logic (Atzei et al., 2017, p. 168). Finally, blockchain governance in and of itself requires mechanism for dispute resolution and protocol adaptation (De Filippi & Loveluck, 2016, p. 8-9). These constraints make making allocations of burden more central, not less.

1.4. Literature Gap and Research Contribution

Across algorithmic accountability, institutional theory, and blockchain governance, there appears to be a blank blind spot: the inability of a coherent framework to specify and operationalize how the responsibilities to provide explainability are determined and allocated among various organizational stakeholders when blockchain-based infrastructure is used to establish accountability. XAI research progressed and produced techniques for generating explanations but says little about how to assign roles, who should verify explanations, or enforce changes at the organizational level (Ribeiro et al., 2016, pp. 1135-1136; Lundberg & Lee, 2017, p. 4765).

Institutional theory explains why legitimacy pressures are important, but it provides limited insights on how to measure explainability requirements, balance heterogeneity in the interests of various stakeholders or implement enforceable rules for allocation in AI systems (DiMaggio & Powell, 1983, p. 151; Scott, 2014, p. 73). Blockchain governance research illustrates the technical feasibility of tamper-resistant log structures and automated compliance but generally focuses on capturing and tracking resource movements rather than with a systematic account of who is responsible for the ongoing work of generating, maintaining, validating, and communicating explanations (Beck et al., 2018, p. 1021; Lumineau et al., 2021, p. 507; Parlak, 2025, p. 1).

The present article of this study attempts to bridge these gaps by formulating the conceptual/theoretical framework that:

1. **Defines explainable burden** as the measurable organizational resource which includes the technical, human and co-ordination costs of producing, maintaining, verifying and conveying explanations for AI decisions.
2. **Identifies stakeholder** role and levels of responsibility in organizational AI systems, to specify how explainability obligations vary among AI developers, data providers, process owners, and auditors depending on their positions in terms of their technical abilities, decision power, and regulatory exposure.
3. **Formalizes explainability burden allocation** by a mathematical model, which quantifies as a function of risk, sophistication, and capacity, explainability expectations of individual stakeholders, which are subject to participation, fairness, capacity, and budget constraints.
4. **Proposes blockchain architecture** based on explainability burden allocation using smart contracts, immutability of decision log and multi-stakeholder verification protocols that will ensure transparency, auditability and enforceability.
5. **Derives testable propositions** that can be formulated between blockchain-based explainability governing and accountability outcomes, stakeholder trust, regulatory compliance and organizational learning.

By combining the technical focus of algorithmic accountability (Diakopoulos, 2016, p. 59; Kroll et al., 2017, p. 637), the legitimacy and responsibility logic of institutional theory (DiMaggio and Powell, 1983, p. 147; Scott, 2014, p. 60), as well as coordination mechanisms in focus in the case of distributed governance (Ostrom, 2010, p. 642; Beck et al., 2018, p. 1028), the framework constitutes a systematic approach on how to deal with the explainability dilemma in organizational AI systems by using blockchain-enabled governance. This framework combines the preoccupation in algorithmic accountability theory with technical transparency, the legitimacy and responsibility structures in institutional theory, and coordination in decentralized governance theory. By connecting these traditions via tangible affordances offered by blockchain technology, I derive a coherent model of managing the explainability dilemma for organizational AI technology.

2. Theoretical Model Development

2.1. Defining "Explainability Burden" Concept

It on explainability burden I introduce, in placing the idea as a basic construct in AI governance. Explainability burden is the sum of the resources-use-cost, costs in talent or skills, and organizational resources, i.e., technical, human, and organizational costs-rather needed to produce, maintain, verify, and communicate explanations for decisions taken by artificial intelligence in such a way that meets the needs of stakeholders and explores regulators/legislators.

Explainability burden includes five different components:

1. Generation Cost (C_gen): Computational and human resources required to produce explanations e.g. by applying the methods for XAI like LIME, SHAP, counterfactual generation (Ribeiro et al., 2016, pp. 1137-1138; Lundberg & Lee, 2017, p. 4766). This includes the topic of execution time of algorithms, infrastructure costs, and data scientist's efforts in (structuring and validating explanation methods.

2. Maintenance Cost (C_main): The ongoing resources that are required to change explanations as models are re-training, data distributions change or as requirements of stakeholders change. AI systems are not static but require constant updating of monitoring and explanations for them to be accurate and relevant (Sculley et al., 2015, p. 2509).

3. Verification Cost (C_ver): The resources that are needed for validation of explanations that also represent faithfully what the models are doing, that are also technically correct and that align with the knowledge of the domain. Verification can be human expert review, automated consistency checks or empirical testing of explanation quality (Lage et al., 2019, p. 60).

4. Communication Cost (C_comm): The labor required to present technical explanations using formats suitable for various audiences of stakeholders - from specific technical documentation for auditors to simple information on why the impacted people should care (Ehsan & Riedl, 2020, p. 461; Miller, 2019, p. 9). Communication must be effective and that means an understanding of stakeholder literacy, preferences and information needs.

5. Coordination Cost (C_coord): The organizational overhead of managing the explainability responsibilities across a range of stakeholders, resolving conflicts among them, ensuring accountability Expenditure on Coordination costs are increased by number of stakeholders, complexity of decision processes and ambiguity of responsibility assignments (Bovens, 2007, p. 458).

The sum of the explainability burdens of all the components taken together for an AI system can be accounted for as:

$$B_{total} = C_{gen} + C_{main} + C_{ver} + C_{comm} + C_{coord} \quad (1)$$

Explainability burden differs from any decision context depending on few factors:

- **Decision Risk:** High-stakes decisions (e.g. medical diagnoses, credit denial, criminal sentencing) have a greater burden of explainability because of the potentially disastrous consequences of getting the decision wrong, and the onus of having to demonstrate a greater level of justification to the relevant stakeholders (Kroll et al, 2017, pp. 636-637).
- **Model Complexity:** More complex models (e.g. deep neural networks with million parameters) require more sophisticated explanation methods and more verification efforts than simpler models (e.g., decision trees, linear regression) (Lipton, 2018, p. 38; Rudin, 2019, p. 3).
- **Stakeholder Diversity:** There are greater communication and coordination costs for systems that are serving different stakeholder groups with different technical literacy and information needs (Gerlings, 2025, p. 27; Ehsan & Riedl, 2020, p. 461).
- **Regulatory Stringency:** Stricter regulatory requirements (e.g. Europe Artificial intelligence (AI) Act High-risk classification) involve higher documentation, conflicting, verification and auditing obligations (European Commission, 2021)

2.2. Stakeholder Roles and Responsibility Layers

Organizational AI systems include a range of stakeholders with varying roles, capabilities and responsibilities for accountability. I come up with four main categories of stakeholders:

1. AI Developers (S_dev): Data Scientists, Machine Learning Engineers, Software Developers who will build, educate and implement the A.I Models. Developers are highly technically experienced in model architecture, training algorithms and XAI methods. Their explainability obligations consist of model design decisions documentation, technical explanation generation, explanation accuracy verification, and explanation system maintenance for evolving models (Gerlings, 2025, p. 54).

2. Data Providers (S_data): Individuals, organizations or systems who conduct training and operational information for the use of AI systems. Data providers can be internal business units, external vendors of data sources, or automated systems to gather the data. Their explainability requirements involve documenting data provenance, quality, data preprocessing, and feature clear choices and disclosing data daily limitations and biases (Gebru et al., 2021, pp. 90-91).

3. Process Owners (S_proc): Business managers that deploy artificial intelligence systems in business processes and make final decisions based on AI recommendations. Business managers, product owners, and operational staff implement artificial intelligence systems in the business process and make final decisions based on AI recommendations. Process owners are domain experts and decision makers with potential knowledge debts when it comes to technical AI. The explainability requirements they obligate them to is to define explainability requirements according to the needs of stakeholders, explaining AI-generated explanations to the people who are affected, and explaining the final decision that has been made by incorporating the recommendations from AI (de Bruijn et al., 2022, p. 6).

4. Auditors (S_audit): Internal compliance officers or external regulators, external audit that checks the compliance of the AI-systems with legal, ethical and organizational standards. Auditors need full access to system documentation, decision records and explanations. Their explainability obligations include that they should check the explanation completeness and accuracy, they should check the compliance with the regulatory requirements, and they should identify accountability gaps (Raji et al., 2020, pp. 37, 41).

These stakeholder roles are operating on three layers of responsibility:

Layer 1 (Technical Layer): Is focused on model development, training and technical explanation generation. Mostly the realm of developers of artificial intelligence, this layer concerns issues such as: What features does the model use? How is it that predictions are calculated? What is its performance on different subgro ups for the model?

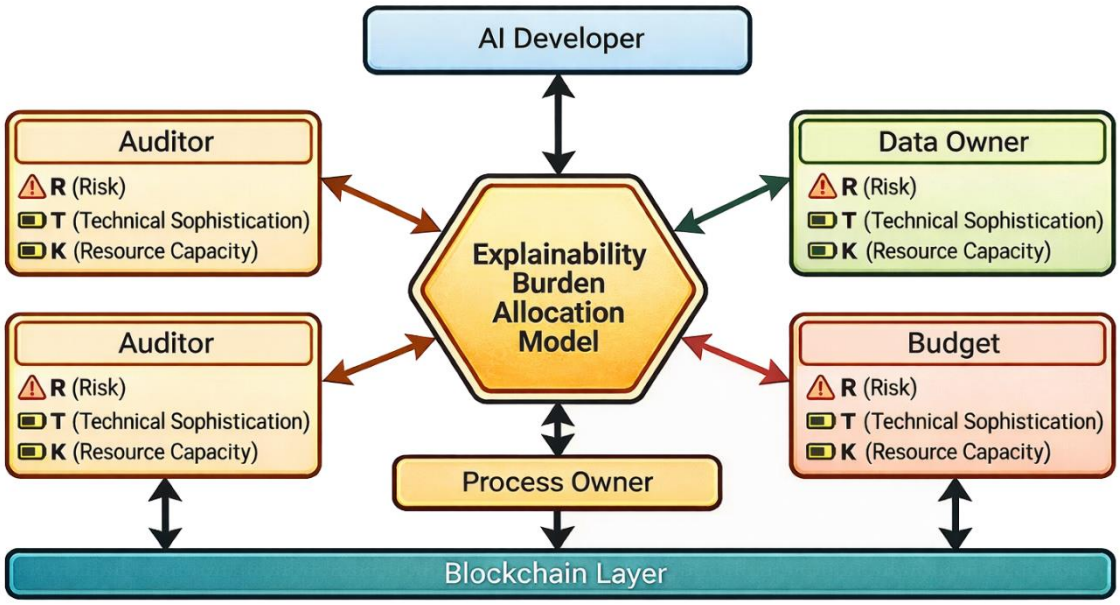
Layer 2 (Operational Layer): Focuses on implementation of Artificial Intelligence systems with regards to processes in organizations and communicating explanations to individuals in the organization. Primarily the domain of the process owners and data providers, this layer looks behind questions like these: Why was this particular decision made? What could the alternative results have been? How are decisions of affected people possible?

Layer 3 (Governance Layer): Layer 3 (Governance Layer): Addresses oversight and validate compliance and enforce an accountability. Mostly the prerogative of the auditor and of top management, this later includes questions about: Does the system comply with regulatory requirements? Responsibility Assignments Are clear, enforceable? What are the means of detecting failures and correcting them?

Figure 1 here gives a broad overview of the explainability burden allocation framework, showing the center of this allocation model along with those connected to the burden of explanation (AI

Developer, Data Owner, Process Owner, Auditor and Budget) by various dimensions of responsibility (Risk, Capacity, Sophistication and Authority). The blockchain layer in the bottom represents the unchangeable infrastructure which stores all explainability commitments and verifications.

Figure 1: *Comprehensive Conceptual Framework Overview*



2.3 Explainability Burden Allocation Model

A formal definition of the explainability burden allocation problem can be stated a follows the following: Given a set of stakeholders S = {s_1, s_2, …, s_n} and a total explainability B_{total}, be the optimal distribution of burden shares $\{b_1, b_2, \ldots, b_n\}$ such that:

1. Each stakeholder's burden is in proportion to their ability, risk and distance to decision-making
2. The allocation meets fairness issues against excessive burden concentration
3. The allocation is implementable against resource constraint of stakeholders
4. The allocation maximizes overall system accountability and legitimacy

I define the explainability expectation of each stakeholder as:

$$E_i = \alpha \cdot R_i + \beta \cdot T_i - \gamma \cdot K_i \quad (2)$$

Where:

- E_i = *Stakeholder* i's expectation of explainability (to what level, to what detail of explanation they expect to be explained)
- R_i = *Stakeholder* i's risk exposure (exposure to potential harm posed by wrong/ unexplained decision making)
- T_i = *Stakeholder* i's technical sophistication (ability to understand and come up with technical explanations)

- $K\_i = Stakeholder\ i$'s resource capacity (budget, personnel and infrastructure available to be used in explainability activities)

The burden share allocation of explainability function allocates burden shares based on these expectations:

$$b_i = \frac{E_i \cdot w_i}{\sum_{j=1}^{n} E_j \cdot w_j} \cdot B_{total} \quad (3)$$

Where w_i is the weight in the allocation of $stakeholder\ i$, depending on their role and their position close to or far from decision-making.

2.4 Mathematical Formulation

In this paper, I express the explainability burden allocation as a constrained optimization problem:

Objective Function: Maximize overall system accountability A, defined as:

$$A = \sum_{i=1}^{n} \alpha_i \cdot \min(b_i, E_i) - \beta \cdot \sum_{i=1}^{n} \max(0, b_i - K_i) \quad (4)$$

Where:

- α_i = Accountability weight for $stakeholder\ i$ (how important is this stakeholder on accountability chain)
- β = Penalty coefficient for any excess of capacity constraints
- The first term rewards allocations that meet stakeholder expectations
- The second term penalizes allocations that are made to exceed stakeholder capacities

Subject to constraints:

Constraint 1 (Participation): Every stakeholder should incur at least some minimum burden if engagement is to happen:

$$b_i \geq b_{min}\ \forall i \in \quad (5)$$

Constraint 2 (Fairness): No stakeholder's disproportionate burden to stakeholder's capacity should be imposed:

$$\frac{b_i}{K_i} \leq \theta \cdot \frac{1}{n} \sum_{j=1}^{n} \frac{b_j}{K_j} \forall i \in \quad (6)$$

Where θ is a parameter of fairness tolerance (e.g., $\theta = 1.5$ means that there is up to 50% deviation from average burden-to-capacity ratio).

Constraint 3 (Capacity): Individual allocations of burden cannot be more than stakeholder capacities:

$$b_i \leq K_i\ \ \forall i \in S \quad (7)$$

Constraint 4 (Budget): Burden allocated must equal burden system:

$$\sum_{i=1}^{n} b_i = B_{total} \quad (8)$$

Constraint 5 (Role-specific minimums): Certain roles of stakeholder have minimum threshold of burden:

- AI Developers: $b_{dev} \geq 0.3 \cdot B_{total}$ (at least 30% of the explanations of technical information are the responsibility of developers)
- Auditors: $b_{audit} \geq 0.15 \cdot B_{total}$ (auditors take minimum burden of 15% of burden of verification)

2.5 Model Assumptions and Limitations

To clarify the scope of the analysis and the conditions to which the allocation results apply, my model is based on a number of important assumptions:

Assumption 1: Explainability burden can be usefully explained and measured. While accurate quantification is difficult some quantification of burden can be made by organizations through resource tracking, time studies and cost accounting (Sculley et al., 2015, p. 2510).

Assumption 2: Stakeholder capacities and expectations are known/are able to be assessed in actual practice, these parameters can be subject to enforced refinement based on the results of pilot implementations, and the opinions and feedback of stakeholders.

Assumption 3: Stakeholders act in good faith and act by allocated responsibilities. The blockchain architecture is capable of enforcement mechanisms yet incapable of eliminating the non-compliance and strategic behavior all together.

Assumption 4: The optimization objective (maximizing accountability) is sufficient for capturing organizational goals. Alternative efforts with other objectives (e.g. minimum cost, maximum stakeholder-satisfaction) may be appropriate in other contexts.

Limitations:

1. **Dynamic environments:** The model is based on relatively stable stakeholder roles and capacities. In an environment where things are changing rapidly, frequent redistribution may be required.
2. **Interdependencies:** The model considers the stakeholder burden to be independent, whereas in reality activities that serve to explain to stakeholders might decrease the burden for other stakeholders (e.g. comprehensive developer documentation decreases auditor verification efforts).
3. **Quality vs. quantity:** The model is concerned with allocation of burden rather than discussing the indirect explanations of actual quality. Additional mechanisms are required to make sure the responsibilities allocated are translated into good explanations.
4. **Cultural and contextual factors:** The model's parameters (weights, thresholds, fairness tolerance), model parameters may have to be adjusted for them to work in different organizational cultures, industries, and regulatory environments.

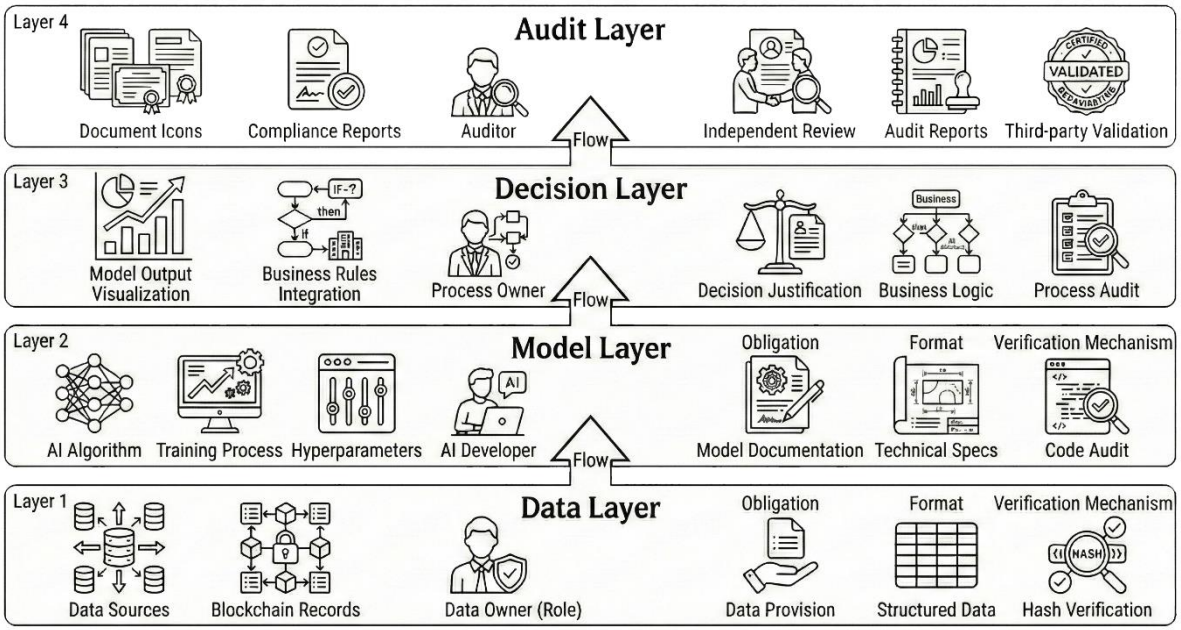
3. Graphical Representation of the Conceptual Model

3.1 Layered Responsibility Map

Explainability accountability can be built in layers of responsibility from the data, model, decision, and audit levels. Data-layer obligations focus on provenance and integrity evidence (e.g. hashed records) to have an explanation of what was used as input with or under which constraint. Model-layer obligations. model layer migrations model layer change tracking model layer design Decision-layer obligations are ways of translating the model outputs and operational justifications as they link the predictions with the organizational rules and human supervision. Audit-layer obligations gather these artifacts into reviewable compliance evidence that allows for either independent verification as well as accountability escalation figures.

The layered responsibility map visualizes how explainability obligations are distributed between (Technical, Operational, and Governance) layers of responsibility and four different categories of stakeholders (Developers, Data Providers, Process Owners, and Auditors).

Figure 2: *Layered Responsibility Map*

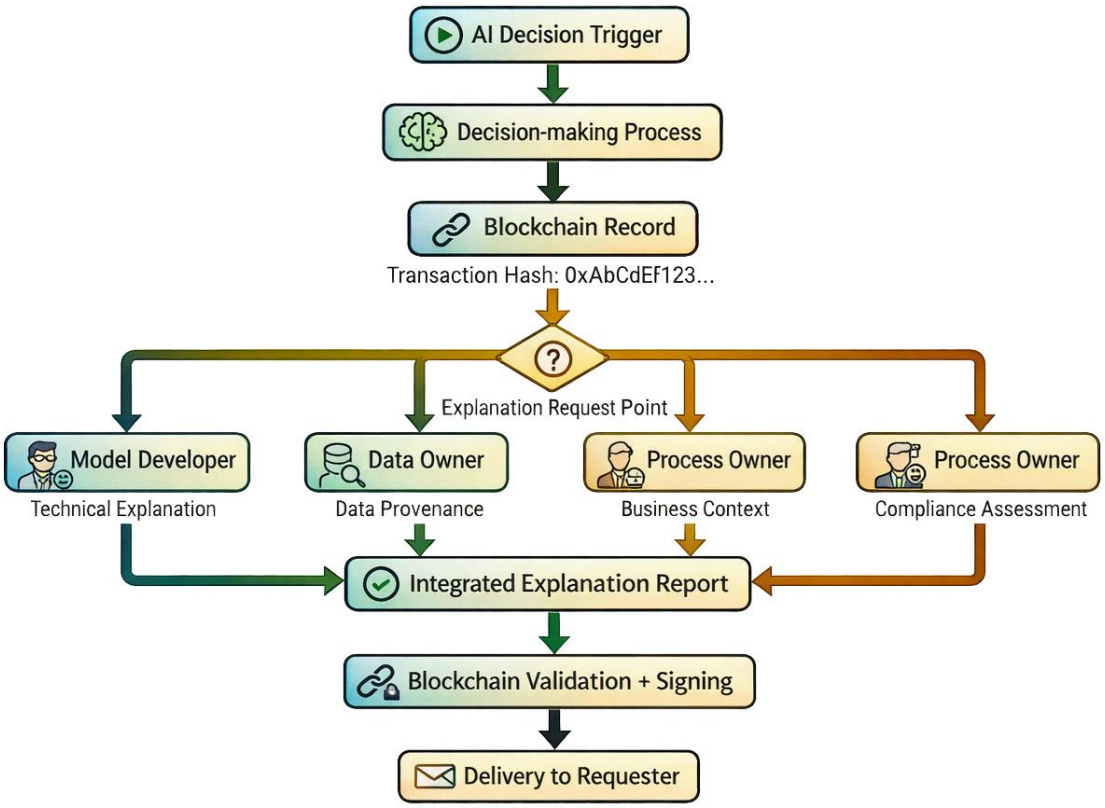


3.2 Explanation Flow Diagram

A decision-to-explanation workflow would need to be equipped with some standard trigger, role-based contribution, consolidation and verification sequence. After an AI decision is produced and bound to an indelible record, an explanation request triggers some of the distributed responsibilities of providing an explanation: the developers of the explanation satisfy technical explanation; the data providers of the source(s) satisfy the explanation of provenance; and the process owners satisfy potential contextual domain and compliance interpretation. These components are combined into a single package of explanation which is then multi-partly attested to, resulting in a traceable chain of contributing and reviewing before disclosure to the requesting subject stakeholder.

The explanation flow diagram shows how explanations are generated, verified, communicated and audited by different stakeholders with each step recorded in blockchain.

Figure 3: *Explanation Flow Diagram*

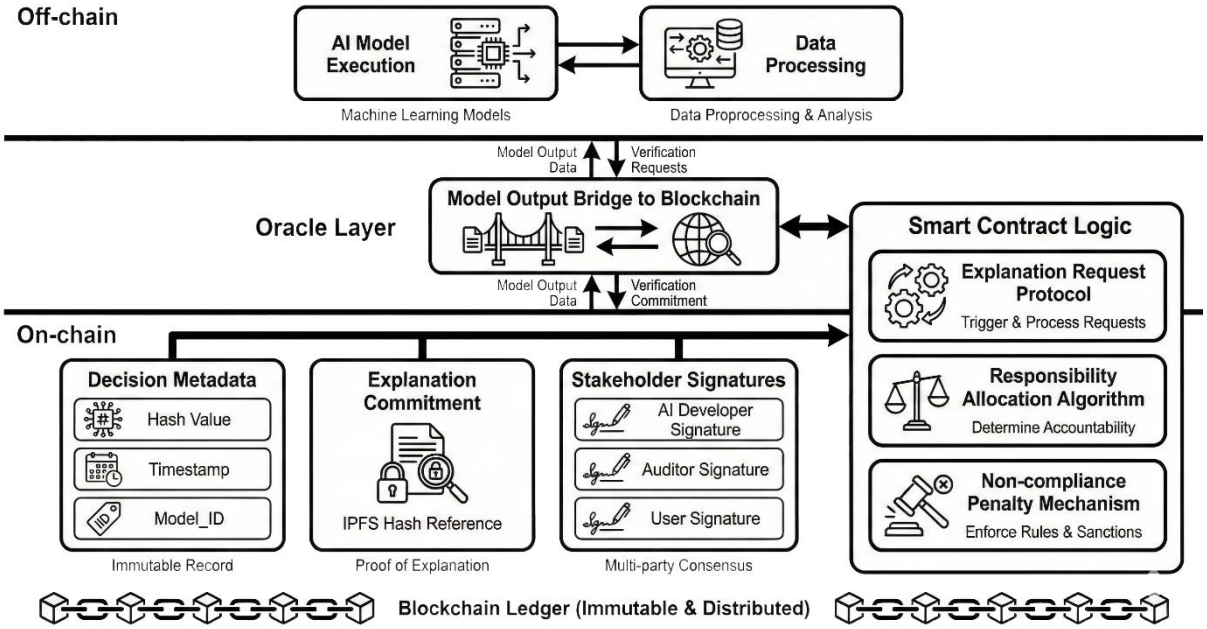


3.3 Blockchain Architecture Schema

A hybrid architecture is an approach to segregating the high volume computation from the verifiable accountability. Model execution, feature engineering and generation of detailed explanation artifacts are all off chain where storage and throughput issues are easier to handle. The identifier of decisions, timestamp, model/version reference, cryptographic commitment to explanation contents (e.g., hashes) and an attestation by stakeholders responsible for the document documenting who produced, reviewed and approved the explanation are all recorded on the blockchain layer. A bridging mechanism is created to connect operational AI systems and log on to the blockchain, triggering the governance process if there is a dispute against a given decision or it needs to undergo review. Smart contracts then operationalize allocation rules by assigning allocations, ahead of time, when obligations are not achieved as well as deadline enforcement is done, and compliance formally registered; and providing for escalation in cases of unmet obligations or capacity limits are exceeded. This ledger design is as an integrity and accountability anchor rather than a data warehouse: It maintains the property of scalability and at the same time makes explanation evidence tamper resistant, auditable among parties and enforceable using transparent, rule-based procedures.

The blockchain architecture implements the explainability burden allocation model through smart contracts and distributed ledger technology.

Figure 4: *Blockchain Architecture for Explainability Governance*



Key Components:

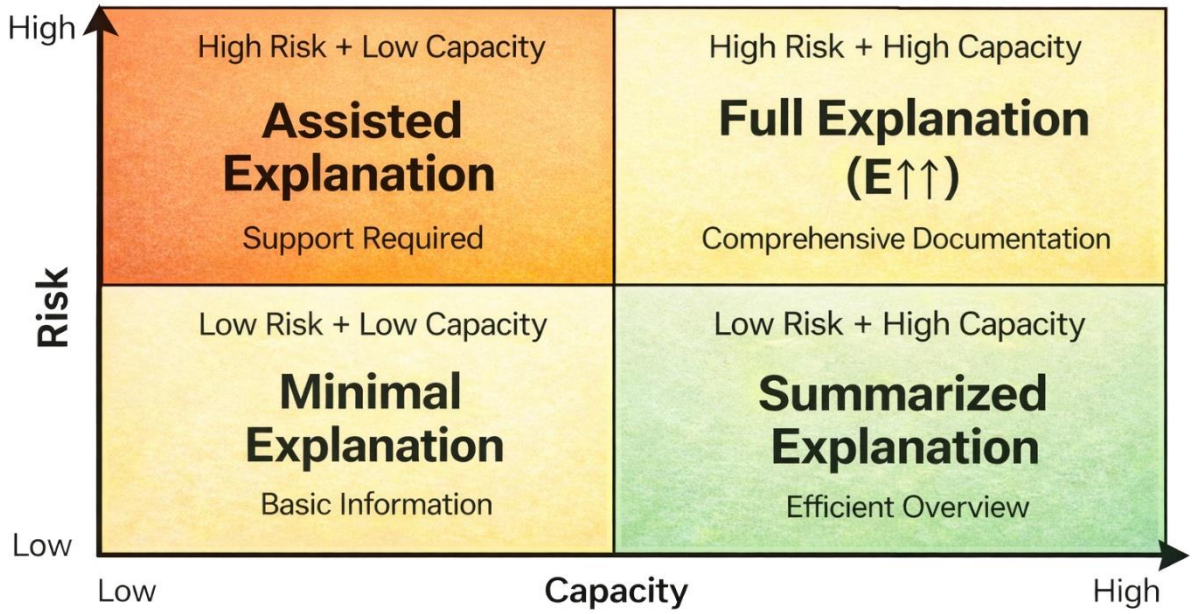
1. **Stakeholder Interface Layer:** Offers role-based interfaces for the various stakeholders to meet their explainability requirements and obtain the relevant information.
2. **Smart Contract Layer:** Provides the logic of allocation in Equations 24, the record of the adherence to the obligations, as well as the implementation of predetermined measures of enforcement in case of non-compliance with the requirements.
3. **Blockchain Ledger Layer:** An immutable record of decisions, explanations, contributions from the stakeholder and results of audit.
4. **Off-Chain Storage Layer:** Contains valuable explanations and documentations in distributed file storage, with cryptographic hashes are noted in blockchain for purposes of verification.

3.4 Decision Tracking and Audit Loop

Explainability intensity may be based on not one size that fits all approaches but a risk vs. capacity logic. High-risk situations necessitate more exegeses in the event of organizational capacities for support production and verification; however, low risk situations may be revealed in a more simplified manner. Where high risk but limited capacity is the case, helps or redistributed explanation production to avoid accountability failure, Where the capacity is high, but the risk is low, there can be a use of summarized explanations for the standardization of practice, without undue overhead being placed on it. This mapping in effect operates burden allocation in terms of a function of exposure and resources that is feasible.

The loop of the decision tracking and audit process shows the ongoing cycle of tracking and improving thanks to explainability governance using blockchain technology.

Figure 5: *Risk-Capacity Matrix for Explainability Burden Allocation*



Allocation Strategy:

- **Quadrant 1 (High Risk, High Capacity):** Explainability burden is maximum to the stakeholders. They have the capability and powerful incentives as per the risk expose.
- **Quadrant 2 (High Risk, Low Capacity):** The stakeholders are not created or generated but explained. Pressure imposed on other stakeholders.
- **Quadrant 3 (Low Risk, High Capacity):** Instead of creating explanations, the stakeholders are provided with them.
- **Quadrant 4 (Low Risk, Low Capacity):** Stakeholders have little responsibility and simply the basic documentation.

4. Propositions and Hypotheses

Using the theoretical framework of the combination of algorithmic accountability, the institutional theory, and blockchain governance, I deduce five provisional suppositions that are testable.

Proposition 1 (Accountability Enhancement): Those organizations that enact blockchain explained allocation of burdening will experience greater algorithmic responsibility as compared to organizations that apply traditional central audit system.

Theoretical Justification: The transparency and irreversibility of the blockchain helps in the existence of auditable audit trails, which deter ex- post manipulation of records of decision-making (Nakamoto, 2008, p. 3; Parlak, 2025, p. 2). Monitoring and enforcement routines can be formalized using smart contracts to minimize accountability gaps owing to an ambiguity of responsibility boundaries (Christidis and Devetsikiotis, 2016, p. 2296). The allocation model also gives out description responsibilities to each decision, which directly responds to coordination failures highlighted in institutional explanation of accountability (Bovens, 2007, p. 458).

Testable Implications: Compared to centralized audit regimes, blockchain-based systems must demonstrate that (a) accountability disputes will be resolved at a faster rate, (b) more frequently explained decisions that are contested will be delivered and (c) accountability gaps in which responsibility is denied or not taken are observed.

Proposition 2 (Stakeholder Trust): An explainability governance based on blockchains will enhance the stakeholder trust in AI decision-making systems, especially the stakeholders with high-risk exposure and with limited technical expertise.

Theoretical Justification: The belief in sociotechnical systems is determined by credible transparency and verifiability (Lumineau et al., 2021, p. 508; Beck et al., 2018, p. 1027). Blockchain decreases the use of organizational assurance because it facilitates independent verification that explained things were created, revised, and documented (Nakamoto, 2008, p. 1). This verification ability is also particularly relevant to non-technical stakeholders, in which it may minimize information asymmetry and a sense of perceived fairness (Gerlings, 2025, p. 54; Ehsan and Riedl, 2020, p. 462).

Testable Implications: Stakeholders in blockchain-based systems report: (a) an increase on perceived fairness and trust, (b) an increase in perceived completeness and accuracy of the explanations and (c) a decrease in perceived organizational opacity.

Proposition 3 (Regulatory Compliance): Organizations that implement blockchain-based explainability burden allocation will have improved rates of regulatory compliance and lower rates of enforcement action than organizations without explainability burden allocation systems.

Theoretical Justification: The EU AI Act and GDPR set requirements of documentation and transparency which can be demanding in operation (European Commission, 2021; European Parliament & Council, 2016). Blockchain-based audit trails can establish and maintain a record of evidence relevant to these requirements (Parlak, 2025, p. 1; Asif et al., 2023, p. 4), while smart contracts can program regular compliance measures instead of relying on these measurements as seasonal activities (Christidis & Devetsikiotis, 2016, p. 2297). The burden allocation model clarifies who is responsible for satisfying certain obligations and is a way of improving traceability under regulatory scrutiny (Kroll et al., 2017, p. 637).

Testable Implications: Compared to other organizations that do not have such systems in place, then adopters are expected to show the following in regulatory audits: (a) better outcomes, (b) faster response to enquiries, and/or (c) less violations and enforcement actions with potential less cost on compliance from lesser work on reconstruction.

Proposition 4 (Organizational Learning): Blockchain explainability governance for organizational learning through: Establishing accessible repositories of decision rationales, patterns of explanations and outcomes for accountability decisions from the design of safe and responsible AI systems.

Theoretical Justification: Organizational learning is ground on systematic capturing/reusing of decision related knowledge (Argote & Miron-Spektor, 2011, p. 1130). Immutable logging of decisions and explanation artifacts can serve to support the retrospective analysis of decision patterns, performance of explanations and resulting accountability outcomes (Parlak, 2025, p. 3). This record can also cut down well-known operational failures in AI deployment through better traceability of moving from iteration to iteration of a system (Sculley et al., 2015, p. 2505), whereas multi-stakeholder verification underlies learning that takes multiple perspectives of heterogeneity in lieu of a solitary managerial viewpoint (Ostrom, 2010, p. 665).

Testable Implications: Blockchain-based systems should be linked with: (a) more frequent model and process improvements based on a historical analysis, (b) using to more quickly detect and correct systematic problems (even a pattern of bias), and (c) enhancing cross-unit knowledge transfer in the area of explanation practice.

Proposition 5 (Legitimacy and Isomorphism): As the institutional legitimacy of explainability governance is increasing, organizations will use these systems in copycat and normative birth, especially when the immediate technical gains in efficiency cannot be assured.

Theoretical Justification: Institutional theory, predicting adoption of legitimacy conferring practices on conditions of uncertainty and external pressure (DiMaggio & Powell, 1983, pp. 150-152; Scott, 2014, p. 73). As benefits of governance are shown by early adopters, peer organizations it is possible that they imitate this approach to signal commitment to a responsible AI (mimetic isomorphism). Professional communities and industry associations can further upgrade these systems as the accepted best (normative isomorphism), and the coercive pressure of regulatory acceptance of blockchain-based audit evidence (Suchman, 1995, p. 574).

Testable Implications: Adoption should: (a) be accelerated following prominent deployments by industry relevant leaders, (b) be correlated with being a member of emerging professional deadly AI networks, and (c) be concentrated in industries faced with prolonged regulatory attention with diffusion specifications decadent signaling lengths aside or ahead of effectiveness justification.

5. Implementation Framework and Managerial Implications

5.1. Implementation Protocol

Operationalizing the model requires the stable governance routines of transforming the allocation equations into governance routines. The protocol begins with the mapping of stakeholders in an organized way over the life cycle of AI (development - deployment - operations - and oversight) and is populated with 4 actor roles (Developers, Data Providers, Process Owners, and Auditors). At this stage, the organization will compile a list of the technical ability of each actor, the resources available and risk exposure and set up a formal set of reporting and communications pathways rules. Next, in conjunction with pilot implementation and resource tracking, the organization estimates both total explainability burden (B_{total}), then it disaggregates it into five activities (generation, maintenance, verification, communication, and coordination), so it does not define explainability as one undifferentiated obligation. With these inputs the parameters of the stakeholders (R_i, T_i, K_i) are measured and the expectations about explainability (E_i) are calculated using Equation 2.

Burden shares (b_i) are then derived by solving the constrained optimization problem (Equations 3-4) after which the fairness and capacity constraints as well as refinement through specific stakeholder consultation should be validated, both the rationale for the allocation and the resulting responsibilities should be documented and formally recognized. Implementation in this scenario then involves writing the logic for allocation and compliance into smart contracts (for example, solidity), in which various mechanisms for monitoring, alerts for noncompliance or capacity surpassing etc. are provided, and sandbox testing is provided before actually deploying the contract in production. Deployment questions include platform choice (transparency to general public vs privacy to private). node and key management Integration with off-chain storage solution such as ipfs/swarm for more detailed explanation artifacts End to end workflow integration with existing AIs. Finally, a process of ongoing monitoring takes into consideration results of allocation, feedback issued from stakeholders, and periodically re-optimization of allocations as capacities, risks, and regulatory requirements have been changed, while lessons learned are preserved and then applied to update the governance procedures as time goes by.

5.2. Architecture and Organizational Readiness

Implementation successes depend on up alignment between the technical architecture and organizational capacity, and governance needs and therefore platform selections need to be guided by 5 interlocking design criteria. First, the platform should have enough processing ability and scalability for the anticipated decision volume. Second, it must support privacy and confidentiality

controls (such as private channels or, where possible, zero-knowledge methods) that can be compatible with organizational and/or regulatory requirements. Third, the smart contract environment must be able to enact allocation rules and conduct continuous monitoring for adherence to the rules. Fourth, the platform must integrate easily with existing enterprise infrastructure, such as databases, APIs, and analytics systems must be in place so accountability evidence can be generated and used for routine workflow. Fifth, it must include credible governance mechanisms when it comes to protocol updates, dispute resolution and parameter recalibration given that risks, capacities and requirements change over time. Because on-chain storage and performance is still limited, these criteria typically mean a hybrid architecture whereby the explanation of the decisions is identified and timestamped on the blockchain, where the decision's explanation is plausible, whilst detailed explanations, supporting documentation, etc. is stored offline, where full integrity is provided via on-chain hashes.

Organizational readiness is as big or bigger than infrastructure. Technically, integration planning, blockchain development competence, and security practices of nodes and keys will be required for organizations. Organizationally, some executive sponsorship, some stakeholder training, and some cross-functional governance on the exceptions and disputes are necessary for implementation. Culturally, organizations should expect their resistance to transparency and work through this with phased pilots in limited and lower-risk areas that prove the utility of auditability and dispute handling.

5.3. Economic and Sectoral Considerations

Blockchain explanations governance should be assessed as a governance investment with value depending on the risk of decisions, intensity of regulation, and the frequency with which explanations are required by different audiences. The critical managerial question is whether and how responsibility can be distributed and addressed in a manner that limits duplication, blame shifting from strategic purposes and simulates and minimizes auditing friction taxation disputes on the side of operations. An appropriate activity-based assessment is to be made in which the effort for generation, maintenance, verification, communication, and coordination are estimated, and ad hoc practices are compared to structured allocations in terms of predictability of workload and evidence readiness.

The implementation costs can be categorized into five parts. They include infrastructure and operational overhead like in node-administration, storage design, and access control, smart contract development with its corresponding independent security review, integration with existing AI/IT systems and data governance pipelines, as well as training and change management so new accountability routines are institutionalized, and ongoing monitoring and maintenance since the allocation rules and parameters continue to evolve. Deployment risk can be reduced in both the phased adoption approach (where it is bounded to a single decision domain) and through consortium or shared services (where fixed costs can be related back and shared them by reusing governance artifacts, audits templates and verification tooling in different units or firms).

Benefits are seen more in terms of governance performance (rather than immediate cost savings): reduced need for audits due to verifiable audit records, accountability with clear chains of control limiting the scope for disputes to escalate, learning loops (through standardized logs) and enhanced trust (where explanation practices are perceived to be consistent and to be subject to enforcement). Sectoral patterns Regarding methods by which stakeholder networks shape allocation In the case of financial services credit scoring, the Developers responsibilities mostly lie on technical explanation and model documentation Process Owners on communication and justification responsibilities Data Providers on provenance and quality documentation responsibilities Auditors pre-occupation on verification The value of blockchain is most evident where there is a need to trail evidential trails for scrutiny. In the case of healthcare diagnostic support, burden falls on

Developers (validation and maintenance) and Process Owners (clinical integration and patient facing accountability), Data Providers and Auditors lean in more on provenance and feeding the brain on principle behind and workflow overview, traceability is for post hoc review. In HR hiring systems, Developers & Process Owners tend to share core work regarding bias testing & candidate-facing explanations, Data Providers tend to be central in regard to pipeline documentation, and Auditors tend to be central in terms of compliance verification, and blockchain tends to reduce ambiguity in case of decision-contestation.

6. Discussion: Contributions, Implications, and Limitations

This article contributes to an intersection among algorithmic accountability, institutional theory, and distributed governance with the introduction of "explainability burden" as a formal construct targeted to aggregate resource cost of generating, maintaining, verifying, communicating and coordinating explanations and refers to a shift in discourse from binary claims of transparency into an exploratory question on how explainability obligations should be distributed. It develops on the earlier studies by tying the issues with respect to technical transparency and the legitimacy of institutions together with decentralized coordination and expressing the division of responsibilities as a formal maximization problem (Equations 2-4) that allows to evaluate the aspects of fairness, efficiency and feasibility with the same accuracy; the earlier qualitative analyses were not able to do that.

In addition, the paper is rephrased as institutional infrastructure rather than a strict technical tool and gaining immutability, transparency, and programmability can rethink accountability practices through coercive, mimetic, and normative pressures and generate implications proposed for accountability tests, trust, compliance, learning, and legitimacy. Finally, the implementation framework in Section 6 offers an integrated set of operational assets, including protocol, architecture/readiness guidance, and economic/sectoral templates, without requiring practitioners to extrapolate the ability to implement from some abstract principles. These assets are fully stated in Section 6 and cannot be reiterated here so as not to create duplication.

The framework suggests concrete directions for policymakers when they develop policies for the governance of AI. Regulations should appreciate that explainability entails legitimate and diverse cost and offer direction for sharing obligations in ways that are both possible due to capacity restrictions; untargeted requirements for "full transparency" have the prospect of turning into a counter-construct when they surpass capability. Policymakers can incentivize verifiable audit systems, for instance if through safe harbors or reduced compliance burden or through expedited path to approval, because using cryptographically verifiable records can contribute to better accountability while reducing monitoring costs. Industry-specific standards for explainability of burden allocation - similar in spirit to capital adequacy requirements - could encourage consistency and help reduce compliance uncertainty; and the model of allocation offers a basis to develop such standards. Finally, regulations should also explicitly endorse multi-stakeholder accountability frameworks (vs. putting the onus on a single actor) and consideration of the transparency-privacy trade-off by lending support to the hybrid architectures (on-chain hashes plus off-chain details) along with privacy-preserving cryptography.

Several limitations are offered to limit generalization. The model is in a state of conceptual-theoretical instead of empirically validated, and the mathematical formulation and the calibration of the parameters both need testing and refinement in actual organizational implementation. Measurement is difficult organizations that might not have accounting systems to quantify the burden components (generation, maintenance, verification, communication, coordination), making calibration and evaluation difficult. The framework has assumed stable roles and capacities of stakeholders, while dynamic AI environments may require frequent reallocation that may undermine stability benefits. Technical constraints such as scalability, privacy and energy

consumption have yet to be addressed and may overturn challenges in high-volume settings. Parameter values (weights, thresholds, fairness tolerance) appear to differ across industries and cultures, implying need for the cross contextual validation. Finally, the model annotates good-faith behavior but as strategic underreporting of capacity or poor-quality explanations may happen, there may be a suggestion for complementary enforcement and incentive mechanisms.

7. Conclusion and Future Research Directions

This article has built an extensive theoretical framework for an explainability burden allocation among organizational stakeholders in artificial intelligence (AI) decision-making systems using blockchain technology as the governance infrastructure. By combining algorithmic accountability theory, institutional theory and distributed governance theory, the research helps to fill an important gap in literature regarding the systematic models of how obligation on generating, maintaining, verifying and communicating AI explanations should be distributed in multi-actor organizational settings.

The paper has five contributions which cumulatively contribute to both the theory and implementable governance design. First, it conceptualizes explainability burden as a measurable organizational resource that consists of different cost elements rather than an unspecified or technical requirement. Second, it framed stakeholder responsibilities and layers of responsibility in AI governance to instill accountability chains and make the chain of responsibility auditable and explicit. Third, it constructs a mathematical optimization model for the allocation of burden with respect to the constraints of fairness and capacity to respond, which makes it possible to systematically evaluate the feasible responsibility distributions. Fourth, it proposes blockchain architecture that can support the implementation of allocation via smart contracts and unaltered audit trial. Fifth, it causes testable propositions which join blockchain-based governance with accountability, trust, compliance, learning, and legitimate conclusions.

Beyond conceptual and formal contributions, the framework has its pragmatic value for organizations when dealing with a globally intricate regulatory environment that encompasses the EU AI Act, GDPR and sector-specific demands. By offering operational protocols, technology selection requirements, criteria for framing cost-benefit considerations and guidance on sectoral application, the paper helps to guide practitioners on how to translate theoretical knowledge into implementable governance systems with a level of enforceability over the long term and over an organization's stakeholder boundaries.

Future Research Directions:

Several research avenues can build on, challenge and expand the framework. A priority is an empirical is validation. Field research passed experiments should test the five propositions and investigate whether the model because of predictions is borne under actual organizational constraints. Longitudinal designs that compare organizations that have and do not have explainability governance via blockchain would be particularly valuable in seeking to identify effectiveness and boundary conditions. Closely related, future work will include developing rigorous approaches for parameter estimation and calibration (including ways of measuring stakeholder attributes and allocation settings across contexts (and where suitable data exists, machine learning approaches could be used to infer robust parameter values from historical decision, audit and compliance records).

A second stream is that of dynamic and quality-sensitive extensions. Current model is specified for a relatively stable environment; expanding it to environments where capacities, risk exposures, and governance requirements change over time would be very helpful for enhancing practical relevance, possibly with the use of adaptive policies. In addition, the framework is responsible but does not directly model explanation quality. Future studies should include explicit quality metrics

(e.g., accuracy, completeness, and comprehensibility) in the allocation objective so that there can be a joint optimization in the burden distribution and explanation performance. Work that provides overlap between allocation and fairness and bias mitigation interventions would provide further support for an integrated responsible-AI governance approach.

A third set of directions deals with context, choice of technology and institutions. Cross-cultural research can help focus on the extent to which such perceptions of "fair" and "legitimate" allocation of burden are influenced by institutional and cultural factors. Comparative studies must also consider design trade-offs crosswise alternative forms of blockchain architecture (public vs private, proof-of-work vs proof-of-stake), and non-block-chain forms of distributed ledger design, to understand the implications from these regarding transparency, privacy and scalability, and the operational complexity.

Finally, the areas to be discussed in the study are regulation, adoption and incentives. Regulatory impact studies can be used to measure the impact on compliance behavior of prescriptive rules versus principles-based standards under the burden of proof. Complementary stakeholder studies should examine the issues such as barriers to adoption, trust-building and legitimacy dynamics among the developers, process owners and affected individuals, and among them and regulators. Deeper economic analysis - especially game-theoretic and mechanism-design - can deal with strategic behavior - e.g., the misreporting of capacity or the low effort of providing explanations. As organizational dependence on AI only grows, mechanisms of accountability that are enforceable will increase in salience. Blockchain-based burden allocation is an interesting direction - but ultimately, it will depend on further theoretical refinement, careful design and strong empirical evaluation to be valuable.

References

Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science, 22*(5), 1123-1137. https://doi.org/10.1287/orsc.1100.0621

Asif, R., Hassan, S. R., & Parr, G. (2023). Integrating a blockchain-based governance framework for responsible AI. *Future Internet, 15*(3), 97, 2-21. https://doi.org/10.3390/fi15030097

Atzei, N., Bartoletti, M., & Cimoli, T. (2017). A survey of attacks on Ethereum smart contracts (SoK). In M. Maffei & M. Ryan (Eds.), *Principles of security and trust. POST 2017* (Lecture Notes in Computer Science, Vol. 10204, pp. 164-186). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-54455-6_8

Banerjee, G., Dhar, S., Roy, S., Syed, R., & Das, A. (2024). Explainability and transparency in designing responsible AI applications in the enterprise. In N. Naik, P. Jenkins, S. Prajapat, & P. Grace (Eds.), *Contributions presented at The International Conference on Computing, Communication, Cybersecurity and AI, July 3-4, 2024, London, UK. C3AI 2024* (Lecture Notes in Networks and Systems, Vol. 884, pp. 420-431). Springer, Cham. https://doi.org/10.1007/978-3-031-74443-3_25

Beck, R., Müller-Bloch, C., & King, J. L. (2018). Governance in the blockchain economy: A framework and research agenda. *Journal of the Association for Information Systems, 19*(10). https://doi.org/10.17705/1jais.00518

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*, 447-468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Buterin, V. (2014). *A next-generation smart contract and decentralized application platform. Ethereum White Paper.* https://ethereum.org/en/whitepaper/

Butt, U. A., Amin, R., Aldabbas, H., Mehmood, M., Shaukat, M. W., & Raza, S. M. (2023). Deploying blockchains to simplify AI algorithm auditing. In *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bahrain* (pp. 1-6). https://doi.org/10.1109/ICETAS59148.2023.10346420

Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the Internet of Things. *IEEE Access, 4*, 2292-2303. https://doi.org/10.1109/ACCESS.2016.2566339

Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., Sirer, E. G., Song, D., & Wattenhofer, R. (2016). On scaling decentralized blockchains. In J. Clark, S. Meiklejohn, P. Ryan, D. Wallach, M. Brenner, & K. Rohloff (Eds.), *Financial cryptography and data security. FC 2016* (Lecture Notes in Computer Science, Vol. 9604, pp. 106-125). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-53357-4_8

De Bruijn, H., Warnier, M., & Janssen, M. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly, 39*(2), 101666, 1-8. https://doi.org/10.1016/j.giq.2021.101666

De Filippi, P., & Loveluck, B. (2016). The invisible politics of Bitcoin: Governance crisis of a decentralized infrastructure. *Internet Policy Review, 5*(3), 1-28. https://doi.org/10.14763/2016.3.427

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56-62. https://doi.org/10.1145/2844110

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147-160. https://doi.org/10.2307/2095101

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Towards a reflective sociotechnical approach. In C. Stephanidis, M. Kurosu, H. Degen, & L. Reinerman-Jones (Eds.), *HCI International 2020—Late breaking papers: Multimodality and intelligence. HCII 2020* (Lecture Notes in Computer Science, Vol. 12424, pp. 449-466). Springer, Cham. https://doi.org/10.1007/978-3-030-60117-1_33

European Commission. (2021). *Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act).* COM(2021) 206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Parliament & Council. (2016). *Regulation (EU) 2016/679 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*, L119/1. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86-92. https://doi.org/10.1145/3458723

Gerlings, J. (2025). *The relevance of explainable artificial intelligence (xAI) in high-risk decisions.* Copenhagen Business School [PhD]. PhD Series No. 35.2025 https://doi.org/10.22439/phd.35.2025

Jarsania, P., Kumar, S., & Patel, R. (2025). TranspareGov-AI: A multi-stakeholder framework for auditable algorithmic decision-making in business processes. In *2025 IEEE International Conference on Artificial Intelligence for Learning and Optimization (ICoAILO), Bali, Indonesia* (pp. 332-337). https://doi.org/10.1109/ICoAILO66760.2025.11156056

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*, 633-705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

Lage, I., Chen, E., He, J., Narayanan, M., Kim, B., Gershman, S. J., & Doshi-Velez, F. (2019). Human evaluation of models built for interpretability. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7*(1), 59-67. https://doi.org/10.1609/hcomp.v7i1.5280

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM, 61*(10), 36-43. https://doi.org/10.1145/3233231

Lumineau, F., Wang, W., & Schilke, O. (2021). Blockchain governance—A new way of organizing collaborations? *Organization Science, 32*(2), 500-521. https://doi.org/10.1287/orsc.2020.1379

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765-4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system.* Bitcoin.org (pp. 1-9). https://bitcoin.org/bitcoin.pdf

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics, 2*(1), 25-42. https://doi.org/10.1007/BF02639315

Ostrom, E. (2010). Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review, 100*(3), 641-672. https://doi.org/10.1257/aer.100.3.641

Parlak, B. (2025). Blockchain-assisted explainable decision traces (BAXDT): An approach for transparency and accountability in artificial intelligence systems. *Knowledge-Based Systems, 307*, 114402, 1-17. https://doi.org/10.1016/j.knosys.2025.114402

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT'20).* Association for Computing Machinery, New York, NY, USA (pp. 33-44). https://doi.org/10.1145/3351095.3372873

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16).* Association for Computing Machinery, New York, NY, USA (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206-215. https://doi.org/10.1038/s42256-019-0048-x

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE Publications.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems, 28*, 2503-2511. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA* (pp. 3-18). https://doi.org/10.1109/SP.2017.41

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *The Academy of Management Review, 20*(3), 571-610. https://doi.org/10.2307/258788

Szabo, N. (1997). Formalizing and securing relationships on public networks. *First Monday, 2*(9). https://doi.org/10.5210/fm.v2i9.548

Tramèr, F., Zhang, F., Juels, A., Reiter, M. K., & Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. In *25th USENIX Security Symposium* (pp. 601-618). USENIX Association. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law, 7*(2), 76-99. https://doi.org/10.1093/idpl/ipx005

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA (pp. 1-18). https://doi.org/10.1145/3351095.3372833