B1 ISSN: 2595-1661

ARTIGO

Listas de conteúdos disponíveis em Portal de Periódicos CAPES

Revista JRG de Estudos Acadêmicos

Página da revista:
https://revistajrg.com/index.php/jrg

ISSN: 2595-1661
Revista JRG de Estudos Acadêmicos

The artificial intelligence (AI) governance dilemma

O dilema da governança da inteligência artificial (IA)

DOI: 10.55892/jrg.v9i20.3182
ARK: 57118/JRG.v9i20.3182

Recebido: 07/04/2026 | Aceito: 17/04/2026 | Publicado *on-line*: 18/04/2026

Alessandro Aveni[1]
https://orcid.org/0000-0001-6266-6818
http://lattes.cnpq.br/0679425851663633
Universidade de Brasília, UnB, DF, Brasil
E-mail: alessandro@unb.br

Abstract

The rapid diffusion of Artificial Intelligence (AI) across organizational and societal domains has exposed a tension between innovation, technological capability, and institutional responsibility. While AI systems are often discussed in terms of efficiency, innovation, or risk, less attention has been devoted to the structural misalignment between those who design, control, deploy, and are affected by these systems. This paper develops a theoretically grounded interpretation of AI governance as a problem of distributed agency. Building on Weber, Luhmann, Foucault, and Arendt, it argues that contemporary AI systems cannot be adequately governed through traditional, actor-centric, or Public Administration bureaucracy frameworks. Instead, they require a shift toward system-level accountability. The analysis highlights how current regulatory approaches, like the European Union's AI Act, only partially address this transformation. The paper concludes by proposing a conceptual model of the ownership–control–responsibility gap and outlining implications for future governance architectures.

Keywords: Artificial Intelligence; Governance; Responsibility; Accountability; EU AI Act.

Resumo

A rápida difusão da Inteligência Artificial (IA) em domínios organizacionais e sociais expôs uma crescente tensão entre a inovação tecnológica e a responsabilidade institucional. Embora os sistemas de IA sejam frequentemente discutidos em termos de eficiência, inovação ou risco, menos atenção tem sido dedicada ao desalinhamento estrutural entre aqueles que projetam, controlam, implementam e são afetados por esses sistemas. Este artigo desenvolve uma interpretação teoricamente fundamentada da governança da IA , um

[1] Bacharel em Administração e Mestre em Geografia pela Universidade de Brasília-UnB, Doutor em Ciências Políticas pela *Universidade Statale de Milano* e em Administração pela *Universidade Cormerciale Luigi Bocconi di Milano* ambas na Itália. Possui também Especialização em Estratégia Empresarial pela Fundação Getúlio Vargas - FGV. Atualmente é Professor de Gestão do Terceiro setor da faculdade Processus, de Empreendedorismo no Centro de Apoio ao Desenvolvimento Tecnológico - CDT/UnB, onde atua também no ensino de Graduação e Pós-Graduação no Mestrado Profissional em Propriedade Intelectual e Transferência de Tecnologia para Inovação – PPG PRONIT/UnB. Em 2022 foi contratado para o projeto 1000 expertos PNRR. Trabalha como consultor na *Regione Molise (Italia)* para transformação digital e racionalização dos processos da Publica Administração.

problema de agência distribuída. Com base em Weber, Luhmann, Foucault e Arendt, argumenta-se que os sistemas de IA contemporâneos não podem ser adequadamente governados por estruturas tradicionais centradas no ator. Em vez disso, exigem uma mudança em direção à responsabilização em nível sistêmico. A análise destaca que as abordagens regulatórias atuais, como a Lei de IA da União Europeia, abordam apenas parcialmente essa transformação. O artigo conclui propondo um modelo conceitual da lacuna entre propriedade, controle e responsabilidade e delineando implicações para futuras arquiteturas de governança.

***Palavras-chave**: Inteligência Artificial, Governança; Responsabilidade; Accontability; EU AI Act.*

1. Introduction

Artificial Intelligence (AI) has transitioned from a technological tool to a pervasive infrastructure shaping economic, organizational, and social processes. Contemporary AI systems—particularly machine learning and generative models—operate through distributed data, adaptive algorithms, and complex human-machine interactions. This evolution challenges traditional governance frameworks rooted in identifiable agents and clear lines of responsibility.

Existing scholarship and regulatory frameworks predominantly conceptualize Information and Communication Technologies (ICT) governance, such as technical and legal definitions like risk management, compliance, accountability, and liability. That regulatory frameworks implicitly assume that AI systems are assistive tools, embedded within identifiable human or organizational decision chains, with defined property and responsibility.

The paper research challenges that assumption. We guess that contemporary AI systems, as organizational decision-making tools, infrastructures, and ICT, whose autonomy and distributed agency, exceed the speech of regulated and uninformed political and legislative governance. Mostly, public actors underestimate innovations' ethical and operational issues, especially AI ones, such as risk management, compliance, accountability, and liability.

Classical organizational and governance theory assumes that decision-making authority is aligned within hierarchy (WEBER, 1978). However, AI systems introduce autonomous or semi-autonomous processes that evolve over time and across institutional boundaries. This paper addresses a central question: *Who governs AI systems when no actor fully controls them?* We argue that AI governance must be understood as a structural governance problem involving distributed agency (when different agents and different organizations are collaborating rather than a central agency often employing an asynchronous workflow**)** and innovation/AI management rather than merely a regulatory or ethical issue.

The AI governance dilemma discussed in this paper arises when organizations and institutions simultaneously act as AI owners, system designers, risk assessors, and primary sources of information for regulators, while regulators depend on these disclosures to exercise oversight.

Within this governance system, responsibility is formally assigned yet substantively diluted, producing a condition of institutional accountability without effective control, responsibility identification, and punishment. That shifts the focus of AI governance from questions of transparency and explainability toward power asymmetries, information dependency, and collective responsibility.

The methodological process integrates normative–analytical reasoning to connect the structural model to accountability, individual protection, and democratic legitimacy. This step is essential to the paper's theoretical contribution, as it shows that the governance dilemma is not only descriptive but normatively consequential.

2. Methodology

This study adopts a conceptual and theoretical methodology. It integrates interdisciplinary perspectives from sociology, political theory, and AI research to construct a coherent analytical framework. Rather than empirical testing, the objective is theory-building and conceptual clarification.

That approach combines:

- Classical sociological theory (Weber, Luhmann)
- Political philosophy (Arendt, Foucault)
- Contemporary AI governance literature

The research uses interdisciplinary theory on AI governance. This modeling shows governance failures result from interactions between organizational ownership, internal oversight, and distributed algorithms.

Thus, the methodology supports, as a result, a theoretical contribution. The formulation of the AI governance dilemma. That is a structural feature of organizational AI ownership rather than a contingent regulatory failure.

3. Discussion

The evidence of failures of AI governance is a fact. Nobody can deny the existence of call centers or a third party not responding to our concerns for a product. Moreover, the Control Public Agency doesn't defend consumers, but corporations, billionaire contracts of technologies are signed by the government without sharing the decision, and over people's decision sovereignty, and predatory applications use your identity and sell your information to marketing and criminals to cheat people.

Thus, AI governance failure is a predictable outcome of organizational ownership under existing legal and bureaucratic regimes. To understand the origin of governance in an AI system, a brief summary of the AI systems ecosystem and typologies can be helpful. Early AI systems were predominantly reactive, operating without internal state or learning capabilities. Examples include rule-based expert systems and game-playing machines such as IBM's Deep Blue, which relied on exhaustive search and handcrafted evaluation functions rather than learning or adaptation (Campbell et al., 2002; Russell and Norvig, 2020; Faria C. L, and Aveni, 2024)

3.1 Governance difference between AI types.

There are different types of AI systems. Several authors distinguish AI systems by their level of generality. Most existing systems fall under Artificial Narrow Intelligence (ANI), optimized for tasks. Artificial General Intelligence (AGI), which would exhibit domain-general cognitive abilities comparable to humans, remains hypothetical, while Artificial Superintelligence (ASI) represents a speculative future stage (Goertzel, 2014; Bostrom, 2014). These distinctions are essential for framing long-term theoretical and governance debates.

Contemporary applications could be defined as predictive or discriminative models. These systems focus on classification, regression, or probabilistic estimation,

forming the backbone of statistical machine learning. Foundational contributions in this area include probabilistic modeling and supervised learning frameworks (Bishop, 2006; Hastie et al., 2009; Murphy, 2012). Unlike generative models, predictive AI infers latent structure or future outcomes from existing observations.

Beyond single-agent settings, multi-agent systems investigate coordination, cooperation, and competition among multiple autonomous entities (Wooldridge, 2009). Agentic AI focuses on systems that perceive the environment, select actions, and optimize behavior. Reinforcement learning (RL) provides the dominant formal framework for such systems, modeling decision-making as a Markov Decision Process (MDP) (Sutton and Barto, 2018). B. In contrast to generative AI, agentic systems are characterized by autonomy, persistence of internal state, and feedback loops, raising distinct challenges in safety, verification, and governance (Russell, 1997).

Generative AI represents a shift from inference to data synthesis. The introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. (2014) marked a turning point, enabling the generation of realistic synthetic data through adversarial training. Subsequent advances include variational autoencoders (Kingma and Welling, 2014) and diffusion-based models (Ho et al., 2020), which improved training stability and output quality.

In parallel, large language models (LLMs) based on transformer architectures demonstrated strong generative capacities across multiple domains, including text, code, and multimodal content (Radford et al., 2019). Highly expressive, generative models are typically non-agentic. They lack persistent goals, autonomous decision-making, and continuous interaction with an environment.

Key characteristics of AI systems also concern the degree of human involvement. Assistive AI systems operate under human supervision and are designed to support, rather than replace, human decision-making (Floridi et al., 2018; Amershi et al., 2019). In contrast, autonomous AI systems act with minimal or no human control, as in robotics, algorithmic trading, and autonomous vehicles. They raise significant ethical and safety concerns (Lin et al., 2011). Finally, multi-agent and collective AI systems extend autonomy to populations of interacting agents, providing powerful tools for modeling social, economic, and organizational dynamics (Axelrod, 1997; Wooldridge & Jennings, 1995).

Another type of AI system is the representational perspective. Symbolic AI (GOFAI) emphasizes explicit rules and logical reasoning (Newell & Simon, 1976; McCarthy, 1969), offering transparency but limited scalability. Connectionist approaches, including deep neural networks, prioritize distributed representations and statistical learning (Rumelhart & McClelland, 1986; LeCun et al., 2015). More recently, neuro-symbolic approaches have emerged as an attempt to integrate learning and reasoning. They are combining the flexibility of neural models with the structure of symbolic representations (Garcez et al., 2009, 2019). Following in Table 1, these types are summarized to show typical legal uses and challenges

Table 1 - AI System Types

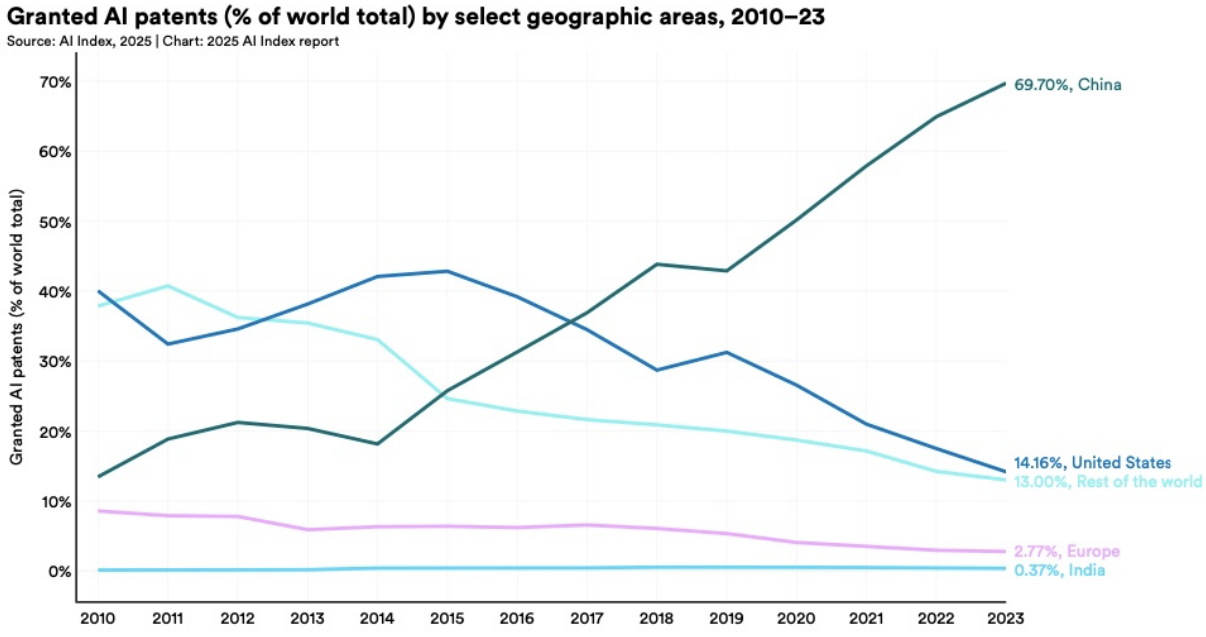
AI Category (Governance Lens)	Defining Features	Typical Public / Legal Uses	Key Governance Challenges	Core Academic References

Reactive & Rule-Based AI	Deterministic, no learning, rule-driven decisions	Tax rules engines, eligibility checks, legal expert systems	Rigidity, bias embedded in rules, responsibility attribution to designers	McCarthy (1969); Newell & Simon (1976); Campbell et al. (2002)
Predictive / Discriminative AI	Statistical inference, risk scoring, classification	Welfare fraud detection, policing risk tools, credit & benefits assessment	Opacity, statistical discrimination, due process and contestability	Bishop (2006); Hastie et al. (2009); Murphy (2012)
Generative AI (Non-Agentic)	Content and data synthesis; no autonomous goals	Drafting legal texts, administrative reports, citizen services	Hallucinations, reliability, authorship, liability for outputs	Goodfellow et al. (2014); Kingma & Welling (2014); Ho et al. (2020)
Large Language Models (LLMs)	General-purpose text and reasoning generation	Regulatory drafting support, legal research, public communication	Accountability gaps, explainability, misuse in decision justification	Radford et al. (2019); Russell & Norvig (2020)
Agentic AI (Single-Agent)	Autonomous goal pursuit, feedback loops, learning over time	Automated enforcement, adaptive resource allocation, robotics	Control, alignment, legal responsibility, auditability	Russell (1997); Sutton & Barto (2018)
Multi-Agent & Collective AI	Interacting autonomous agents; emergent outcomes	Traffic systems, market simulations, policy modeling	Unpredictability, systemic risk, collective responsibility	Axelrod (1997); Wooldridge (2009)
Symbolic AI	Explicit logic and rules; interpretable reasoning	Legal reasoning engines, compliance verification	Limited scalability; rule maintenance	McCarthy (1969); Newell & Simon (1976)
Connectionist / Deep Learning AI	Opaque statistical representations; high performance	Image recognition, surveillance, biometric systems	Explainability deficits, fundamental rights risks	Rumelhart & McClelland (1986); LeCun et al. (2015)
Neuro-Symbolic AI	Hybrid learning + reasoning	Explainable decision support, regulated AI systems	Technical complexity; verification standards	Garcez et al. (2009); Garcez et al. (2019)
Assistive AI (Human-in-the-Loop)	Human oversight retained; advisory role	Administrative decision support, judicial assistance	Automation bias, unclear responsibility boundaries	Floridi et al. (2018); Amershi et al. (2019)
Autonomous AI	Minimal or no human oversight	Autonomous vehicles, algorithmic trading, defense systems	Liability gaps, safety, constitutional accountability	Lin et al. (2011); Amodei et al. (2016)
ANI / AGI / ASI (Generality Axis)	Narrow vs general vs superhuman capabilities	Policy foresight and long-term regulation	Anticipatory governance, existential and systemic risks	Goertzel (2014); Bostrom (2014)

Source: the author Alessandro Aveni. Alessandro@unb.br

The differentiation makes it difficult to clarify who owns and who operates the AI system, and, consequently, who is responsible for it. An AI system is not only a tool or a license but many. The consequence is that, while no global registry and indexes exist in WIPO (Aveni 2024), research evidence suggests that large language models, foundation models, agentic and autonomous systems, AI used in public administration, finance, healthcare, and defense are owned by private organizations (corporate, public, institutional). That is the only conclusion if we base our logic on the following Research report chart (Stanford 2025).

Graph 1 - Granted AI patents



Source Stanford University AI index report 2025 (STANFORD, 2025, pg.44)

One can also infer that if the main AI registration patent belongs to China (70%) and the USA (14%), because of the economic system of the two countries, there is a high probability that there is no individual production and commercialization of the AI systems alone. It is probable that AI systems are rented or purchased by the Government of China and big corporations in the USA to assemble a package to be sold to other corporations or the Public Administration to scale the system.

According to Index Report 2025 (Stanford 2025, p. 47), the proportion of notable AI models originating from industry has steadily increased over the past decade, growing to 90.2% in 2024. the top contributors were Google(7),OpenAI(7models), and Alibaba(6). Since 2014, the number of notable AI models by organization, 2024. Google has led with 187 notable models, followed by Meta (82) and Microsoft (39). Among academic institutions, Carnegie Mellon University (25), Stanford University (25), and Tsinghua University (22) have been the most prolific since 2014 (Stanford 2025 p.50).

Moreover, in 2024, most AI models were released via API access (32.8%) and an open-source hybrid system (AVENI 2025a). Thus, another issue to measure ownership is to define the ownership of the components of the AI system, notably the training and data subset. That is completely owned by the corporation or other institutions. The Api and open access are allowed, but must be maintained.

Ownership, in the actual property framework system, always depends on the register. Because of the complexity of the AI systems framework, there is no possibility for individual owners to influence the AI systems used all over the internet. Thus, the ownership hypothesis of AI systems from Organizations and Institutions, and the purchase or rent of AI systems by PA, must be validated as a milestone of the present paper to be trusted. To identify organizations and institutions as the only possible governance actors and responsible for AI systems is the first step in the discussion of AI governance.

Following the discussion above on types and property,AI systems differ significantly in structure and governance implications. Table 1 summarizes key categories.

Table 2. AI system typologies and governance challenges

Type	Characteristics	Governance Issue
Rule-based systems	Deterministic, explicit logic	Clear accountability
Machine learning	Data-driven, adaptive	Opaque decision-making
Generative AI	Produces novel outputs	Ownership ambiguity
Multi-agent systems	Distributed interactions	Coordination failure
Autonomous systems	Independent operation	Responsibility gaps

Source the author. Alessandro Aveni. Alessandro@unb.

Moreover, the shift from deterministic to probabilistic and generative systems increases uncertainty and reduces traceability, complicating governance.

3.2. Theoretical Framework of actual AI Governance.

AI governance is the structured effort to ensure that autonomous, adaptive, and opaque AI systems operate safely, ethically, and in accordance with human and institutional objectives, while establishing clear channels of responsibility and accountability. Some classic theoretical positions about governance are listed below.

a) **Weberian Perspective**

Weber's theory of bureaucracy emphasizes rational-legal authority and hierarchical accountability (Weber, 1978). AI systems challenge this model by decentralizing decision processes and reducing direct human oversight.

As AI systems become agentic and autonomous, this attribution becomes increasingly problematic. Decision-making is no longer exclusively located within a human office or role, but distributed across algorithmic processes that are neither legal persons nor morally accountable. This development intensifies what Weber described as the "iron

cage" of rationalization, where procedural efficiency expands while substantive responsibility is defined but becomes diluted.

b) Luhmann's Systems Theory

Luhmann conceptualizes society as composed of self-referential systems operating through communication (Luhmann, 1995). AI can be interpreted as a subsystem that processes information autonomously, complicating external control.

From this viewpoint, the widespread adoption of AI contributes to a shift from responsibility to decision absorption: decisions are stabilized and legitimized by systems, while the question of "who is responsible" becomes secondary or even irrelevant. Agentic AI thus accelerates a structural tendency toward governance without accountable subjects, where responsibility is diffused across technical, organizational, and legal subsystems.

c) Foucault and Govern-mentality

Foucault's notion of governmentality highlights how power operates through dispersed mechanisms rather than centralized authority (Foucault, 1991, 2007). AI systems exemplify this shift, embedding governance within technical infrastructures.

Following Weber Luhmann, Foucault, the AI governance shifts responsibility to the control process and compliance. Agentic AI systems intensify this dynamic by enabling automated interventions in real time, often without transparency or contestability. Responsibility is displaced from visible authority to opaque technical digital applications, reinforcing what has been described as algorithmic or digital governmentality.

d) Arendt and the root of Responsibility

Arendt's analysis of responsibility emphasizes the ethical limits of bureaucratic systems (Arendt, 1958, 1963). AI extends these limits by diffusing agency across human and non-human actors. For Arendt, responsibility arises from human action in a shared public space, grounded in judgment and plurality. AI autonomous and agentic systems do not act; they execute processes without judgment, intention, or moral accountability. The delegation of decisions to AI systems, therefore, risks producing what Arendt would characterize as a dilemma or responsibility without actors. All outcome effects are real, but no one can be held accountable.

Comparing the analysis, one can perceive that the actual system accountability becomes procedural (regulations without operative implementations) rather than substantive; there is liability risk fragmentation across developers, deployers, and users, and human oversight may become symbolic, especially for agentic AI. This is not accidental, but it is a structural feature of organizational ownership of AI. Moreover, ownership is not accountability. Formal ownership by a legal entity does not guarantee effective control, meaningful oversight, moral or political responsibility. Generative and Agentic AI require System-Level Governance (SLG).

In other words, individual fault attribution is insufficient for cumulative harms, automated action chains, and multi-agent environments. When organizations deploy AI, define compliance, and supply audit information. The regulation risks most advanced proposed by the AI EU act, becoming self-referential, not corrective (Aveni 2025b, Campos Faria and Aveni 2024b).

Thus, there is a structural mismatch between contemporary AI systems and institutions' responsibility. Addressing this mismatch requires not only technical

safeguards (control risks) but a rethinking of governance paradigms capable of operating in contexts of distributed, non-human decision-making. That could be defined by a regulation that will encompass all the issues. In AI-mediated organizations, governance can no longer rely on hierarchical accountability models but must be understood as an emergent property of sociotechnical systems, where responsibility shifts from individual agents to institutional design, system architecture, and upstream decision-making (Aveni 2025c).

3.3. Example of governance failure frameworks. The Eu AI act.

"Who decided? How does control, influence, and responsibility emerge across humans and machines over time?", among many, Weber's theory could be used for formal authority and legitimacy, Foucault to discuss power as behavioral modulation, and Arendt to discuss responsibility and judgment. However, according to (Aveni and De Carvalho 2017), the Collingridge Dilemma (Technology Governance) underlines temporal and control paradoxes[2].

All the cited theories implied an architectural governance (design constraints), institutional responsibility (organizational duty, not individual blame), and adaptive oversight (continuous monitoring and recalibration). Taken together, these perspectives suggest that different types of AI correspond and could be summarized as three main issues to distinct governance problems: 1) Predictive and assistive AI challenge responsibility at the procedural level (Weber), but remain largely compatible with existing institutional frameworks (Floridi), 2) Generative AI complicates epistemic responsibility, as meaning and content are produced without clear authorship (Foucault) 3) Agentic and autonomous AI pose the most radical challenge, undermining the very conditions under which responsibility, judgment, and accountability have traditionally been conceived (Luhmann, Foucault, Arendt).

The European Union's regulation system control, compliance, and liability has emerged as a global leader in AI regulation. The European Union proposed the Artificial Intelligence Act (EU AI Act), which adopts a risk-based approach to AI governance. While the EU Act and the other regulations bound to it provide a comprehensive regulatory framework, its underlying assumptions about responsibility, control, and accountability remain implicitly grounded in modern institutional models that precede contemporary developments in agentic and generative AI.

The EU AI Act largely reflects a Weberian rational-legal model of governance in which responsibility is formally assigned to identifiable actors (providers, deployers, and users), embedded within bureaucratic and legal hierarchies. High-risk AI systems are subject to ex ante conformity assessments, documentation requirements, and post-market monitoring, reinforcing procedural rationality and traceability.

This approach is well-suited to predictive, assistive, and decision-support AI, where human oversight remains meaningful, and responsibility can be attributed to specific organizational roles. The bureaucratic model has changed over time, but it is the fundamental structure of actual Public Administration; however, as AI systems become increasingly autonomous and agentic, Weberian accountability encounters its limits. The decision-making processes are no longer fully transparent, nor reducible to discrete human actions. That depends on who rules and how the bureaucratic system is ruled. The result we guess is a growing, deep gap between formal responsibility and effective control of the individual, and society suffers.

[2] In general terms : "technology changes exponentially, but social, economic, and legal systems change incrementally."

From a Luhmannian standpoint, the EU AI Act can be interpreted as an attempt to stabilize expectations within functionally differentiated systems (law, economy, administration). It doesn't restore moral responsibility, compliance mechanisms, risk classifications, and technical standards function. But it is a tool of decision absorption. The scope of that type of governance allows institutions to continue operating despite the opacity and complexity of AI systems.

In this sense, accountability under the EU AI Act is less about identifying responsible subjects and more about ensuring systemic continuity. Agentic AI systems intensify their dynamic by embedding decision-making capacities directly into technical infrastructures, distancing outcomes from individual accountability. Liability regimes risk, even when formally present, becoming symbolic rather than corrective.

Foucault's concept of governmentality sheds light on the risk-based logic at the core of the EU AI Act. The EU AI Regulation classifies systems according to their potentially risky impact on populations, emphasizing control, standardization, and continuous improvements. This approach aligns particularly well with predictive and generative AI, which function as management's tool through classification, profiling, and content modulation.

The focus on risk mitigation may obscure deeper power asymmetries produced by the large-scale deployment of AI, especially when agentic systems are capable of autonomous intervention. Responsibility is displaced from political decision-making to technical compliance, reinforcing a form of algorithmic governmentality where automated control replaces accountability. However, technical compliance is subject to organizational control. Today, not all organizations have clear and good accountability, compliance, and a liability-trusted system. Corruption and strong lobbies' manipulations are present in totalitarian regimes and democracies too.

Arendt's theory highlights the fundamental dilemma. The legal responsibility presupposes human judgment, yet AI systems increasingly operate without it. While the EU AI Act mandates human oversight, this requirement often functions as a formal safeguard rather than a substantive guarantee of judgment and responsibility. Involving autonomous or self-learning AI systems, the liability becomes undefined across developers, deployers, and users. That produced what Arendt described as a condition of responsibility without actors. Harmful outcomes may occur without any single agent exercising intentional action or moral judgment, challenging the very foundations of liability law.

The conclusion is that the EU AI Act represents an insufficient response to contemporary AI systems (Aveni 2025b, Campos Faria and Aveni 2024), control, accountability, and liability, or as a governance system. Future AI governance may require a shift from individual defense and actor-based liability toward system-level accountability, control, and responsibility, for instance, including Public administration, which we must remember is excluded in the EU AI Act, to have a collective responsibility, mandatory insurance schemes, and continuous institutional oversight (Aveni 2024)

4. Results of the discussion.

4.1 The AI Governance Dilemma and Solutions.

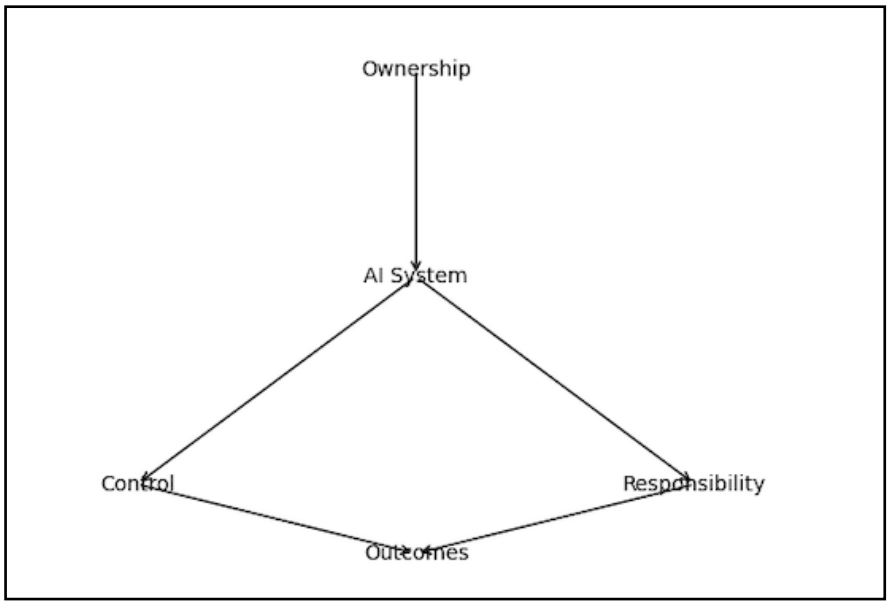
At the core of AI governance lies a structural tension that cannot be reduced to legal or ethical shortcomings. Rather, it reflects a deeper transformation in how action, decision, and consequence are distributed across socio-technical systems.

In traditional organizational settings, ownership, control, and responsibility tend to overlap or, at the very least, remain traceable. In AI-driven environments, however, these dimensions drift apart. Developers design architectures without controlling real-world deployment contexts; organizations deploy systems they only partially understand; users interact with outputs that emerge from opaque processes.

This misalignment can be conceptualized as an ownership–control–responsibility gap or a governance dilemma. Figure 2 below illustrates how ownership (legal or economic), control (technical and operational influence), and responsibility (normative and legal accountability) are not aligned along a single axis. Instead, they form a fragmented structure in which causal influence and accountability diverge.

This condition also has a constraint. It is possible to predict innovation and AI outcomes. We know, resonating with the Collingridge dilemma (Collindrige, 1980), that at early stages, intervention is possible but knowledge is limited; at later stages, knowledge increases but systems become resistant to change. In other words, the prediction of an innovation today depends on the assessment and initial predictions, but these could fail; in general, they are not assessed after the innovation spreads, and are not compared to the initial status. The ownership persists and changes the impacts all over society (Aveni and De Carvalho 2017)

Figure 1 - Actual AI governance system



Source: the author. Alessandro Aveni. Alessandro@unb.br.

A central finding emerging from the comparison of AI typologies, theory, and AI act analysis is that not all AI systems pose the same governance problem. The regulation does not cover all issues. Predictive and assistive AI systems largely conform to the assumptions embedded in current regulatory models. Their outputs can be audited, their deployment contexts are relatively stable, and human oversight remains substantively meaningful. In these cases, Weberian models of rational-legal accountability retain their validity.

Generative and agentic AI systems introduced forms of decision-making that are diffuse, opaque, and temporally extended. Generative models challenge epistemic

accountability by producing content without clear authorship, intention, or stable meaning. Agentic systems go further by autonomously selecting actions over time, thereby eroding the link between decision, actor, and responsibility. Existing liability frameworks, which presuppose identifiable agents and causal chains, struggle to accommodate these dynamics.

The risk-based logic of the EU AI Act aligns with broader trends of algorithmic governmentality. By focusing on population-level risks rather than individual acts, regulation increasingly operates through continuous monitoring, classification, and optimization. This approach is effective in managing large-scale systems but may obscure questions of power, asymmetry, and contestability. The EU AI Act places strong emphasis on formal safeguards: documentation, risk classification, conformity assessments, and human oversight requirements. While these mechanisms enhance transparency at an organizational level, the analysis suggests that they may function primarily as procedural substitutes for responsibility rather than as guarantees of effective control. The regulation needs to be completed.

From a Luhmannian perspective, these mechanisms enable institutions to absorb decision-making complexity without resolving the underlying accountability deficit. Compliance becomes a means of stabilizing expectations, not of restoring agency. This raises the risk that governance regimes may achieve legal robustness while remaining normatively fragile, particularly in high-impact applications involving autonomous AI. Agentic AI systems intensify this dynamic by enabling automated interventions that occur below the threshold of political visibility. In such contexts, responsibility is displaced from public deliberation to technical infrastructures, reinforcing a model of governance oriented toward control rather than accountability. This shift raises concerns about democratic legitimacy, especially when AI systems shape access to resources, opportunities, or rights.

According to Arendt, the delegation of decision-making to AI systems can also be interpreted as producing outcomes without judgment. Even when legal responsibility is formally assigned, the absence of human deliberation undermines the moral and political foundations of accountability. This is particularly evident in scenarios where harm emerges cumulatively or probabilistically, rather than from discrete decisions. The resulting condition—responsibility without actors—poses a fundamental challenge to liability law and regulations. Fragmentation of responsibility across developers, deployers, and users risks normalizing harm as a systemic byproduct rather than an actionable failure. In this sense, advanced AI systems expose a limit case for modern institutions of responsibility.

Thus, incremental adjustments to existing liability regimes may be insufficient. Instead, the governance of generative and agentic AI may require a shift toward system-level accountability mechanisms. These could include collective responsibility models, mandatory insurance schemes, independent oversight bodies, and continuous auditing infrastructures that operate beyond individual fault attribution.

Such approaches address harms that emerge from the actual complex sociotechnical systems. Importantly, they also align more closely with the distributed nature of agency in contemporary AI, acknowledging that responsibility can no longer be fully localized. We are demonstrating that AI governance challenges are not merely technical or legal, but structural and conceptual. The preceding analysis suggests the following operational recommendation: differentiate accountability regimes by AI typology or move beyond a uniform accountability model and explicitly differentiate

governance requirements for predictive/assistive AI,generative AI, agentic, and autonomous AI.

Table 3 - Key points by AI generative e agent systems

AI Type	Governance Characteristics	Challenges for Accountability	Theoretical Lens / Insights	Recommended Approach
Generative AI	Produces content without clear authorship, intention, or stable meaning	Epistemic accountability: authorship and intent unclear; causal chains diffuse	Luhmann: stabilizes complexity without resolving accountability; Arendt: decisions without judgment	Require system-level mechanisms: collective responsibility, continuous auditing, differentiated governance
Agentic / Autonomous AI	Autonomously selects actions over time; diffuse and temporally extended decision-making	Responsibility displaced from actors; cumulative/probabilistic harm; conventional liability insufficient	Luhmann: institutions absorb complexity without restoring responsibility; Arendt: moral and political accountability absent	System-level accountability: independent oversight, mandatory insurance, continuous monitoring; shift beyond individual fault

Source: the author Alessandro Aveni. Alessandro@unb.br

4.2. The AI Governance Limits and Structural Misalignment, toward System-Level Accountability

The European Union's AI Act represents a risk-based regulatory framework (European Parliament, 2024 European Commission, 2021,). While innovative, it retains a firm-centric logic, assigning responsibility primarily to providers and deployers.
However, this approach underestimates:

- Distributed data ecosystems
- Model reuse and fine-tuning
- Emergent behaviors in complex systems

As a result, legal responsibility does not fully align with actual causal influence.

Building on the preceding analysis, the paper advances the following propositions:

- Structural Decoupling**:**
In AI-driven socio-technical systems, ownership, control, and responsibility are structurally decoupled, leading to systematic accountability gaps that cannot be resolved within actor-centric governance frameworks.

- Distributed Agency**:**
As the degree of system autonomy and interaction increases, causal agency becomes distributed across human and non-human actors, reducing the explanatory and normative adequacy of traditional liability models.

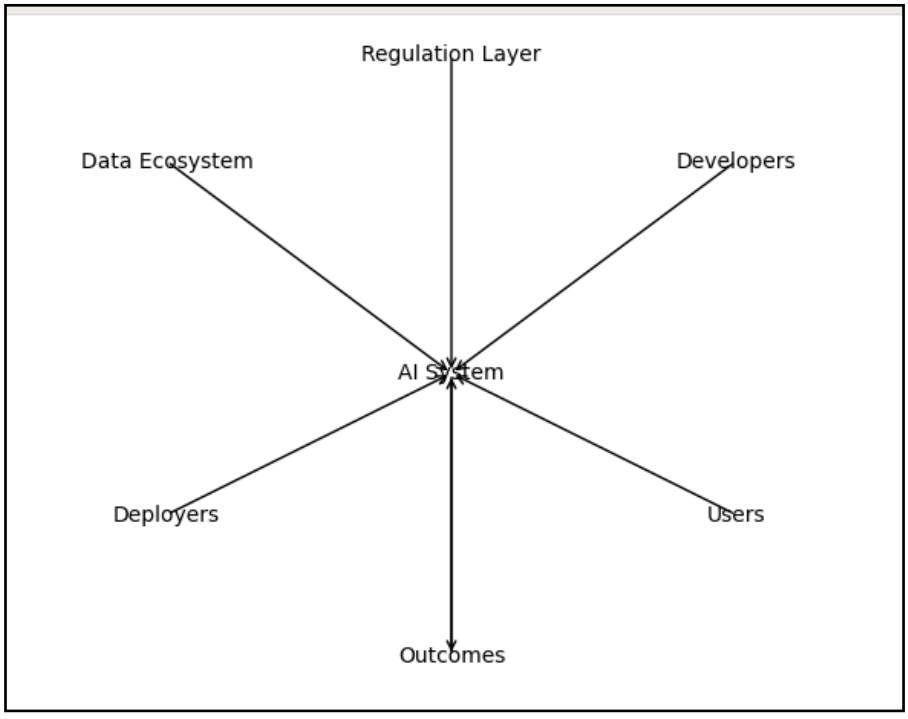
-System-Level Governance Necessity**:**
Effective governance of advanced AI systems requires a shift from individual attribution of responsibility to system-level accountability mechanisms encompassing lifecycle monitoring, collective liability, and institutional oversight.

If the governance problem is structural, solutions cannot rely exclusively on identifying a single responsible actor. What is required is a shift in perspective—from responsibility as attribution to responsibility as an emergent property of systems.

This model emphasizes three key aspects:

1. Multiplicity of actors: no single entity determines system behavior.
2. Feedback dynamics: outcomes reshape inputs (data, usage, optimization).
3. Layered governance: regulation operates as an overlay rather than a direct control mechanism**.**

Figure 2. The AI Governance New Framework



Source: The author.Alessandro Aveni. Alessandro@unb.br

From this perspective, accountability should be reinterpreted along three dimensions, namely: Ex-ante responsibility (design and training choices), In-process responsibility (monitoring and intervention), Ex-post responsibility (liability and redress)

Rather than assigning responsibility to isolated actors, governance frameworks should distribute obligations across these phases. Thus, policy design should move beyond firm-centric regulation, incorporate ecosystem-level analysis, enable adaptive governance, and promote transparency and traceability. Finally, a list of punctual implications and conflicts is presented to be solved by a new governance system.

Table 4: AI Systems Ownership, Control, and Governance Implications

Dimension	Key Points	Governance / Legal Implications
Legal vs. Actual Ownership	AI is not a legal subject; it cannot own property or bear liability. Ownership is legally attributed to a human or organizational legal entity holding economic exploitation rights (not epistemic or contributory creators).	Law requires a clearly identifiable legal subject for liability, licensing, and compliance (e.g., EU AI Act), even when real control is distributed.
Registration & Attribution	Every AI system must be registered or associated with a legal person (individual or organization).	Formal ownership enables enforcement but may mask internal responsibility fragmentation.
Empirical Ownership Distribution	Organizations own ~95–98% of modern AI systems; individuals ~2–5%, mainly open-source, prototypes, or lightweight tools.	Governance frameworks must primarily target organizational actors rather than individual developers.
Drivers of Ownership Concentration	High capital, data access, compute infrastructure; dominance of Big Tech and large research institutions.	Structural asymmetry reinforces regulatory focus on large-scale AI providers and deployers.
Advanced Models as Consolidated Assets	State-of-the-art models (e.g., GPT-4/5, Gemini, Claude) are proprietary organizational assets.	Raises issues of market power, access control, and systemic risk governance.
Ownership–Control Alignment	Owners are often also users and operators of AI systems within the same organization.	Blurs separation between internal governance and external regulatory oversight.
Governance Loop	Organizations control AI internally, while public authorities regulate externally and ex post.	Creates accountability gaps when internal roles and responsibilities are unclear or informal.
Liability under EU AI Act	Liability assumes clear hierarchical roles within organizations.	Mismatch between legal assumptions and organizational reality increases compliance risk.
Responsibility Dispersion	Decision-making distributed across teams, units, and consortia.	Amplifies "decision absorption" (Luhmann), weakening individual accountability.

Ownership as Basis of Responsibility	Ownership is a prerequisite for legal responsibility, but not sufficient for effective accountability.	Necessitates complementary governance mechanisms beyond formal ownership.
Normative Arcology (Foucauldian Lens)	Governance occurs through law plus internal compliance, contracts, and private rules.	Power and discipline extend beyond public regulation into organizational micro-governance.
Policy Option Transparency	Public registries of AI ownership and internal governance roles.	Enhances traceability and public oversight.
Policy Option Ownership Accountability Standards	Formal distinction between license holders, infrastructure owners, and decision governors.	Clarifies responsibility allocation inside complex organizations.
Policy Option Institutional Counterbalancing	Strengthened independent auditing and supervisory authorities.	Reduces conflicts of interest between AI owners, users, and regulators.

Source: the author Alessandro Aveni. alessandro@unb.br

5. Concluding Remarks

The research contributes to the emerging literature on AI governance by advancing a structural, theory-driven account of responsibility in socio-technical systems. The present study reframes governance as a problem of distributed agency and systemic misalignment.

The paper integrates classical sociological and political theory, Weber, Luhmann, Foucault, and Arendt, into a unified analytical framework for AI governance. This cross-theoretical synthesis extends existing approaches that typically rely on either normative ethics (Floridi et al., 2018) or technical safety (Amodei et al., 2016), but rarely connect them to foundational theories of organization and power.

The paper introduces the concept of the ownership–control–responsibility gap as a structural property of AI systems. Unlike existing discussions of the "accountability gap," which often remain descriptive, this formulation provides a conceptual mechanism explaining why attribution fails under conditions of distributed causality.

The paper also develops a system-level governance model that shifts the unit of analysis from individual actors to interacting components within an ecosystem. This perspective contributes to ongoing debates in platform governance and complex systems theory by emphasizing feedback loops, multi-actor interactions, and layered regulation. The model implies the formulation of three explicit theoretical propositions within the tradition of theory-building research in organization studies. By articulating testable claims regarding structural decoupling, distributed agency, and governance necessity, the study opens pathways for empirical validation and comparative institutional analysis.

Finally, the critique of the European Union's AI Act highlights a broader limitation in current regulatory paradigms: the persistence of firm-centric responsibility models in contexts where agency is inherently distributed. This argument contributes to policy-oriented literature by suggesting the need for collective and lifecycle-based accountability mechanisms. Taken together, these contributions reposition AI governance from a question of rule-setting to a problem of institutional design under conditions of systemic complexity.

AI governance represents a fundamental challenge to traditional concepts of responsibility and control. By framing AI as a distributed socio-technical system, this

paper highlights the need for new governance paradigms. Future research should further develop empirical and institutional models to operationalize system-level accountability.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., … Horvitz, E.. Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference, 1–13. 2019.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. Concrete problems in AI safety. 2016. arXiv preprint arXiv:1606.06565.

Axelrod, R.. The complexity of cooperation: Agent-based models of competition and collaboration. Princeton University Press. 1997.

Arendt, Hannah. The human condition. Chicago: University of Chicago Press, 1958.

Arendt, H. Eichmann in Jerusalem: A report on the banality of evil. New York, NY: Viking Press. 1963.

Aveni. A. Overview of AI International and Brazil regulations. Revista da Presidência v. 27 n. 142 2025b. https://revistajuridica.presidencia.gov.br/index.php/saj/article/view/3248 in https://revistajuridica.presidencia.gov.br/index.php/saj/issue/view/153

Aveni A.Free, open source, and paid programs offer. Increasing social surplus. Revista Processus de Estudos de Gestão, Jurídicos e Financeiros v. 16 n. 50 2025a. https://periodicos.processus.com.br/index.php/egjf/article/view/1370

Aveni A. Define artificial intelligence outcomes as intellectual property, a collective right. Revista Processus de Estudos de Gestão, Jurídicos e Financeiros. v.XV, p.1 - 18, 2024.

Aveni A, De Carvalho S.S.M. Avaliação de patentes e inovação DE métodos e problemas. Cad. Prospec., Salvador, v. 10, n. 4, p. 639-649, out./dez. 2017. D.O.I.:http://dx.doi.org/10.9771/cp.v10i4.23018

Bishop, C. M.. Pattern recognition and machine learning. Springer. 2006.

Bostrom, N.. Superintelligence: Paths, dangers, strategies. Oxford University Press. 2014.

Campbell, M., Hoane, A. J., & Hsu, F.. Deep Blue. Artificial Intelligence, 134(1–2), 57–83.. 2002. https://doi.org/10.1016/S0004-3702(01)00129-1

Collindrige, David. The social control of technology. London: Frances Pinter, 1980.

European Commission (EC). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). 2021. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

European Parliament and Council of the European Union (EP). Artificial Intelligence Act (final text, pending implementation). 2024

Campos Faria L. And Aveni A. Clarify Artificial Intelligence (AI) decision models' rights in the Intellectual Property (IP) system. Revista JRG de Estudos Acadêmicos. v.7, p.e141033, 2024a.

Campos Faria L. And Aveni A. Desafios e perspectivas da Inteligência artificial na análise da concorrência do Poder Público. Revista JRG de Estudos Acadêmicos. v.7, p.e141035, 2024b. https://revistajrg.com/ index.php/jrg][doi:10.55892/ jrg.v7i14.1035

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Vayena, E.AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689–707. 2018. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., & Cowls, J. A. Unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). 2019. https://doi.org/10.1162/99608f92.8cd550d1

Foucault, Michel. Security, territory, population. New York: Palgrave, 2007.

Foucault, Michel. Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), The Foucault effect: Studies in governmentality (pp. 87–104). Chicago, IL: University of Chicago Press. 1991.

Garcez, A. d., Lamb, L. C., & Gabbay, D. (2009). Neural-symbolic cognitive reasoning. Springer.

Garcez, A. d., Besold, T. R., Raedt, L. D., Földiak, P., Hitzler, P., Icard, T., … Silver, D. Neural-symbolic learning and reasoning: Contributions and challenges. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 9482–9489. 2019.

Goertzel, B. Artificial general intelligence: Concept, state of the art, and future prospects. Journal of Artificial General Intelligence, 5(1), 1–48. 2014.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y.. Generative adversarial nets. Advances in Neural Information. 2014

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer.

Ho, J., Jain, A., & Abbeel, P. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851. 2020.

Jensen, Michael; Meckling, William. Theory of the firm. Journal of Financial Economics, v. 3, p. 305–360, 1976.

Kingma, D. P., & Welling, M. Auto-encoding variational Bayes. International Conference on Learning Representations (ICLR). 2014.

Latour, Bruno. Reassembling the social. Oxford: Oxford University Press, 2005.

lacuna, Y., Bengio, Y., & Hinton, G. Deep learning. Nature, 521(7553), 436–444. 2015. https://doi.org/10.1038/nature14539

Lin, P., Abney, K., & Bekey, G. A. Robot ethics: The ethical and social implications of robotics. MIT Press. 2011.

Luhmann, Niklas. Social systems. Stanford: Stanford University Press, 1995.

McCarthy, J. Programs with common sense. Stanford University AI Laboratory. 1969.

Maslej, Nestor et al. AI Index Report 2025. Stanford: HAI, 2025.

Murphy, K. P. Machine learning: A probabilistic perspective. MIT Press. 2012.

Newell, A., & Simon, H. A.. Computer science as empirical inquiry: Symbols and search. Communications of the ACM, 19(3), 113–126. 1976. https://doi.org/10.1145/360018.360022

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I.. Language models are unsupervised multitask learners. OpenAI Technical Report. 2019.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. Parallel distributed processing: Explorations in the microstructure of cognition. MIT Press. 1986.

Russel, Stuart; Norvig, Peter. Artificial intelligence: a modern approach. 4. ed. Pearson, 2020.

Russell, S. J.. Rationality and intelligence. Artificial Intelligence, 94(1–2), 57–77. 1997. https://doi.org/10.1016/S0004-3702(97)00026-8

Stanford University. The AI Index 2025 Annual Report. By Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Toby Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak. "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025. https://doi.org/10.48550/arXiv.2504.07139

Weber, Max. Economy and society: An outline of interpretive sociology (G. Roth & C. Wittich, Eds.). Berkeley, CA: University of California Press, 1978

Sutton, R. S., & Barto, A. G. Reinforcement learning: An introduction (2nd ed.). MIT Press. 2018.

Wooldridge, M. An introduction to multiagent systems (2nd ed.). Wiley. 2009.

Wooldridge, M., & Jennings, N. R. Intelligent agents: Theory and practice. Knowledge Engineering Review, 10(2), 115–152. 1995.