Scholedge International Journal of *Business Policy & Governance*
ISSN 2394-3351 Vol. 13, Issue 02 (2026)
DOI: https://dx.doi.org/10.19085/sijbpg130201

Who Is Accountable? When AI Makes the Wrong Decision? Rethinking Corporate Governance

Ibrahim Abdullahi
University of Uyo, Uyo, Akwa-Ibom State, Nigeria

ABSTRACT

Corporate boards now sit atop decision architectures that no longer belong to them alone. Machine learning systems screen credit applications, price insurance risk, recommend mergers, flag fraud, and increasingly shape the strategic judgments that directors and executives once reached through experience and deliberation. When one of these systems produces a harmful, discriminatory, or commercially damaging outcome, existing governance doctrine struggles to identify who answers for it. This paper examines the widening gap between algorithmic decision-making and the accountability structures built for human agents. Drawing on agency, stakeholder, stewardship, institutional, and resource dependence theories, together with enterprise risk management and responsible AI scholarship, the paper argues that accountability for algorithmic harm cannot rest on a single actor or a single governance layer. Responsibility instead needs to be distributed across the people and functions that design, approve, deploy, and supervise an AI system, with each layer answerable for a distinct category of failure: design flaws, oversight lapses, deployment misjudgment, and monitoring neglect. The paper's central contribution is a multi-level Corporate AI Accountability Governance Framework that assigns differentiated responsibilities to the board, executive leadership, a dedicated AI governance committee, risk management, internal audit, technology teams, external vendors, and regulators. The framework is built around a feature that conventional governance controls were never designed to handle: AI systems continue to change after deployment, so a one-time approval cannot substitute for ongoing supervision. The paper closes with practical implications for boards preparing for algorithmic oversight, a comparative reading of regulatory expectations across the European Union, the United States, the United Kingdom, and selected Asia-Pacific economies, and a research agenda for scholars working on the next phase of digital corporate governance.

Keywords: AI governance, board accountability, algorithmic transparency, fiduciary duty, enterprise risk management, corporate digital governance

INTRODUCTION

In 2018, Amazon shelved an internal recruiting tool after discovering that it had taught itself to downgrade resumes containing the word "women's," a pattern the model had absorbed from a decade of hiring data skewed toward men. No single engineer wrote a rule that said "reject female candidates." The bias emerged from historical data, propagated through a scoring system, and reached job applicants before anyone at the company noticed. Three years later, Zillow wound down its Zestimate-based home-buying operation after an algorithm mispriced thousands of houses, forcing a write-down that eliminated roughly a quarter of the company's workforce and cost hundreds of millions of dollars. Neither failure involved a rogue employee or a single bad decision made in a boardroom. Both involved a system that had been approved, deployed, and left to run with insufficient scrutiny of how its outputs were actually being used.

These episodes are not curiosities from the margins of corporate life. They sit at the center of a question that corporate governance scholarship has been slow to confront directly: when an organization delegates part of its judgment to a machine learning system, who carries the legal, ethical, and commercial responsibility if that judgment goes wrong? Traditional governance theory answers accountability questions by tracing decisions back to identifiable humans, a chief executive who approved a strategy, a board that failed to ask the right question, a manager who ignored a warning sign. Algorithmic decision-making disrupts that tracing exercise. A model's output is the product of data selection choices made months earlier, feature engineering decisions made by a vendor, training procedures that few people inside the company can fully explain, and deployment choices made by business units under commercial pressure. Responsibility is distributed across time and across organizational boundaries in a way that older governance models did not anticipate.

The literature on AI and corporate governance has grown quickly, but it remains split across disciplines that rarely speak to one another. Legal scholars have examined whether reliance on AI recommendations satisfies or breaches directors' duty of care (Mertens, 2023; Coulson-Thomas, 2023). Information systems researchers have studied accountability as a design property of algorithmic systems, asking how interpretability and audit trails can be built into machine learning pipelines (Lehner, Ittonen, Silvola, Ström, & Wührleitner, 2022). Business ethics scholars have treated algorithms as value-laden actors whose design choices carry moral weight (Martin, 2019). Public administration researchers have proposed accountability as a relational concept involving an actor, a forum, and a set of standards against which conduct is judged (Busuioc, 2021). Each of these literatures offers a partial answer. None of them, on its own, gives boards and executives a workable map of who should do what when an AI system misfires.

This paper attempts to build that map. Rather than asking the increasingly stale question of whether AI can or should make decisions inside organizations, since in practice it already does, the paper asks how governance systems need to evolve once that dependence becomes routine. The analysis proceeds in three movements. First, it revisits the theoretical foundations of corporate governance, agency theory, stakeholder theory, stewardship theory, institutional theory, and resource dependence theory, and asks what each theory implies about accountability once decisions are partly algorithmic. Second, it examines the specific governance challenges that arise from AI-assisted decision-making: opaque models, vendor dependency, data governance failures, bias, cybersecurity exposure, and the compliance gaps that regulators in different jurisdictions are now trying to close. Third, it proposes an original governance framework that assigns distinct, overlapping, and mutually reinforcing responsibilities to eight organizational actors, from the board of directors down to the external AI vendors whose products increasingly sit inside corporate decision chains.

The argument is conceptual rather than empirical. It draws on a substantial body of recent scholarship, spanning law, information systems, business ethics, and management, together with the emerging regulatory record from the European Union's AI Act, the United States' sectoral approach, the United Kingdom's principles-based framework, and comparable initiatives in Singapore and elsewhere in the Asia-Pacific region. The intended contribution is not a checklist. It is a way of thinking about accountability that treats AI governance as a permanent feature of corporate life rather than a temporary technology problem that will settle down once the tools mature.

LITERATURE REVIEW

Scholarship on AI and corporate governance has developed along four fairly distinct tracks, and it helps to separate them before asking how they might be combined.

The first track is legal and doctrinal. Mertens (2023) offers one of the more careful treatments of what happens to board-level decision-making once AI enters the room, concluding that existing company law

categories, duty of care, duty of loyalty, business judgment protection, were built for human deliberation and do not map cleanly onto algorithmic inputs. A related strand of legal scholarship examines fiduciary duty directly, arguing that corporate officers face a genuine dilemma: failing to use available AI tools may itself become a breach of the duty of care, while over-reliance on those same tools, without adequate understanding of how they work, can breach the same duty from the opposite direction. Zhao and Gómez Fariñas (2023) extend this line of reasoning to sustainability-linked decisions, showing how algorithmic inputs complicate a board's ability to demonstrate that a decision was reached through an informed and deliberate process, one of the central tests underlying business judgment protection in most common law jurisdictions.

A second track, drawn mostly from accounting, auditing, and information systems research, treats accountability as something that has to be engineered into a system rather than assumed after the fact. Lehner et al. (2022) examine how accounting and audit functions grapple with AI-based decision-making, arguing that ethical oversight cannot be bolted onto a model after deployment; it has to be part of how the model is specified, trained, and monitored. Cobbe, Lee, and Singh (2021) push this further with a framework for what they term reviewable automated decision-making, insisting that accountability requires systems to be built so that their outputs can be reconstructed and examined after the fact, not merely explained in the abstract. This design-first perspective matters for governance scholars because it shifts part of the accountability question away from the boardroom and toward the technical teams that build and maintain these systems, a shift that traditional governance theory, focused on principals and agents at the top of the organization, does not fully accommodate.

A third track comes from business ethics and public administration, and it treats algorithms less as tools and more as actors with moral and administrative standing. Martin (2019) argues that algorithms function as structuring agents in organizational decisions, delegated certain tasks the way a human employee might be delegated a task, which means the firm bears responsibility for the values embedded in the algorithm's design even when no single person intended the resulting harm. Busuioc (2021) approaches the same problem from a public administration lens, defining accountability as a relationship between an actor and a forum in which the actor must explain and justify conduct against a set of standards, and showing why algorithmic systems complicate all three elements: the actor is often distributed across a vendor and an internal team, the forum struggles to obtain a comprehensible explanation, and the standards themselves are often unsettled. Johnson (2021) adds a useful caution here, warning against a slide toward what she calls a diffusion of responsibility, in which so many hands touch a system that no one ends up answerable for its output. Kaminski (2019), writing about the European Union's General Data Protection Regulation, describes this as a shift toward "binary governance," combining individual rights with collaborative, ongoing oversight rather than relying on either alone, a framing that anticipated much of what later AI-specific regulation would attempt.

The fourth track is regulatory and increasingly urgent. The European Union's AI Act, formally Regulation (EU) 2024/1689, introduces binding, risk-tiered obligations for organizations that develop or deploy AI systems, with the heaviest obligations, including conformity assessments, documentation, and human oversight requirements, attached to systems classified as high-risk (European Parliament and Council of the European Union, 2024). The United States has taken a different route, relying on the National Institute of Standards and Technology's AI Risk Management Framework as voluntary guidance built around four functions, govern, map, measure, and manage, that organizations are expected to weave into existing risk and compliance structures rather than treat as a stand-alone program (National Institute of Standards and Technology, 2023). The Organisation for Economic Co-operation and Development's AI Principles, first adopted in 2019 and updated in 2024, function as a values baseline that both the EU and US frameworks explicitly reference, emphasizing human-centered values, transparency, robustness, and

accountability (OECD, 2019/2024). Singapore's Personal Data Protection Commission and Infocomm Media Development Authority have developed a Model AI Governance Framework built around explainability, auditability, and traceability, an approach that several other Asia-Pacific regulators have since drawn on.

Recent management scholarship has begun to connect these tracks directly to boardroom practice. Coulson-Thomas (2023) argues that boards cannot delegate AI oversight entirely to executives, since the strategic and reputational stakes of algorithmic failure sit squarely within the board's monitoring responsibility, even as day-to-day management of AI systems properly belongs to executive teams. Work published in AIB Insights frames the challenge as one of integrating AI into board decision practices without abdicating the board's own judgment, treating risk management and compliance as a standing item for board attention rather than a one-time briefing (Algorithmic Boardroom, AIB Insights). A recent dual-dimensional framework for board-level AI integration argues that boards need to calibrate both how autonomous an AI system is allowed to be and how deeply it is structurally embedded in decision workflows, treating these as two separate dials rather than a single question of whether to adopt AI at all.

Taken together, these four tracks point toward a shared conclusion even though none of them states it explicitly: accountability for algorithmic decisions cannot be resolved by identifying a single responsible party. It has to be distributed, and the distribution has to match the actual points in a system's lifecycle where failure can originate, design, data, deployment, and ongoing monitoring. What the literature has not yet produced is a framework that translates this insight into a workable allocation of responsibility across the specific organizational functions that most companies already have in place. That is the gap this paper addresses.

THEORETICAL FOUNDATIONS

Corporate governance theory was built to answer a narrower question than the one AI now poses, but each of its major strands still has something to say once the question is expanded.

Agency theory, in its classical form, treats the corporation as a nexus of contracts between principals, typically shareholders, and agents, typically managers, whose interests may diverge (Jensen & Meckling, 1976). The theory's central concern is monitoring cost: how much it costs principals to verify that agents are acting in their interest, and what mechanisms, incentive alignment, board oversight, disclosure, reduce that cost. AI complicates agency theory in an interesting way. It is often introduced precisely to reduce monitoring costs, an algorithm can flag anomalies faster than a human auditor and more consistently than a distracted manager. Yet the same technology introduces a new layer of agency risk, since the system itself becomes an intermediary whose behavior the human agents who deployed it may not fully understand. A model can behave in ways that serve neither the shareholder nor the manager's interest, simply because it was trained on data that no longer reflects current conditions, or because it optimizes for a proxy metric that diverges from the outcome the organization actually cares about. Agency theory, applied to AI, therefore needs an additional category of cost: not just the cost of monitoring humans, but the cost of monitoring the tools that monitor humans.

Stakeholder theory offers a different lens. Freeman's (1984) foundational argument holds that the firm has obligations to a wider set of constituencies than shareholders alone, employees, customers, suppliers, communities, and that managers must balance these interests rather than optimize for one at the expense of the rest. Applied to AI governance, stakeholder theory does useful work in identifying who bears the cost when an algorithm goes wrong. The rejected loan applicant, the job candidate screened out by a biased hiring tool, the customer whose data was mishandled by a poorly governed recommendation engine, these are stakeholders whose interests rarely appear on a board agenda unless something has already gone wrong. Stakeholder theory pushes governance scholars to ask not just who is accountable inside the firm,

but to whom that accountability is owed, a question agency theory, focused inward on the principal-agent relationship, tends to underweight.

Stewardship theory sits in productive tension with agency theory. Where agency theory assumes managers require monitoring because their interests diverge from owners', stewardship theory assumes that managers, given the right structural conditions, will act as stewards of organizational value even without heavy external control (Davis, Schoorman, & Donaldson, 1997). This matters for AI governance because it suggests an alternative to pure compliance-based oversight. If technology teams and executives are structured and incentivized as stewards of the organization's long-term reputation and stakeholder trust, rather than merely as agents to be monitored, they may be more likely to flag a model's weaknesses proactively rather than waiting for an external audit to surface them. The practical governance implication is that accountability frameworks built entirely around control and sanction may crowd out the very stewardship behavior that catches problems early, a tension the framework proposed later in this paper tries to hold rather than resolve by assumption.

Institutional theory explains why companies in the same industry often adopt strikingly similar AI governance practices even before regulation forces them to. DiMaggio and Powell's (1983) account of institutional isomorphism, coercive, mimetic, and normative, describes how organizations converge on similar structures under pressure from regulators, competitors, and professional norms respectively. Much of the current wave of AI ethics committees and responsible AI principles looks like a mimetic response: companies adopt similar governance vocabulary, sometimes before they have built the underlying capability to act on it. Mittelstadt (2019) makes a related point from a different angle, warning that principle-based AI ethics frameworks tend to proliferate without producing consistent practice, precisely because organizations adopt the language of accountability for legitimacy reasons while the harder operational work lags behind. Institutional theory is useful here because it helps explain a pattern practitioners often notice anecdotally: a company's AI governance documents can look considerably more mature than its actual practice.

Resource dependence theory adds a further dimension that the other frameworks miss. Pfeffer and Salancik's (1978) argument that organizations depend on external resources, and adjust their governance structures to manage that dependence, applies with particular force to AI, since most companies do not build their own foundation models. They depend on external vendors for the core technology, and that dependence creates a governance problem that agency theory, built around internal principal-agent relationships, does not directly address. When a vendor's model changes behavior after a silent update, or when a vendor's training data turns out to embed a bias the client company never audited, the client's board may be accountable to its own shareholders and regulators for a failure that originated entirely outside the organization's walls. Resource dependence theory suggests that governance structures need to extend outward to manage vendor relationships as a genuine risk category, not merely a procurement question.

Two more contemporary frameworks complete the picture. Teece's (2007) work on dynamic capabilities, the capacity to sense change, seize opportunity, and reconfigure resources, offers a useful frame for thinking about why static, one-time AI approvals are inadequate: an organization's capability to govern AI has to be renewed continuously as models are retrained and business conditions shift, not established once and left in place. Enterprise risk management scholarship, formalized in the COSO (2017) framework, provides the operational vocabulary, risk appetite, risk tolerance, control environment, that this paper draws on when describing how AI risk should be integrated into existing risk governance structures rather than treated as a separate silo.

No single theory, on its own, produces a complete account of AI accountability. Agency theory explains why monitoring costs rise. Stakeholder theory identifies who is harmed when monitoring fails. Stewardship theory cautions against over-relying on control mechanisms alone. Institutional theory explains why governance responses can look more polished than they are. Resource dependence theory extends the accountability perimeter beyond the firm's own walls. Read together, these theories converge on a single implication that the framework in Section 7 tries to operationalize: accountability for algorithmic decisions has to be distributed across multiple organizational levels, each answerable for the type of failure closest to its own function, rather than concentrated in any one governance body.

EMERGING GOVERNANCE CHALLENGES IN AI-DRIVEN DECISION-MAKING

Several distinct governance problems arise once AI systems move from back-office automation into decisions that shape strategy, risk exposure, and stakeholder outcomes.

Executive reliance on AI recommendations. A recurring finding across the legal literature is that executives face a genuine bind when a model recommends a course of action. Ignoring the recommendation without documented justification can expose an executive to criticism for disregarding available information; following it without understanding its basis can expose the same executive to criticism for abdicating independent judgment. This is not a problem that better technology solves on its own. It is a governance design problem: organizations need documented protocols for when a model's recommendation can be overridden, by whom, and what record must be kept of the reasoning either way.

Autonomous decision systems. As systems move from decision support toward greater autonomy, actually executing trades, approving small loans, or adjusting prices in real time, the gap between the moment of design and the moment of harm widens. A model approved eighteen months ago may now be operating in market conditions its training data never anticipated. Krakowski and Raisch's work on the automation-augmentation tension, discussed in the corporate boards literature, points to a related risk: organizations that push autonomy too far too quickly often lose the tacit human judgment that would have caught an edge case the model was never trained to handle.

Opaque algorithms. Many of the models generating the most commercially significant recommendations, particularly deep learning systems, are difficult to explain even to the technical staff who built them. This is not merely an inconvenience for compliance reporting. It directly undermines the "informed decision" standard that underlies business judgment protection in most jurisdictions, since a director cannot be said to have exercised informed judgment over a recommendation whose basis cannot be reconstructed (Zhao & Gómez Fariñas, 2023). Explainability tools have improved, but a persistent tension remains between model performance and interpretability, and boards rarely have the technical fluency to evaluate that trade-off themselves.

Vendor responsibility. Most organizations license rather than build their core AI capability, which means governance failures can originate in a supply chain the board never directly examines. A vendor's silent model update, a change in training data, or an undisclosed subcontracting arrangement can alter system behavior without the client organization's knowledge. Resource dependence theory suggests this is not a peripheral concern but a structural one: contracts, audit rights, and change-notification clauses need to be treated as governance instruments, not merely legal boilerplate.

Data governance. Every downstream harm traceable to an AI system, biased outcomes, privacy violations, discriminatory pricing, usually has an upstream data governance failure behind it. Data lineage, consent management, and representativeness testing are technical disciplines, but they are also

governance questions, since a board that has not asked how training data was sourced and validated has effectively delegated a strategic risk decision to whoever assembled the dataset.

Bias and discrimination. The Amazon hiring case referenced in the introduction illustrates a pattern that recurs across lending, insurance underwriting, and criminal justice risk assessment: models trained on historical data reproduce and sometimes amplify the biases embedded in that history. Martin (2019) frames this as a design responsibility rather than an unfortunate side effect, arguing that firms cannot claim ignorance of biases that predictable auditing would have surfaced.

Cybersecurity risks. AI systems introduce attack surfaces that traditional cybersecurity governance was not built to address, including data poisoning, where an attacker manipulates training data to corrupt a model's future behavior, and adversarial inputs designed to fool a deployed model without altering its underlying code. Recent taxonomies of adversarial machine learning attacks catalogue these risks in detail and argue that they require dedicated governance attention distinct from conventional information security controls.

Compliance failures. Multiple regulatory regimes now impose overlapping and sometimes inconsistent obligations, documentation requirements under the EU AI Act, risk management expectations under the NIST framework, sector-specific rules for financial services or healthcare, and emerging state-level rules in the United States. A compliance function built only for financial and data-protection regulation is unlikely to catch the gaps specific to algorithmic decision-making unless its mandate is explicitly extended.

Shareholder accountability and board oversight. Shareholders increasingly expect boards to demonstrate active oversight of algorithmic risk, not merely delegate it to management, a shift visible in the growing frequency of AI-related risk disclosures in corporate reports and the emergence of AI oversight as a stated committee responsibility in some public company proxy statements. Audit committees, traditionally focused on financial controls, are being asked to extend their remit to algorithmic risk without always having the technical grounding to do so credibly, a gap several of the frameworks discussed above flag as unresolved.

Internal controls and risk governance. Traditional internal control frameworks assume a relatively stable system whose behavior can be tested once and monitored periodically. AI systems that continue to learn or that are retrained on new data violate that assumption, which means control testing needs to become continuous rather than periodic, a shift most internal audit functions have not yet fully absorbed.

Read together, these challenges share a common structure. Each one arises at a different point in an AI system's lifecycle, design, data sourcing, deployment, ongoing operation, and each implicates a different organizational function. This is the empirical basis for the multi-level framework proposed in Section 7: a single governance body cannot plausibly own all of these failure modes at once.

ACCOUNTABILITY ACROSS ORGANIZATIONAL LEVELS

If governance challenges arise at different points in an AI system's life, accountability has to be examined at the same granularity, level by level, rather than assigned wholesale to "the company" or "the board."

At the board level, the relevant duty is oversight, not operational control. Directors are not expected to understand the mathematics of a gradient-boosted model, but they are expected to ask whether the organization has a credible process for approving, testing, and monitoring AI systems that materially affect strategy or risk. The Chinese Journal of Comparative Law's recent treatment of board monitoring functions argues for an explicit, legally grounded human-in-the-loop requirement at the board level, precisely

because algorithmic decision-making otherwise erodes the board's ability to demonstrate the kind of informed engagement that satisfies fiduciary standards. A board that receives a single annual briefing on "AI risk" and treats the topic as closed for the year is not meeting this standard, regardless of how sophisticated the briefing sounds.

At the executive level, accountability concerns implementation judgment: whether AI recommendations were integrated into decisions with appropriate skepticism, whether override protocols were followed, and whether executives escalated concerns about model behavior rather than absorbing them quietly into business-as-usual. The tension identified earlier, between the risk of ignoring useful AI recommendations and the risk of over-relying on them, sits squarely at this level, and it cannot be resolved by policy alone. It requires executives who are willing to say, in a documented way, why they did or did not follow a system's recommendation.

A dedicated AI governance committee, whether a board subcommittee or a cross-functional executive body, increasingly appears in practice as the mechanism that translates board-level principles into operational standards. Its accountability is coordination: making sure that data governance, model risk, cybersecurity, and legal compliance are not each addressing AI risk in isolation, since a fragmented approach tends to produce the exact gaps that later cause a harmful outcome to go undetected until a customer, regulator, or journalist finds it first.

Risk management functions carry accountability for classification and monitoring: identifying which AI applications are high-risk under regulatory definitions such as those in the EU AI Act, setting risk appetite for algorithmic decision-making the way they already do for credit or market risk, and running the continuous testing that static, once-a-year control reviews cannot achieve. Enterprise risk management scholarship built on the COSO framework offers the natural home for this work, since AI risk is, in most respects, a new category layered onto an existing discipline rather than something requiring an entirely separate structure.

Internal audit carries a distinct accountability: independent verification that the controls risk management claims to have in place are actually operating as described. This is where the "reviewable automated decision-making" standard proposed by Cobbe, Lee, and Singh (2021) becomes operationally important. An auditor cannot verify a control that produces no reconstructable trail, which means internal audit's accountability depends directly on whether earlier design choices, made by technology teams, built in the ability to reconstruct a decision after the fact.

Technology teams carry accountability for the technical integrity of the systems they build and maintain: data quality, model validation, documentation, and the interpretability trade-offs discussed in Section 4. This is the level at which Martin's (2019) argument about algorithms as structuring agents lands most directly. If a model embeds a biased proxy variable, the responsibility for that choice sits with the people who selected the variable, even if the harm only becomes visible much later, at the point of deployment.

External AI vendors carry accountability that is often contractually defined but poorly enforced in practice. Resource dependence theory's insight applies directly here: a client organization's board bears ultimate accountability to its own stakeholders, but it can and should push part of the operational accountability onto vendors through audit rights, change-notification obligations, and shared liability provisions. The absence of such provisions in many current commercial AI contracts is, in itself, a governance gap.

Regulators, finally, carry a different kind of accountability, not for a specific company's decisions, but for the clarity and consistency of the standards against which all of the levels above are judged. Kaminski's

(2019) account of binary governance is instructive here: regulation that combines individual rights with ongoing collaborative oversight tends to produce more consistent corporate behavior than either pure rule-based compliance or pure self-regulation, since it gives companies a stable reference point without requiring an unrealistic degree of technical prescription from the regulator.

This level-by-level account matters because it resists two tempting but flawed simplifications. The first flaw is concentrating all accountability at the board, which sounds appropriately serious but is operationally meaningless, since boards cannot supervise model training data on a weekly basis. The second flaw is diffusing accountability so broadly, "the whole organization is responsible", that Johnson's (2021) diffusion-of-responsibility problem sets in and no one is actually answerable when something goes wrong. The framework in Section 7 is built to avoid both traps by giving each level a distinct, checkable responsibility rather than a shared, vague one.

COMPARATIVE POLICY AND REGULATORY PERSPECTIVES

Regulatory approaches to AI governance differ enough across jurisdictions that a board operating across borders cannot simply adopt one national standard and assume it satisfies the rest.

The European Union has taken the most prescriptive route. The AI Act, formally Regulation (EU) 2024/1689, classifies AI systems by risk tier and imposes binding obligations on providers and deployers of high-risk systems, including conformity assessments, technical documentation, human oversight measures, and post-market monitoring (European Parliament and Council of the European Union, 2024). This sits alongside the General Data Protection Regulation, which already restricts fully automated decisions that produce legal or similarly significant effects on individuals unless specific safeguards are in place. Together, these instruments push EU boards toward a compliance posture that is closer to product safety regulation than to the principles-based codes more familiar in corporate governance. The practical implication for boards operating in or selling into the EU is that AI governance cannot be treated purely as an internal risk management exercise; it now carries direct regulatory exposure with penalties that can reach into the tens of millions of euros for the most serious violations.

The United States has so far avoided a single comprehensive federal AI statute, relying instead on a combination of sectoral regulation, existing consumer protection and anti-discrimination law, and voluntary technical guidance. The NIST AI Risk Management Framework, published in 2023, is the closest thing to a national reference point, organized around four functions, govern, map, measure, and manage, that are explicitly designed to be woven into an organization's existing risk and compliance architecture rather than stood up as a parallel structure (National Institute of Standards and Technology, 2023). Because the framework is voluntary, its practical force comes less from legal compulsion and more from its adoption as a de facto standard among federal contractors, insurers evaluating corporate risk, and litigants seeking to establish a reasonable standard of care after something has gone wrong. This produces a governance environment that rewards documented process even in the absence of a specific legal mandate, since demonstrating alignment with a recognized framework has become a meaningful defense in itself.

The United Kingdom has taken a third path, favoring a principles-based, sector-led approach in which existing regulators, financial, competition, data protection, apply a shared set of cross-sectoral AI principles within their own domains rather than a single AI-specific statute imposing uniform rules. This gives UK boards more interpretive flexibility than their EU counterparts but correspondingly less certainty about how a given regulator will apply the principles to a specific case, a trade-off that several governance practitioners have described as shifting more of the burden onto the board's own judgment rather than onto a fixed rulebook.

Selected Asia-Pacific economies illustrate a further variation. Singapore's Model AI Governance Framework, developed jointly by its data protection and infocomm regulators, emphasizes explainability, auditability, and traceability as practical design requirements rather than abstract principles, giving companies a reasonably concrete template to build against. Comparable initiatives elsewhere in the region tend to combine elements of the EU's risk-tiering language with the more voluntary, guidance-based posture closer to the American approach, producing a patchwork that multinational boards need to map carefully rather than assume converges on a single global standard.

What unites these otherwise divergent approaches is a shared emphasis on human oversight, transparency, and continuous risk management rather than a one-time technical certification. The OECD AI Principles function as something like a common vocabulary underneath all four regimes, referenced explicitly in the EU AI Act and echoed in the NIST framework's language on trustworthy AI (OECD, 2019/2024). For multinational boards, the practical governance lesson is not to pick the strictest regime and apply it everywhere, though that is sometimes the simplest operational choice, but to build an internal framework flexible enough to demonstrate compliance with the specific documentation, oversight, and audit expectations of each jurisdiction in which the company operates. This is precisely the gap the framework proposed in the next section is designed to fill: a structure general enough to sit underneath multiple regulatory regimes, while specific enough to assign real responsibility inside the organization.

A PROPOSED CORPORATE AI ACCOUNTABILITY GOVERNANCE FRAMEWORK

The preceding sections point toward a single design requirement: accountability for algorithmic decisions has to be distributed across the organizational levels where different types of failure originate, and each level's responsibility has to be specific enough to be checked, not merely stated as a shared aspiration. This section proposes such a structure, the Corporate AI Accountability Governance Framework, built around eight actors and four accountability domains.

The four domains, drawn from the failure patterns identified in Section 4, are design accountability (was the system built with adequate data, testing, and documentation), approval accountability (was the system authorized through an informed process that understood its risks), deployment accountability (was the system used appropriately in the specific business context it was applied to), and monitoring accountability (is the system's ongoing behavior being checked against the assumptions under which it was approved). Every one of the eight actors below carries primary responsibility for at least one domain and a supporting responsibility for at least one other, which is what distinguishes this framework from a simple responsibility matrix: no actor is accountable for everything, but no actor is accountable for nothing either.

Board of Directors holds primary approval accountability at the strategic level and supporting monitoring accountability. Its specific responsibilities include approving a written AI risk appetite statement, requiring documented board-level review of any AI system classified as high-risk under applicable regulation before deployment, and requiring at least semi-annual reporting on algorithmic incidents, near-misses, and model drift, not merely an annual overview briefing. This satisfies the human-in-the-loop standard argued for in recent governance scholarship without requiring directors to acquire technical modeling skills they do not need (Chinese Journal of Comparative Law treatment of board monitoring functions).

Executive Leadership holds primary deployment accountability and supporting design accountability. Chief executives and business unit leaders are responsible for ensuring that override protocols exist and are actually used when a model's recommendation conflicts with business judgment, that overrides are documented with reasoning, and that executive compensation and performance incentives do not implicitly reward ignoring model-flagged risks in pursuit of short-term commercial targets, a subtle but important safeguard against the diffusion-of-responsibility problem Johnson (2021) warns about.

AI Governance Committee, whether structured as a board subcommittee or a cross-functional executive body reporting to the board, holds primary coordination accountability across all four domains. Its role is not to duplicate the work of risk management, internal audit, or technology teams, but to ensure those functions are working from a shared risk taxonomy, a shared incident reporting process, and a shared view of which AI systems are currently in use across the organization, an inventory that surprisingly few companies maintain in complete form even today.

Risk Management holds primary monitoring accountability and supporting approval accountability. Its specific tasks include classifying AI systems by risk tier consistent with applicable regulatory frameworks, setting quantitative and qualitative risk appetite thresholds specific to algorithmic decision-making, and running continuous, rather than periodic, testing regimes appropriate for systems that can change behavior after deployment, extending the enterprise risk management discipline formalized in the COSO (2017) framework into this new risk category.

Internal Audit holds primary verification accountability, checking that the controls risk management and technology teams claim to operate are actually functioning, and that AI systems meet the reviewability standard needed for meaningful audit (Cobbe, Lee, & Singh, 2021). Internal audit's independence from the teams that build and deploy AI systems is what gives its sign-off credibility to the board and, where relevant, to external regulators.

Technology Teams hold primary design accountability, covering data sourcing and validation, model testing prior to deployment, documentation sufficient for both internal audit and external regulatory review, and ongoing technical monitoring for model drift, the tendency of a model's real-world accuracy to degrade as conditions change from those reflected in its training data. This is the level at which Martin's (2019) argument about algorithms as structuring agents becomes an operational requirement rather than a philosophical claim: technology teams need documented sign-off procedures that make explicit which design choices were made and why.

External AI Vendors hold contractually defined accountability that should extend to audit rights, change-notification obligations when a model or its training data is materially updated, and shared liability provisions proportionate to the vendor's role in a harmful outcome. Resource dependence theory's insight that external dependence is itself a governance risk, not merely a commercial one, supports treating vendor accountability as a distinct category rather than folding it into ordinary procurement terms.

Regulators sit outside the organization but complete the framework by supplying the external standard against which all of the above is measured, the risk classifications, disclosure requirements, and enforcement mechanisms that give the internal framework its teeth. Kaminski's (2019) binary governance concept is useful here as a design principle for regulators themselves: combining individual rights protections with collaborative, ongoing oversight tends to produce steadier corporate behavior than either pure rule-following or pure self-regulation.

The framework's distinguishing feature, relative to existing governance models, is its explicit treatment of AI systems as objects that continue to change after approval. Conventional governance controls assume that a decision, once approved, remains stable until deliberately revisited. A model retrained on new data, or a vendor's system updated without notice, violates that assumption. The framework therefore builds monitoring accountability into every level rather than treating it as a separate, final stage that only risk management or internal audit need worry about. A board that reviews AI risk once a year, an executive team that only checks override logs during an annual audit, and a technology team that tests a model once before launch and never again are each, in their own way, applying a static governance mindset to a system that does not behave statically. The framework is designed to close that mismatch by making monitoring a

continuous, shared responsibility distributed across all eight actors rather than a single checkpoint owned by one of them.

MANAGERIAL IMPLICATIONS

For practitioners, the framework translates into a fairly specific set of near-term actions, though the emphasis will differ depending on organizational size and regulatory exposure.

Boards should insist on a written AI risk appetite statement distinct from, though connected to, existing risk appetite statements for financial and operational risk. Without this document, board discussions about AI tend to default to anecdote, a single high-profile failure at a competitor, rather than a structured assessment of the company's own exposure. Boards should also require that any AI system materially affecting pricing, credit, hiring, or safety decisions be reviewed before deployment, not after a problem surfaces, and that this review be documented in a form an external auditor or regulator could later examine.

Executive teams need to build override protocols that are actually usable under time pressure, since a protocol that requires a lengthy sign-off process will simply be bypassed when a decision needs to be made quickly, recreating the very risk the protocol was meant to control. Executives should also resist the temptation to treat a vendor's assurance of "responsible AI" as a substitute for internal due diligence; the resource dependence literature is unambiguous that outsourcing a capability does not outsource accountability for its consequences.

Risk and audit functions should build an AI system inventory before attempting anything more sophisticated. It is difficult to overstate how often organizations, when asked directly, cannot produce a complete list of the AI and machine learning systems currently influencing decisions across their business units. Classification by risk tier, consistent with the EU AI Act's categories even for companies outside EU jurisdiction, since the categories are becoming a de facto international vocabulary, should follow from that inventory rather than precede it.

Technology teams should treat documentation as a governance deliverable, not an afterthought produced only when a regulator asks for it. The reviewability standard proposed by Cobbe, Lee, and Singh (2021) is achievable in practice, but only if it is designed in from the start; retrofitting explainability onto a system already in production is considerably harder and often incomplete.

Finally, organizations negotiating AI vendor contracts should treat audit rights and change-notification clauses as non-negotiable for any system classified as high-risk, even when vendors resist on grounds of protecting proprietary model details. A workable compromise, increasingly seen in practice, involves third-party audit arrangements that protect vendor intellectual property while still giving the client organization the assurance it needs to satisfy its own board and regulators.

POLICY RECOMMENDATIONS

Several recommendations follow for regulators and standard-setting bodies, distinct from the managerial implications above.

Regulators across jurisdictions would benefit from converging on a shared definition of what counts as a "high-risk" AI application in a corporate governance context, since the current patchwork, EU risk tiers, sector-specific US rules, principles-based UK guidance, forces multinational boards to maintain parallel compliance tracks that add cost without necessarily adding safety. Convergence does not require identical rules, but a shared taxonomy would reduce the compliance burden considerably.

Standard-setting bodies should develop clearer guidance on the boundary between vendor and client accountability, an area where current frameworks, including the EU AI Act's distinction between providers

and deployers, offer a starting point but leave considerable ambiguity for the many organizations that customize a vendor's base model for their own use, blurring the line between the two categories.

Regulators should also consider requiring, rather than merely encouraging, board-level disclosure of AI governance structures in a form comparable to existing disclosure requirements for financial risk and, in some jurisdictions, climate risk. Voluntary disclosure has produced uneven practice, with some companies providing detailed AI governance sections in their annual reports and others saying little beyond generic references to "responsible AI," a gap that mandatory minimum disclosure standards could narrow.

Finally, professional bodies responsible for director education and audit standards should update their curricula to reflect the accountability distribution proposed in this paper, so that directors and auditors entering these roles are equipped to ask the right questions about algorithmic risk rather than treating it as a specialist topic reserved for a chief technology officer.

FUTURE RESEARCH AGENDA

Several open questions deserve sustained empirical attention that this conceptual paper cannot resolve on its own.

Longitudinal studies tracking how boards actually allocate AI oversight responsibilities in practice, as opposed to how governance codes suggest they should, would test whether the framework proposed here matches observed behavior or whether some other allocation emerges informally. Comparative empirical work examining whether companies operating under the EU AI Act's binding obligations show measurably different governance maturity than companies operating under the United States' voluntary framework would help settle an open debate about whether binding regulation actually improves practice or merely improves documentation.

Research on the vendor-client accountability boundary would be particularly valuable given how quickly commercial AI procurement has grown; case-based research examining how liability was actually allocated after a documented vendor-related AI failure, rather than how contracts theoretically allocate it, would sharpen the resource dependence argument advanced in Section 3. Similarly, research connecting stewardship theory to AI governance empirically, testing whether organizations with stronger stewardship cultures actually detect and self-report algorithmic problems earlier than organizations relying primarily on external audit and compliance pressure, would help resolve the tension flagged in this paper between control-based and stewardship-based approaches to oversight.

Finally, as AI systems move toward greater autonomy in decision execution rather than mere recommendation, scholars will need to revisit whether the accountability framework proposed here, built around distinct human-owned domains of design, approval, deployment, and monitoring, remains adequate once a meaningful share of decisions no longer pass through a human checkpoint at all before taking effect.

CONCLUSION

Corporate governance has always rested on the assumption that a traceable human being sits behind every significant organizational decision, someone who can be asked to explain a judgment and held accountable for its consequences. Algorithmic decision-making does not eliminate that assumption, but it stretches it thin enough that older governance habits, a single annual technology briefing, a compliance checkbox at launch, an assumption that a vendor's assurances substitute for internal diligence, are no longer sufficient. The central argument of this paper is that accountability for AI-related harm has to be distributed deliberately across the organizational levels where different types of failure actually originate, design,

approval, deployment, and monitoring, rather than left to default to whichever function happens to be closest when something goes wrong. The Corporate AI Accountability Governance Framework proposed here is one way of making that distribution explicit and checkable, assigning distinct responsibilities to boards, executives, a coordinating governance committee, risk management, internal audit, technology teams, external vendors, and regulators, while building continuous monitoring into every level rather than treating it as a single closing step. Whether this particular allocation proves durable is a question for the empirical research the paper's final section calls for. What seems less open to debate is the underlying premise: as long as AI systems keep changing after they are approved, governance structures built around one-time sign-offs will keep failing to catch problems until customers, employees, or regulators find them first.

REFERENCES

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825–836.

Cobbe, J., Lee, M. S. A., & Singh, J. (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 598–609).

Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2017). *Enterprise risk management: Integrating with strategy and performance*.

Coulson-Thomas, C. (2023). AI, executive and board leadership, and our collective future. *Effective Executive*, 26(3), 5–29.

Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management Review*, 22(1), 20–47.

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.

European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*.

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.

Johnson, D. G. (2021). Algorithmic accountability in the making. *Social Philosophy and Policy*, 38(2), 111–127.

Kaminski, M. E. (2019). Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *Southern California Law Review*, 92, 1529–1616.

Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: Ethical challenges and normative thinking. *Accounting, Auditing & Accountability Journal*, 35(9), 109–135.

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835–850.

Mertens, F. (2023). *The use of artificial intelligence in corporate decision-making at board level: A preliminary legal analysis*. SSRN Working Paper.

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1(11), 501–507.

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)*. U.S. Department of Commerce.

Organisation for Economic Co-operation and Development (OECD). (2019, updated 2024). *OECD principles on artificial intelligence*.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*. Harper & Row.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.

Zhao, J., & Gómez Fariñas, B. (2023). Artificial intelligence and sustainable decisions. *European Business Organization Law Review*, 24, 1–39.