Stanford – Vienna
Transatlantic Technology Law Forum

A joint initiative of
Stanford Law School and the University of Vienna School of Law

European Union Law Working Papers

No. 135

Fiduciary Duties and Business Judgment Rule 2.0 in the EU AI Act Age

Lillà Montagnani and Maria Lucia Passador

2026

European Union Law Working Papers

Editors: Siegfried Fina and Roland Vogl

About the European Union Law Working Papers

The European Union Law Working Paper Series presents research on the law and policy of the European Union. The objective of the European Union Law Working Paper Series is to share "works in progress". The authors of the papers are solely responsible for the content of their contributions and may use the citation standards of their home country. The working papers can be found at http://ttlf.stanford.edu.

The European Union Law Working Paper Series is a joint initiative of Stanford Law School and the University of Vienna School of Law's LLM Program in European and International Business Law.

If you should have any questions regarding the European Union Law Working Paper Series, please contact Professor Dr. Siegfried Fina, Jean Monnet Professor of European Union Law, or Dr. Roland Vogl, Executive Director of the Stanford Program in Law, Science and Technology, at:

Stanford-Vienna Transatlantic Technology Law Forum
http://ttlf.stanford.edu

Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305-8610

University of Vienna School of Law
Department of Business Law
Schottenbastei 10-16
1010 Vienna, Austria

About the Authors

Maria Lillà Montagnani is a Full Professor of Commercial Law, IT and IP Law, Bocconi University, Milan, Italy; Former Fellow, Transatlantic Technology Law Forum (TTLF), Stanford Law School.
Maria Lucia Passador is an Assistant Professor of Corporate Law and Financial Markets Regulation, Bocconi University, Milan, Italy; Adjunct Professor, SUPSI; Affiliate, Transatlantic Technology Law Forum (TTLF), Stanford Law School.
Although the paper is the product of a joint collaboration, authorship of Paragraphs I and II is attributable to Lillà Montagnani, whereas Paragraphs IV, V and VI were authored by Maria Lucia Passador; Paragraph III was jointly authored.

General Note about the Content

Suggested Citation

This European Union Law Working Paper should be cited as:
Lillà Montagnani and Maria Lucia Passador, Fiduciary Duties and Business Judgment Rule 2.0 in the EU AI Act Age, Stanford-Vienna European Union Law Working Paper No. 135, http://ttlf.stanford.edu.

Copyright

Abstract

This Article examines how the EU Artificial Intelligence Act unsettles the justificatory foundations of fiduciary deference in corporate law. Although the Act does not directly regulate corporate boards, it institutionalizes procedural expectations of oversight, traceability, and anticipatory risk assessment that alter the institutional conditions under which fiduciary judgment is assessed.

The Article argues that the challenge posed by algorithmic decision-making is not primarily one of compliance or technical competence, but of justification. As corporate decisions are increasingly mediated by opaque systems that directors neither fully understand nor meaningfully interrogate, the premises on which judicial deference—embodied in the Business Judgment Rule—rests become contestable.

By analyzing the AI Act as a process-oriented governance regime, the Article shows how its procedural mandates risk recasting fiduciary oversight as a largely formalized practice divorced from substantive judgment. Against this backdrop, it reconceptualizes directors' duties of care and loyalty under conditions of algorithmic reliance, identifying an emergent expectation of demonstrable stewardship over AI systems.

The Article concludes that fiduciary deference remains normatively defensible only where algorithmic tools mediate, rather than substitute for, human judgment. Absent such engagement, the Business Judgment Rule risks insulating not discretion, but abdication.

CONTENTS

Plus ça change, plus c'est la même chose.

Jean-Baptiste Alphonse Karr, Les Guêpes (1849)

INTRODUCTION

Corporate decision-making is undergoing a quiet but consequential transformation. Across domains long regarded as the core terrain of fiduciary judgment—workforce management, credit allocation, compliance monitoring, enterprise risk assessment, and strategic planning—corporate boards increasingly rely on algorithmic systems to inform, structure, and, in some cases, effectively determine outcomes. These systems no longer merely assist deliberation. They increasingly shape the decisional architecture within which judgment is exercised. Yet corporate law continues to extend fiduciary deference on the assumption that board decisions remain attributable to human judgment exercised under conditions of uncertainty. That assumption is no longer self-evident.

The challenge posed by artificial intelligence in the boardroom is not, at its core, a question of technological adoption or regulatory compliance. It is a problem of justification.[1] Doctrines such as the Business Judgment Rule ("BJR") rest on an implicit epistemic premise: that directors, even when wrong, are the primary authors of the decisions for which they are held accountable.[2] Judicial abstention has traditionally been justified by deference to this human judgment—bounded, fallible, but institutionally situated. As corporate decision-making becomes increasingly mediated through opaque, adaptive systems whose internal logic directors neither fully understand nor meaningfully interrogate, the normative foundations of fiduciary deference come under strain. The central question this Article addresses is therefore not whether directors may rely on AI, but whether

[1] *See generally* Martin Petrin, *Corporate Management in the Age of Artificial Intelligence*, 2019 *Colum. Bus. L. Rev.* 965 (2019); Katja Langenbucher, *AI Judgment Rule(s), U. Chi. L. Rev. Online*, https://lawreview.uchicago.edu/online-archive/ai-judgment-rules

[2] *See generally* Bernard S. Sharfman, *The Importance of the Business Judgment Rule*, 14 N.Y.U. J.L. & Bus. 27 (2017).

fiduciary deference remains conceptually and normatively coherent under conditions of algorithmic mediation.

This Article argues that the European Union Artificial Intelligence Act ("AI Act")[3] marks a critical regulatory inflection point that renders this problem newly visible. The significance of the AI Act for corporate law does not lie in its formal scope. The Act does not directly regulate corporate boards, nor does it purport to revise fiduciary doctrine. Its importance lies elsewhere: in the density and explicitness with which it institutionalises procedural expectations of oversight, traceability, human supervision, and anticipatory risk management. By targeting the internal decision architectures through which organisational choices are made, the AI Act exemplifies a broader shift toward process-oriented—indeed, epistemic—governance. These techniques do not merely regulate technology. They reshape the institutional conditions under which corporate judgment is exercised, documented, and evaluated.

Although formally addressed to AI providers, deployers, and users, the AI Act inevitably implicates board-level governance. Directors are exposed indirectly insofar as they authorise, supervise, or rely upon AI systems whose lawful use presupposes compliance with extensive procedural obligations—particularly in relation to high-risk systems deployed in employment, credit allocation, and strategic risk management. The result is a reconfiguration of the environment in which fiduciary judgment operates, even in the absence of any explicit reform of corporate law doctrine. What changes is not the content of fiduciary duties, but the background assumptions that have traditionally justified deference to their exercise. Even in jurisdictions where the AI Act does not apply, its governance architecture crystallizes a structural problem fiduciary law already faces: how to justify deference when judgment is increasingly mediated by systems that directors do not fully control or understand.

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Certain Union Legislative Acts, 2024 O.J. (L 1689) 1.

For the purposes of this Article, references to "boards" and "directors" concern the governing bodies of large and medium-sized corporate users of AI systems, particularly publicly listed companies and enterprises for which strategic and oversight decisions are plausibly mediated by algorithmic tools. The problem, however, is not whether directors may delegate analytical or operational tasks to complex systems. Corporate law has never required directors to personally perform every task underlying a board decision. The problem is whether judicial deference remains legitimate when judgment itself is increasingly channelled through systems whose operations are opaque, adaptive, and only partially intelligible to human overseers.

Existing scholarship has examined the relationship between AI and corporate governance from multiple angles. Early contributions framed AI primarily as a tool for automating routine functions, improving efficiency, mitigating bias, and enhancing compliance.[4] More recent work has explored AI's potential to reconfigure—and in some accounts to supplant—the deliberative and strategic capacities of boards.[5] Proponents emphasise algorithmic consistency, improved decisional accuracy, and reduced agency costs, sometimes drawing analogies to algorithmic tools in judicial or administrative decision-making.[6] At the same time, sustained critiques warn that delegating strategic and oversight functions to algorithmic systems risks hollowing out the normative foundations of corporate governance, particularly the role of human judgment in balancing legal, ethical, and commercial considerations.[7] Parallel debates surrounding technocratic "Board Service Providers" reflect similar anxieties about governance reduced to functional optimisation.[8]

---

[4] Petrin, *supra* note 1, at 965.
[5] Langenbucher, *supra* note 1.
[6] Respectively, Maria Lucia Passador, *From White Noise to Sound Decisions: Overcoming Noise in Corporate Law*, 13 Am. U. Bus. L. Rev. 265, 306 (2024); Mayank Shukla, *The Impact of AI on Improving the Efficiency and Accuracy of Managerial Decisions*, 12 Int'l J. for Rsch. Applied Sci. & Eng'g Tech. 830 (2024); Sreeram Mullankandy, *Transforming Data into Compliance: Harnessing AI/ML to Enhance Regulatory Reporting Processes*, 3 J. Artificial Intell. & Gen. Sci. 62 (2024).

[7] Petrin, *supra* note 1, at 965.
[8] Stephen M. Bainbridge & M. Todd Henderson, *Boards-R-Us: Reconceptualizing Corporate Boards*, 66 Stan. L. Rev. 1051, 1061, 1064–68 (2014); *see also* Stephen M. Bainbridge & M. Todd Henderson, *Outsourcing the Board* 190–202 (Cambridge Univ. Press 2018); Ronald J. Gilson & Jeffrey N. Gordon, *Board 3.0—An Introduction*, 74 Bus. Law. 351, 361–63 (2019).

Concerns about opacity, bias embedded in training data, and the ethical limitations of machine inference further complicate this picture. Even when operating consistently, algorithmic systems lack the contextual sensitivity required for complex, multi-stakeholder decisions.[9] Yet directors are increasingly expected to embrace AI-mediated decision-making without a stable doctrinal or procedural framework capable of addressing these challenges. This institutional lacuna generates legal uncertainty and divergent corporate responses: some firms aggressively deploy AI to secure competitive advantage, while others proceed cautiously, wary of legal and reputational risk. Although AI is now embedded in boardroom processes, it remains only loosely connected to formal governance structures. Ad hoc responses—such as the creation of technology or AI committees—have emerged, but their authority, scope, and fiduciary significance remain underdeveloped.

Against this backdrop, the AI Act operates as a regulatory shock. Unlike earlier regimes such as the General Data Protection Regulation ("GDPR"),[10] which primarily addressed informational flows and individual rights ex post, the AI Act targets organisational decision-making ex ante. By institutionalising requirements of human oversight, systemic risk anticipation, documentation, and auditability, it exemplifies a shift from informational compliance to epistemic governance. It is this shift that renders the AI Act doctrinally salient for fiduciary law. By proceduralising oversight and formalising stewardship over decision-making systems, the Act indirectly but materially reframes the conditions under which fiduciary discretion claims legitimacy.

This Article advances three core claims. First, it argues that the AI Act should be understood as a paradigmatic example of process-based governance that reshapes the environment of corporate decision-making, even though it does not formally regulate boards. Second, it shows that the

[9] Eric A. Posner & Shivam Saran, *Judge AI: Assessing Large Language Models in Judicial Decision-Making*, Univ. of Chi. Coase-Sandor Inst. for L. & Econ. Rsch. Paper No. 25-03 (Jan. 15, 2025), https://ssrn.com/abstract=5098708.
[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), 2016 O.J. (L 119) 1.

procedural mandates embedded in the AI Act risk transforming fiduciary oversight into a performative exercise—what Michael Power famously described as "rituals of verification"—in which documentation and auditability may substitute for substantive judgment.[11] Third, and most importantly, this Article contends that no existing scholarship has examined how procedural AI governance destabilises the justificatory foundations of fiduciary deference itself.[12] The challenge posed by algorithmic governance is not primarily one of compliance or doctrinal insufficiency, but of justification: whether courts can continue to abstain from review when judgment is increasingly mediated by systems that directors do not meaningfully control or understand.

The Article does not propose new fiduciary duties, nor does it advocate the erosion of the BJR. Its claim is more modest, and more fundamental. By altering the informational conditions under which corporate judgment is exercised, the AI Act refracts regulatory expectations into fiduciary doctrine. In this environment, the duties of care and loyalty acquire a modified significance. Care increasingly operates against heightened expectations of technological literacy, procedural vigilance, and informed oversight. Loyalty evolves as directors must weigh efficiency gains against the ethical, stakeholder, and institutional implications embedded in algorithmic systems. The Article captures these developments through the analytical lenses of "AI due care" and "AI loyalty oversight," not as new doctrinal categories, but as reinterpretations of existing fiduciary standards under conditions of algorithmic mediation.

Focusing on high-stakes, board-mediated uses of AI—particularly in employment, credit allocation, and strategic risk modelling—this Article shows that the continued legitimacy of fiduciary deference depends on whether algorithmic tools mediate, rather than substitute for, human judgment. Where

[11] Michael Power, The Audit Society: Rituals of Verification (Oxford Univ. Press 1997).

[12] *See, e.g.,* Alfred R. Cowger Jr., *Corporate Fiduciary Duty in the Age of Algorithms*, 14 *J.L. Tech. & Internet* ___ (2022–2023) (examining how algorithmic decision-making reshapes the content of fiduciary duties); Bart Custers, Henning Lahmann & Benjamyn I. Scott, *From Liability Gaps to Liability Overlaps: Shared Responsibilities and Fiduciary Duties in AI and Other Complex Technologies*, *AI & Soc'y* (2025), https://doi.org/10.1007/s00146-024-02137-1 (addressing the allocation of responsibility and fiduciary concepts in complex AI systems).

oversight remains substantive, contextual, and institutionally traceable, deference remains normatively defensible. Where it does not, the BJR risks insulating not discretion, but abdication.

This Article proceeds as follows. Paragraph I analyses the EU Artificial Intelligence Act as a paradigmatic instance of process-oriented, epistemic governance, showing how—despite its formal inapplicability to corporate boards—it reshapes the conditions under which fiduciary discretion is exercised and judicial deference is justified. The following Paragraphs examine how the Act's procedural mandates translate into board-level governance exposure, focusing on directors' oversight responsibilities in high-stakes, AI-mediated decision-making contexts such as workforce management and consumer-facing operations. Paragraphs III and IV develop the concepts of AI due care and AI loyalty oversight as analytical lenses for reinterpreting fiduciary duties under conditions of algorithmic mediation, emphasising technological literacy, anticipatory risk governance, and stewardship over algorithmic systems as prerequisites of legitimate oversight. Paragraph V paragraph turns to the BJR, arguing that its justificatory foundations increasingly depend on demonstrable stewardship over algorithmic decision-making infrastructures rather than uncritical reliance on automated outputs. Paragraph VI shows that fiduciary deference endures in the age of AI—but only under altered conditions in which judgment remains substantively human, institutionally traceable, and worthy of deference.

I. The EU Artificial Intelligence Act

This Part does not offer a comprehensive account of the EU Artificial Intelligence Act. Nor does it proceed on the assumption that the Act's political durability, enforcement trajectory, or global normative authority will determine its significance for corporate law. The relevance of the AI Act for fiduciary doctrine lies elsewhere. It lies in the density and explicitness with which the Act proceduralises oversight, formalises risk governance, and institutionalises human control as a legal

expectation—thereby making visible a governance logic that fiduciary law has already begun, often implicitly, to absorb.[13]

It is this shift—from outcome-oriented regulation to epistemic governance—that renders the AI Act uniquely salient for fiduciary doctrine.

The central claim advanced here is straightforward: although the AI Act is not addressed to corporate boards, it materially alters the justificatory environment in which fiduciary discretion is exercised and judicial deference is sustained. By embedding documentation, traceability, and structured oversight into organisational decision architectures, the Act reshapes the conditions under which directors can plausibly claim to have exercised judgment. This shift is doctrinally salient not because it creates new fiduciary duties, but because it recalibrates the background assumptions that traditionally justify judicial abstention under doctrines such as the BJR.

This Part makes four moves. First, it identifies the AI Act as a paradigmatic example of process-oriented technological governance—one that targets internal decision architectures rather than discrete outcomes. Second, it examines how the Act's expansive definition of AI and its risk-classification framework extend procedural governance across a wide range of organisational technologies. Third, it shows how the Act's transparency and documentation requirements risk transforming oversight into a largely performative exercise of verification. Fourth, it situates this dynamic within a broader transnational convergence around auditability and procedural legitimacy, sharpening the role of fiduciary duties as a counterweight to symbolic compliance.

A. The AI Act as a Process-Based Governance Regime

The AI Act constitutes the most procedurally articulated regulatory intervention to date within the European Union's technological governance landscape. Although it does not explicitly designate

[13] See generally, *See, e.g.,* James Sayles, *AI Governance and Oversight Model*, in *Principles of AI Governance and Model Risk Management* 183–208 (James Sayles ed., Apress 2024).

corporate directors as its principal regulatory addressees, its operative provisions nonetheless exert a substantial legal and strategic influence on firms—particularly listed companies operating in AI-intensive sectors. This influence does not derive from the Act's normative finality. Indeed, the AI Act remains contested and institutionally fragile. Its doctrinal relevance for corporate law lies instead in the way it externalises governance expectations through dense procedural mandates.

Formally sector-neutral, the Act nonetheless reshapes the justificatory environment surrounding board-level decision-making. Its obligations indirectly but significantly sharpen expectations around due diligence, anticipatory risk management, and accountability in AI-mediated organisational processes. Directors are implicated not because the Act imposes duties upon them directly, but because the lawful deployment of AI systems increasingly presupposes organisational compliance with governance practices that boards authorise, supervise, and rely upon.

The Act's extraterritorial scope reinforces this dynamic. By applying to both EU-based and third-country actors whose systems affect the EU market, the AI Act complicates compliance strategies for multinational enterprises and intensifies pressure to recalibrate governance structures across jurisdictions.[14] What emerges is not merely a compliance challenge, but a strategic governance constraint: boards are increasingly required to operate within decision environments structured by proceduralised oversight regimes, even where those regimes do not formally address fiduciary actors.

The AI Act's overarching ambition—to foster "trustworthy" and "human-centric" AI[15]—translates into a complex matrix of obligations that render visible the limits of formal legal compliance. Risk-tiered classification, mandatory transparency, and stringent penalties create incentives for

[14] Maria Lucia Passador, *Transcending Boundaries in the Age of International Corporate and Financial Law*, forthcoming 39 Emory Int'l L. Rev. (2025), https://ssrn.com/abstract=4713225 (last visited May 8, 2025).

[15] *See*, on the one hand, Johann Laux, Sandra Wachter & Brent Mittelstadt, *Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk*, 18 Reg. & Governance 3 (2024); Johann Laux, *Institutionalised Distrust and Human Oversight of Artificial Intelligence: Towards a Democratic Design of AI Governance under the European Union AI Act*, 39 AI & Soc'y 2853 (2024); *and*, on the other hand, Ross P. Buckley, Dirk A. Zetzsche, Douglas W. Arner & Brian Tang, *Regulating Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 Sydney L. Rev. 43 (2021), https://ssrn.com/abstract=3831758.

organisational reconfiguration touching strategic planning, capital allocation, and internal control systems.[16] The significance of these features for fiduciary doctrine lies not in their technical detail, but in how they condition the way judgment is exercised, recorded, and later justified.

B. Definition, Risk Classification, and the Expansion of Governance Infrastructure

One of the Artificial Intelligence Act's most consequential design choices is its deliberately broad yet open-textured definition of an "AI system" in Article 3(1). Under that Article, an AI system is defined functionally as a machine-based system designed to operate with varying degrees of autonomy, that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. This seven-element formulation, while capacious, intentionally captures a wide range of algorithmic and data-driven systems as within the Act's regulatory scope. [17]

Recognising the potential uncertainties inherent in such an open-textured provision, the Commission has published non-binding Guidelines to assist stakeholders in determining whether a given software system qualifies as an AI system under Article 3(1). These Guidelines explicate each constituent element, emphasise that *inference* (rather than mere rule execution) and a meaningful degree of autonomy are critical thresholds, and aim to delineate the outer bounds of the definition so as to avoid unintended inclusion of traditional or purely deterministic software.[18]

---

[16] Claudio Novelli, Mariarosaria Taddeo & Luciano Floridi, *Accountability in Artificial Intelligence: What It Is and How It Works*, 39 *AI & Soc.* 1871 (2024), https://doi.org/10.1007/s00146-023-01635-y.

[17] Theodore S. Boone, *The Challenge of Defining Artificial Intelligence in the EU AI Act*, 6 J. Data Prot. & Privacy 180 (2023).

[18] Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act), COMM. OF THE EUR. COMM'N (July 29, 2025) (non-binding guidance on Article 3(1) definition of "AI system"), available at *https://ai-act-service-desk.ec.europa.eu/sites/default/files/2025-08/commission_guidelines_on_the_definition_of_an_artificial_intelligence_system_established_by_regulation_eu_20241689_ai_actenglish_nf2skcqfrtjdfggjavcodopcwz4_112455.PDF*.

Doctrinally, the Act's definitional architecture matters because it anchors the set of technologies that may be subject to substantive regulatory duties. Even though the Guidelines operate as an interpretive and containment mechanism, the combination of the statutory definition and its administrative interpretation extends regulatory reach well beyond what would be captured by narrow conceptions of "advanced" or "learning" AI. As a result, firms deploying tools with modest levels of adaptiveness or autonomy may nonetheless fall within the regime's compliance perimeter, giving rise to an enforcement infrastructure in which procedural oversight is a presumptive condition of legitimacy across a broad swath of computational systems.

The Act's risk classification model compounds this effect. Article 6 and Annex III designate high-risk applications across critical domains such as employment, financial services, healthcare, law enforcement, and infrastructure. Although the regime formally allows for limited contextual exclusions, its application-based structure offers only constrained flexibility to reflect differences in technical sophistication, functional centrality, or epistemic risk. A system that filters CVs using keyword analysis may therefore attract governance obligations comparable to those imposed on fully autonomous hiring platforms. Such relative uniformity risks over-regulating low-stakes deployments while failing to engage substantively with the challenges posed by more complex systems that resist rigid typologies.[19]

Taken together, definitional breadth and risk-tiering construct a dense matrix of procedural obligations that demand not only legal interpretation but critical scrutiny of their symbolic function within corporate governance. These mechanisms channel organisational behaviour toward documentation, certification, and auditability, raising the possibility that compliance may become decoupled from substantive oversight.

[19] Ayushi Tripathi, *Study on Impact of AI for Recruitment and Selection of Skilled Professionals in the Organization* (May 4, 2024) (unpublished manuscript), SSRN, https://ssrn.com/abstract=4817290 (accessed May 8, 2025).

C. Transparency, Documentation, and the Risk of Performative Oversight

Transparency and documentation requirements further entrench this procedural logic. Articles 10 through 14 impose obligations to disclose training data sources, ensure model interpretability, and establish human oversight protocols. While grounded in legitimate accountability aims, these requirements place significant burdens on organisations already subject to disclosure-intensive regimes—most notably under the Corporate Sustainability Reporting Directive.[20] The intersection of overlapping frameworks heightens the risk of regulatory duplication, administrative fatigue, and escalating compliance costs.

More importantly for fiduciary doctrine, these documentation demands are symptomatic of a broader transformation: compliance with the AI Act increasingly entails structural governance recalibration. High-risk systems must undergo ex ante risk assessments, continuous post-market monitoring, and third-party conformity evaluations. These processes require substantial investments in personnel, technical infrastructure, and internal control mechanisms. Certification, reminiscent of the EU's medical device regime, is designed to ensure safety and reliability but may functionally constrain organisational agility.[21]

Yet even these safeguards risk institutionalising what Michael Power famously described as "rituals of verification."[22] Checklists, audit trails, and conformity assessments may satisfy regulatory expectations while leaving the internal logic of machine-learning systems opaque. In such settings, governance risks becoming a performance—formally compliant but substantively hollow. This dynamic is amplified for multinational enterprises forced to choose between harmonising global

[20] Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Directive 2013/34/EU, as regards corporate sustainability reporting, 2022 O.J. (L 322) 15.
[21] Juhamatti Huusko, Ulla-Mari Kinnunen & Kaija Saranto, *Medical Device Regulation (MDR) in Health Technology Enterprises – Perspectives of Managers and Regulatory Professionals*, 23 *BMC Health Serv. Rsch.* 310 (2023).
[22] Power, *Audit Society*, supra note 11.

governance frameworks around EU standards or maintaining fragmented, jurisdiction-specific regimes.[23]

The symbolic dimension of these practices is illuminated by Rose and Miller's account of "technologies of government": dashboards, metrics, and evaluative protocols that render complex socio-technical systems administratively visible.[24] These instruments do not merely reflect organisational practices; they actively construct them, translating ambiguity into apparent order. In doing so, they may obscure the very risks they purport to govern. The AI Act thus risks entrenching a model of procedural compliance that satisfies formal oversight requirements without ensuring that fiduciary judgment remains substantive and engaged.

D. Normativity, Technocracy, and the Fiduciary Counterweight

Crucially, the normative assumptions embedded in the AI Act's procedural mandates—what constitutes acceptable risk, bias, or meaningful human oversight—are not neutral. As Sheila Jasanoff has argued, it is essential to make explicit "the normative that lurks within the technical."[25] The AI Act encodes value-laden judgments about fairness, accountability, and control under the guise of technical governance.

This critique is not a rejection of the Act's framework. It is a warning against technocratic complacency in its implementation. Governance must not only be seen to function; it must function meaningfully. In this context, fiduciary duties assume renewed significance. As the AI Act and its transnational counterparts channel corporate behaviour through procedural incentives, directors must resist the gravitational pull of symbolic compliance. The duty of care must require interrogation

[23] Michel Foucault, *Governmentality*, in *The Foucault Effect: Studies in Governmentality* 87, 87–104 (Graham Burchell, Colin Gordon & Peter Miller eds., Univ. of Chi. Press 1991) (describing modern governance as operating through indirect techniques—such as statistics, surveillance, and visibility—that shape the conditions of action rather than issuing direct commands).
[24] Nikolas Rose & Peter Miller, *Political Power Beyond the State: Problematics of Government*, 43 *Brit. J. Soc.* 173, 183 (1992).
[25] Sheila Jasanoff, *Technologies of Humility: Citizen Participation in Governing Science*, 41 *Minerva* 223 (2003).

of the assumptions embedded in AI design and deployment. The duty of loyalty must entail stewardship over the normative integrity of the systems through which strategic decisions are mediated.

Regulatory developments outside the EU reinforce this trajectory. In the United States, proposed algorithmic accountability initiatives[26] and emerging SEC guidance emphasise design-time accountability and risk anticipation.[27] In the United Kingdom, the ICO promotes AI impact assessments,[28] while the FCA articulates expectations around model risk management.[29] Although these regimes vary in scope and enforceability, they converge around a shared normative orientation: organisations are expected to institutionalise governance mechanisms before harms materialise.

This convergence consolidates a global grammar of procedural legitimacy—dashboards, risk matrices, conformity assessments, internal model inventories. Yet these artefacts often operate at a remove from the systems they govern. The diffusion of such structures risks flattening normative complexity, treating socio-technical dilemmas as issues of procedural regularity rather than ethical and institutional judgment.

It is precisely here that fiduciary law must intervene. Fiduciary responsibility is not fulfilled by legal formalism alone. It requires reasoned discretion, critical reflection, and sustained engagement with systems whose normative implications exceed what technical documentation can capture. The AI Act thus serves as a revealing case study: not because it dictates fiduciary outcomes, but because it exposes the conditions under which fiduciary deference can no longer be assumed.

---

[26] *See*, e.g., Algorithmic Accountability Act of 2023, H.R. 5628, 118th Cong. (2023), https://www.congress.gov/bill/118th-congress/house-bill/5628 (accessed June 3, 2025).

[27] U.S. Securities and Exchange Commission, *AI Compliance Plan – Addressing Termination of Non-Compliant Applications* (2023), https://www.sec.gov/files/sec-ai-compliance-plan.pdf (accessed June 3, 2025).

[28] UK Information Commissioner's Office, Guidance on AI and Data Protection & AI and Data Protection Risk Toolkit, ICO (2023), https://ico.org.uk/ (accessed 10 February).

[29] Financial Conduct Authority, FCA Approach to AI, FCA (updated Dec. 5, 2025), https://www.fca.org.uk/ (accessed 10 February).

In this setting, fiduciary law becomes the site at which procedural legitimacy must be translated back into accountable judgment—if judicial deference is to remain normatively defensible.

II. The Role of Directors in Complying with the AI Act

The enactment of the AI Act constitutes a regulatory inflection point for corporate governance within the European Union. Although the legislation does not expressly address company boards, its ramifications for directors' roles and legal obligations are both far-reaching and profound. In particular, the integration of AI systems—especially those classified as high-risk—necessitates a reconceptualization of directors' fiduciary duties, requiring not only adherence to general corporate standards but also compliance with the specific mandates of the AI Act.

Rooted in Articles 16 and 114 of the Treaty on the Functioning of the European Union (TFEU), the AI Act pursues a dual mandate: the protection of fundamental rights (including data protection, non-discrimination, and human dignity), and the establishment of a functional internal market for AI technologies. As such, the responsibilities of corporate directors extend beyond ethical discretion or strategic foresight; they directors are increasingly exposed to legal and governance expectations requiring them to ensure that AI deployment coheres with the regulatory objectives codified in the AI Act. These obligations encompass, inter alia, due diligence, transparency, and risk governance. Accordingly, the implementation of AI technologies has shifted from a matter of operational optimisation to a domain of legal accountability.[30]

This elevation of legal responsibility is most clearly manifest in the regulation of high-risk AI systems—those deployed in employment, financial services, critical infrastructure, or consumer

[30] Lucía Bosoer, Marta Cantero Gamito & Ruth Rubio-Marín, *Non-Discrimination and the AI Act*, in *Law and Digitalization* (Marco Casadei ed., Arazandi 2023), https://ssrn.com/abstract=4666071 (accessed May 8, 2025); Simona Demková, *The EU's Artificial Intelligence Laboratory and Fundamental Rights*, in *Redressing Fundamental Rights Violations by the EU: The Promise of the 'Complete System of Remedies* (Melanie Fink ed., Cambridge Univ. Press forthcoming), https://ssrn.com/abstract=4566098 (accessed May 8, 2025); Samuele Bertaina et al., *Fundamental Rights and Artificial Intelligence Impact Assessment: A New Quantitative Methodology in the Upcoming Era of AI Act*, 56 *Comp. L. & Sec. Rev.* 106101 (2025).

interfaces—which frequently reside within a firm's core operations. Owing to their potential to impinge upon fundamental rights and undermine market fairness, such systems are subject to a heightened regulatory regime. Their lawful use presupposes the establishment of internal audit mechanisms and human-centric oversight structures designed to preserve reviewability and ensure accountability. Article 14 imposes continuous oversight obligations across the AI system's lifecycle, while Article 29 mandates a fundamental rights impact assessment prior to deployment, thereby obliging directors to pre-emptively identify and mitigate risks to dignity, data protection, and equality.[31]

Taken together, these provisions construct a compliance architecture that is both anticipatory and retrospective. Impact assessments function as ex ante safeguards against harm, particularly in domains such as employment, credit, or essential services, while Article 25 imposes rigorous ex post record-keeping obligations, including the preservation of technical specifications, datasets, and decision-making logic. This interplay of anticipatory diligence and evidentiary rigour makes governance not only legally traceable but normatively grounded. From a liability standpoint, these mechanisms intensify the directors' duty to demonstrate that organisational conduct meets the diligence threshold embedded in the AI Act's regulatory framework.

Non-compliance with these standards exposes both firms and individual directors to an array of legal and reputational risks. These range from administrative sanctions and civil liability to heightened regulatory scrutiny and diminished stakeholder trust. One salient area of exposure is the failure to evidence proactive compliance—for example, by neglecting documentation or omitting risk assessments—which may be construed as symptomatic of systemic governance failures. Where AI-generated outcomes produce harm, directors may face claims of dereliction of oversight duties, particularly if adequate review, monitoring, or mitigation protocols are absent.

---

[31] Samuele Bertaina et al., *Fundamental Rights and Artificial Intelligence Impact Assessment: A New Quantitative Methodology in the Upcoming Era of AI Act*, 56 *Comp. L. & Sec. Rev.* 106101 (2025).

Accordingly, the AI Act's operational impact must be situated within the directors' sphere of responsibility. Two domains emerge as particularly salient: workforce management and consumer-facing operations. While not exhaustive, these fields illustrate how the AI Act reconfigures directors' obligations at the intersection of legal compliance and strategic implementation.

The first domain—algorithmic workforce management[32]—illustrates the legal intricacies that arise when AI is used in recruitment, performance evaluation, surveillance, or termination. While such applications offer efficiency, they are classified as high-risk systems under Article 6 and Annex III, and thereby subject to elevated scrutiny. This designation reflects the disruptive potential of AI to recalibrate employer–employee relations, skew informational symmetry, and erode procedural fairness.

In routine recruitment scenarios, accountability may be diffused; however, where AI is used in high-stakes hiring—such as executive appointments or board succession—directors are more directly exposed to scrutiny. In such settings, the plausibility of judicial deference depends not on outcome neutrality, but on whether directors can demonstrate substantive engagement with the systems mediating their decisions.

Article 10 requires that such systems uphold principles of fairness and non-discrimination, which implies robust safeguards in model training, explainability of outcomes, and rights of contestation for affected candidates.

Particularly contentious is the use of AI for performance monitoring and surveillance. Tools such as keystroke monitoring and emotion detection software intensify concerns over privacy, autonomy, and individual dignity. Article 6 designates real-time employee evaluation systems as high-risk, while certain practices—such as emotion recognition—are subject to categorical bans save for narrowly defined exceptions. Article 26 further imposes implementation duties, including advance disclosure,

[32] For a definition, see Simone Baiocco et al., *The Algorithmic Management of Work and Its Implications in Different Contexts* (Eur. Comm'n Joint Rsch. Ctr. 2022), JRC129749.

adherence to provide guidance, and human supervision under Article 14. These obligations are reinforced by jurisprudence from the European Court of Human Rights, notably *López Ribalda v. Spain*,[33] which underscores the impermissibility of covert surveillance in violation of Article 18 ECHR.[34]

Corporate governance must therefore ensure that compliance is not merely formal, but substantively protective of labour rights. This entails providing workers with accessible channels to challenge AI-generated assessments and precluding coercive surveillance practices. Failure to establish transparency mechanisms, appeal procedures, and fairness audits risks infringing on rights enshrined in data protection and employment law.

Disciplinary or dismissal decisions effected via AI merit equal attention. Article 50 of the AI Act demands explainability and fairness, undergirded by human oversight.[35] Given the EU Charter's protection against unfair dismissal, directors must verify that such decisions meet procedural and normative thresholds. This may necessitate consultation with trade unions and the institution of internal review structures.

From a fiduciary perspective, the adoption of measures such as AI ethics boards, external audits, and targeted governance training functions less as a checklist of best practices than as evidence of substantive oversight. Given the convergence of risk across legal regimes—from GDPR to employment law—a proactive, multi-tiered approach to governance is essential.[36]

---

[33] López Ribalda v. Spain, Apps. Nos. 1874/13 & 8567/13, Eur. Ct. H.R. (Jan. 9, 2018). See also Michele Molè, *The Quest for Effective Fundamental Labour Rights in the European Post-Pandemic Scenario: Introducing Principles of Explainability and Understanding for Surveillance Through AI Algorithms and IoT Devices* (19th Int'l Conf. in Commemoration of Marco Biagi, *Work Beyond the Pandemic: Towards a Human-Centred Recovery*, May 1, 2022), Univ. of Groningen Faculty of Law Rsch. Paper No. 26/2022, https://ssrn.com/abstract=4099663.

[34] Prohibition F states that an AI system for emotion recognition must infer emotions of natural persons in the areas of workplace and education institutions. An AI system for emotion recognition in situations related to the workplace and education is also covered by this prohibition.

[35] Thomas Gils, *A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act* (June 14, 2024) (unpublished manuscript), forthcoming in *The EU Artificial Intelligence (AI) Act: A Commentary* (C. N. Pehlivan, N. Forgó & P. Valcke eds., Kluwer L. Int'l), https://ssrn.com/abstract=4865427.

[36] As to the privacy aspects, see Graham Greenleaf, *EU AI Act: The 2nd Most Important Data Privacy Law* (May 30, 2024) (unpublished manuscript), https://ssrn.com/abstract=4913686; Nathalie A. Smuha, *The Paramountcy of Data Protection Law in*

The second key domain is consumer-facing operations, encompassing both algorithmic profiling and dynamic market strategies. AI technologies used for personalisation, recommendation, or pricing shape public engagement and may influence access to core goods and services. Article 5 prohibits manipulative or subliminal techniques, especially those targeting vulnerable individuals, while Article 50 mandates disclosure when consumers interact with AI. These provisions embed a normative commitment to autonomy and fair dealing, demanding that oversight systems guarantee transparency, contestability, and redress.

Directors must therefore ensure that AI systems used in consumer settings are intelligible, subject to human override, and do not exploit behavioural bias. Risk can be mitigated through early regulatory engagement and the implementation of public-facing compliance standards. Such measures may also support a cooperative posture during audit or enforcement procedures.[37]

Finally, the AI Act confers investigatory and supervisory powers upon national authorities, heightening expectations around procedural readiness and internal control. For directors, this reinforces the imperative to institutionalise compliance as a core function of governance—not through mere documentation, but through substantive alignment with legal and ethical obligations. Investment in AI compliance personnel, legal oversight, and technical governance structures is no longer an optional efficiency measure. It is, rather, an expression of fiduciary responsibility: a duty to ensure that algorithmic decision-making unfolds within an accountable, intelligible, and rights-respecting institutional framework.

What emerges is not a conflict between law and technology, but a growing misalignment between fiduciary deference and the conditions on which that deference has historically relied.

---

the Age of AI (Acts), in *Two Decades of Personal Data Protection. What Next?* (EDPS ed., Publ'ns Off. of the Eur. Union 2024), https://ssrn.com/abstract=4874388.

[37] Tara K. Giunta & Lex Suvanto, *Board Oversight of AI* (Harv. L. Sch. F. on Corp. Governance, Sept. 17, 2024), https://corpgov.law.harvard.edu/2024/09/17/board-oversight-of-ai/ (accessed May 8, 2025).

III. Towards a Duty of AI Due Care

AI due care is not offered as a freestanding doctrinal category. It is a heuristic—an analytical lens for re-reading the duty of care under conditions of algorithmic mediation. The claim is that AI systems alter the cognitive and informational architecture of board decision-making, thereby re-specifying what "reasonable care" plausibly requires and what courts can normatively presume when extending deference.

Rather, it uses AI due care as an analytical lens through which to reinterpret established fiduciary standards under altered cognitive and informational architecture surrounding board decision-making and created by algorithmic decision-making. This evolving standard re-specifies the content of directors' duties under conditions of algorithmic mediation. While corporate law has long imposed a general duty of care upon directors—requiring that they act with the diligence and skill reasonably expected of someone in their position[38]—this new form of care calls for the oversight of a qualitatively different category of risk: systemic, technical, and often inscrutable.

This reconceptualisation responds to a context unlike that contemplated by traditional fiduciary frameworks. As Paul Humphreys has shown, many AI systems exhibit epistemic opacity:[39] their internal operations resist explanation, even to those responsible for their use. In corporate governance, this opacity undermines the premise that directors can exercise meaningful oversight

[38] Carsten Gerner-Beuerle, Philipp Paech & Edmund-Philipp Schuster, *Study on Directors' Duties and Liability* 74–118 (London Sch. of Econ., Apr. 2013) (prepared for the Eur. Comm'n, DG Internal Mkt. & Servs.), https://eprints.lse.ac.uk/50438/1/__Libfile_repository_Content_Gerner-Beuerle%2C%20C_Study%20on%20directors%E2%80%99%20duties%20and%20liability%28lsero%29.pdf (providing a comprehensive comparative analysis of directors' duties, liability regimes, enforcement mechanisms, and cross-border issues across EU Member States aimed at informing EU corporate governance policy).

[39] On the concept of epistemic opacity, see Paul Humphreys, *The Philosophical Novelty of Computer Simulation Methods*, 69 *Phil. Sci.* 615, 618–19 (2009); Eva Schmidt, Paul Martin Putora & Rianne Fijten, *The Epistemic Cost of Opacity: How the Use of Artificial Intelligence Undermines the Knowledge of Medical Doctors in High-Stakes Contexts* (2023), https://cris.maastrichtuniversity.nl/en/publications/the-epistemic-cost-of-opacity-how-the-use-of-artificial-intellige (accessed May 8, 2025).

through deliberation and informed inquiry. As algorithmic systems shape more decisions, the conditions of fiduciary judgment shift, no longer grounded in human understanding but increasingly mediated by inscrutable models. Socio-legal accounts of non-human agency underscore how technical systems may shape institutional decision-making without being directly accountable.[40] The duty of AI due care must therefore confront procedural adequacy, as well as the shifting locus of authority and expertise away from human deliberation toward machine-led inference. This article shows that the normative justification for deference becomes increasingly contestable where algorithmic reliance substitutes for, rather than mediates, human judgment.

Hence, it demands that directors engage not only with legal and commercial concerns but also with the design, function, and governance implications of AI. These AI systems are not merely tools—they are decision-making structures with systemic effects, mediating organisational choices in ways that often elude human comprehension, auditability, and normative clarity. Directors are therefore expected to govern not only the outcomes of AI deployment, but also the internal logics, limitations, and unintended consequences of these systems. This requires approaching AI not as a discrete technical solution but as a strategic and procedural constraint—one that implicates board-level accountability and long-term corporate legitimacy.

Viewed through the lens of AI due care, the traditional duty of care spans legal compliance, ethical alignment, technical literacy, and strategic foresight. It is grounded in the recognition that AI technologies are not ethically neutral or operationally autonomous: they are deployed, shaped, and constrained by institutional choices. Accordingly, directors are cast not as passive overseers, but as accountable custodians of technological governance.

[40] Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* 128 (Oxford Univ. Press 2005).

A. The AI Due Care's First Dimension: Technological Literacy as a Fiduciary Imperative

Directors are no longer mere stewards of financial performance or strategic intent; they are gatekeepers of systems whose logics are often inaccessible, whose outputs may elude retrospective justification, and whose long-term consequences remain structurally unpredictable. In this evolving landscape, *technological literacy* must be recast not as an ancillary skill, but as a fiduciary imperative: a threshold condition for lawful and legitimate corporate oversight.

To be clear, the expectation is not technical fluency. The law does not ask directors to parse code or audit neural architectures. Rather, the emerging standard is one of cognitive adequacy—a duty to develop, or to procure, the interpretive capacity necessary to assess whether algorithmic systems are consistent with the organisation's objectives, ethical commitments, and legal constraints. This includes familiarity with model typologies, data provenance, fairness metrics, and the governance implications of opacity. Directors must not only ask the right questions; they must know which questions matter.[41]

This shift challenges the presumptions that underpin doctrines like the BJR.[42] Where traditional governance models rested on the assumption of bounded rationality—that directors could make decisions under conditions of partial information and uncertainty—the introduction of AI tools has created conditions of unbounded abstraction. Directors now confront systems whose outputs cannot always be explained, whose optimisation goals may not be fully understood, and whose behaviour may evolve post-deployment. This dynamic introduces a delegation paradox: as directors delegate critical functions to systems, they do not fully comprehend the foundational premise of independent judgement—so central to doctrines such as the BJR—begins to erode.

---

[41] Michael W. Peregrine & Alya Sulaiman, *The Argument for Strong Board Oversight of Artificial Intelligence* (CLS Blue Sky Blog, July 28, 2023), https://clsbluesky.law.columbia.edu/2023/07/28/the-argument-for-strong-board-oversight-of-artificial-intelligence/.

[42] Langenbucher, *supra* note 1, 10–11.

This erosion of judgment is compounded by a tendency, noted by Sheila Jasanoff, to place excessive trust in what she terms "technologies of hubris"—systems that promise precision, objectivity, and control while masking the normative choices and uncertainties embedded in their design.[43] In corporate governance, this mindset risks transforming technological literacy into a false proxy for diligence. Directors may feel reassured by the presence of advanced tools, dashboards, and compliance protocols, even as those systems encode opaque assumptions that go unchallenged. Fiduciary oversight thus requires more than comprehension: it demands a posture of critical interrogation, aware that algorithmic objectivity is often constructed rather than neutral.

The normative implications of this shift are already visible in statutory frameworks. Under Section 174 of the UK Companies Act 2006, directors must exercise the care, skill and diligence of a reasonably diligent person, judged according to both an objective standard and the specific knowledge and experience of the director in question.[44] This dual limb must now be read expansively. Objectively, all directors must possess a baseline capacity to interrogate the operational and ethical implications of AI deployment. The significance of this claim does not lie in the content of the standard—which remains formally unchanged—but in its *function*. What changes is not what directors must do, but what courts can legitimately presume when extending deference. Where directors lack the capacity to interrogate algorithmic systems, the assumption that judgment has been exercised collapses, even if formal procedures are observed. Subjectively, those with technical or delegated authority bear an elevated duty to ensure that AI systems are not merely effective, but *governed*—that their design, oversight, and auditability align with principles of procedural fairness and accountable discretion.

These obligations are not abstract. The risks are structural and cumulative. Consider the phenomenon of algorithmic path dependency, whereby AI-generated outputs—once routinised—

[43] Jasanoff, *Technologies of Humility*, supra note 25.

[44] Gerner-Beuerle et al., *Study on Directors' Duties and Liability*, supra note 38, 91-92.

begin to shape internal metrics, strategic direction, and boardroom expectations.[45] Such feedback loops can entrench decisions that appear rational in isolation, but which, over time, harden into unexamined orthodoxy. The boardroom, far from acting as a site of independent scrutiny, risks becoming an echo chamber that merely amplifies the logics embedded in code—logics that directors do not fully understand, yet may allow to shape strategic judgment unchallenged.

Algorithmic systems introduce opacity that complicates the exercise of fiduciary judgment and weakens the assumptions underpinning deference. Consider a stylised but increasingly plausible scenario: being trained on historical performance data, the system gradually deprioritises rural areas and reallocates human resources toward more densely populated zones. Efficiency metrics improve. Profit margins climb. The board is pleased. Yet no inquiry is made into the demographic consequences of these shifts; no fairness audit is conducted; no scenario planning addresses the risk of regulatory or reputational blowback. Months later, a parliamentary inquiry exposes a pattern of geographic exclusion and employment suppression. When called to account, the board cannot reconstruct how the system's parameters were set, what values they reflected, or why no corrective governance structures were in place. The failure is not of intent but of structure—a collapse of fiduciary attention in the face of technological abstraction.

This is not an argument for perfection. It is an argument for procedural vigilance. Directors must build governance architectures capable of scrutinising systems that are simultaneously powerful and inscrutable. This means ensuring the availability of independent advice, instituting internal review protocols, mandating traceability of outputs, and demanding narrative justifications alongside statistical results. It also means resisting the allure of automation as a proxy for objectivity. Fiduciary

[45] Derek E. Bambauer & Michael Risch, *Worse Than Human?*, 53 *Ariz. St. L.J.* 1091 (2021) (examining the comparative risks and accountability implications of AI decision-making versus human judgment in legal and regulatory contexts); see also Derek E. Bambauer & Michael Risch, *Worse Than Human?* (TPRC49: The 49th Rsch. Conf. on Commc'n, Info. & Internet Pol'y), Ariz. Legal Stud. Discussion Paper No. 21-22, https://ssrn.com/abstract=3897126.

responsibility cannot rely on surface metrics or assumed technical validity. Algorithmic recommendations are not facts; they are artefacts—constructed, contingent, and contestable. Where boards fail to interrogate these systems, they do not merely incur risk. They abdicate responsibility.[46] And in doing so, they jeopardise not only the firm's interests but the legal fabric that grants them authority. For fiduciary law is not indifferent to method. It recognises that how decisions are made matters as much as what those decisions are. And in an era of machine-led governance, that principle must be reaffirmed—not in code, but in conduct.

B. The AI Due Care's Second Dimension: A Proactive and Forward-Looking Risk Governance

Whereas legal risks have never been entirely static, AI systems introduce a new class of risk—emergent, recursive, and feedback-sensitive—whose evolution is often triggered by real-time interactions, data volatility, and opaque system behaviours.[47] This dynamic quality—often masked by AI washing—challenges traditional compliance framework and demands anticipatory governance models.[48] Under conditions of algorithmic mediation, a reasonable board process increasingly entails implementing anticipatory frameworks that embed risk assessment, human-in-the-loop oversight, and clear escalation pathways throughout the AI lifecycle. In high-risk domains—as delineated by Article 6 and Annex III of the AI Act—this includes third-party conformity assessments, continuous post-market monitoring, and robust internal accountability mechanisms. These are not merely technical requirements but instruments of fiduciary risk stewardship. Importantly, this governance

[46] Maria Lucia Passador, *Governance Gambits and Business Judgment in In/Out-Sourcing Tactics*, forthcoming in *Am. Bus. L.J.* (2025) (arguing that boards risk becoming "theatre boards" in which formal adherence substitutes for meaningful oversight, and that uncritical reliance on external consultants hollows out fiduciary responsibility by displacing substantive judgment).

[47] David Rhinesmith, Brad Marcus & Sarah Schaedler, *AI Washing Enforcement Continues, Highlighting Risks to Companies and Investors* (Harv. L. Sch. F. on Corp. Governance, July 19, 2024), https://corpgov.law.harvard.edu/2024/07/19/ai-washing-enforcement-continues-highlighting-risks-to-companies-and-investors/ (accessed May 8, 2025).

[48] Natalia Díaz-Rodríguez et al., *Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation*, 99 *Info. Fusion* 101896 (2023), https://doi.org/10.1016/j.inffus.2023.101896.

approach aligns with the UK duty to promote the success of the company under section 172 of the UK Companies Act 2006, which mandates consideration of long-term consequences and the interests of broader stakeholder constituencies.[49] In the AI context, stakeholders may include not only shareholders, but also data subjects, employees, end-users, and communities affected by algorithmic decisions. Accordingly, AI governance must shift from narrow legal defensibility to a more participatory and ethically attuned model of risk anticipation—one that reflects the normative ambitions of the AI Act and accounts for who may be affected, in what ways, and under what systemic conditions.

The contours of the latter may be further delineated by noting that, under U.S. corporate law, the duty of oversight has undergone substantial development since *In re Caremark International Inc. Derivative Litigation*,[50] in which the Delaware Chancery Court held that directors only breach their fiduciary duties where they wholly fail to establish any reporting or information systems, or to monitor such systems once in place. Although described as "possibly the most difficult theory in corporation law upon which a plaintiff might hope to win a judgment," subsequent Delaware case law has significantly broadened its practical application.[51] This expansion is most clearly exemplified in *Marchand v. Barnhill*, where the Delaware Supreme Court found that directors of a food company had breached their oversight duties by failing to implement any board-level mechanism to monitor food safety—a "mission-critical" risk.[52] Notably, the court emphasised that the mere absence of structured oversight—even in the absence of specific red flags—may suffice to establish liability. In

[49] David Kershaw & Edmund Schuster, *The Purposive Transformation of Corporate Law*, 69 Am. J. Comp. L. 478, 488–89 (2021), https://doi.org/10.1093/ajcl/avac004.

[50] In re Caremark Int'l Inc. Derivative Litig., 698 A.2d 959 (Del. Ch. 1996).
See also Robert C. Bird, *Caremark Compliance for the Next Twenty-Five Years*, 58 Am. Bus. L.J. 63 (2021).

[51] Gregory A. Markel, Daphne Morduchowitz & Matthew C. Catalano, *A Director's Duty of Oversight after Marchand in "Caremark" Case* (Harv. L. Sch. F. on Corp. Governance, Jan. 23, 2022), https://corpgov.law.harvard.edu/2022/01/23/a-directors-duty-of-oversight-after-marchand-in-caremark-case/ (accessed May 8, 2025);

[52] Katherine M. King, *Marchand v. Barnhill's Impact on the Duty of Oversight: New Factors to Assess Directors' Liability for Breaching the Duty of Oversight*, 62 B.C. L. Rev. 1925 (2021).

the context of AI governance, *Marchand* underscores the imperative for boards not only to remain vigilant in the face of known risks but also to proactively structure oversight mechanisms around foreseeable, high-impact areas of corporate exposure, including the deployment of opaque algorithmic systems in decision-making processes.

A comparable emphasis on proactive oversight is also present in continental European jurisdictions. Under German corporate law, §93(1) of the *Aktiengesetz* (AktG) imposes a rigorous standard of care on management board members (*Vorstandsmitglieder*), requiring them to act with the diligence of a prudent and conscientious director ("*die Sorgfalt eines ordentlichen und gewissenhaften Geschäftsleiters*").[53] This standard—interpreted by the *Bundesgerichtshof* (BGH) as encompassing both organisational duties (*Organisationspflichten*) and internal monitoring obligations[54] is likely to extend to technological risks in complex operating environments. This includes a duty of oversight (*Aufsichtspflicht*), which obliges directors to monitor the company's affairs, internal controls, and compliance mechanisms on an ongoing basis. The German *Aufsichtspflicht* differs from the Delaware model in its statutory formalism and codified compliance expectations, but it converges in function: directors are expected to establish and maintain structures that detect and mitigate risks, particularly when those risks arise from systems that may not be easily intelligible or auditable.

Algorithmic systems challenge the assumptions underpinning traditional frameworks, as they often operate in ways that defy intuitive understanding, resist effective auditing, and evolve in real time. The U.S. model now mandates affirmative steps to ensure board-level engagement with operational risk, while the German regime imposes ex ante obligations to establish internal safeguards through

[53] Gerner-Beuerle et al., *Study on Directors' Duties and Liability*, supra note 38, at A323; Tamo Zwinge, *Have Directors' Duties of Care and Skill Become More Stringent? What Has Driven This Development? Is This Development Beneficial? An Analysis of the Duty of Care in the UK in Comparison to the German Duty of Care* (Oct. 20, 2009) (unpublished manuscript), https://ssrn.com/abstract=1591590 (accessed June 7, 2025).

[54] For the standard of care under § 93(1) AktG, see BGH [Bundesgerichtshof] Apr. 21, 1997, II ZR 175/95 (*ARAG/Garmenbeck*) (Ger.); BGH Dec. 6, 2001, 1 StR 215/01 (*SSV Reutlingen*) (Ger.); BGH June 19, 2012, II ZR 243/11 (Ger.); BGH Oct. 12, 2016, 5 StR 134/15 (*HSH Nordbank*) (Ger.); BGH Jan. 27, 2021, II ZR 84/20 (Ger.). See also BGH Mar. 25, 1996, II ZR 263/94 (*Siemens/Nold*) (Ger.).

defined structures of responsibility and documentation. In both jurisdictions, the trend is toward institutionalised vigilance over increasingly inscrutable systems. So, while doctrinal labels vary—*oversight*, *Sorgfaltspflicht*, *Aufsichtspflicht*—, these doctrines respond to analogous governance pressures, albeit through distinct institutional logics. Courts in both common law and civil law worlds are increasingly holding directors to a standard that is not satisfied by structural formalism alone, but requires affirmative inquiry into the *substance, limitations, and risks* of complex technologies. As AI continues to permeate corporate decision-making infrastructures, fiduciary accountability will be redefined less by traditional categories and more by the director's capacity to steward systems whose consequences they can neither delegate nor defer.

The duty of AI due care synthesises and extends these developments. It incorporates the Delaware court's insistence on "mission-critical" risk oversight, while aligning with the German emphasis on proactive institutional design. In addition, it also introduces a distinctive requirement: *epistemic engagement with algorithmic systems themselves*. In contrast to the reactive or structural orientation of traditional oversight doctrines, AI due care demands *content-level awareness* of how decision-making systems function, where opacity arises, and what accountability structures are in place. It is not enough that systems exist—they must be understood and challengeable.

Taken together, these comparative insights reveal a pressing need to recalibrate the application of the directors' duty of care in response to the specific challenges posed by AI. This reframing gives rise to a converging justificatory logic across fiduciary systems, which is essential to ensure that corporate governance remains both legally credible and operationally effective in the age of algorithmic risk

C. The AI Due Care's Third Dimension: Embedding Rule-of-Law Principles in AI Oversight

Although the AI Act is technical in structure, it is underpinned by constitutional values such as transparency, non-discrimination, explainability, and human dignity.

Directors who rely uncritically on algorithmic outputs—or delegate oversight entirely to technical teams—risk breaching fiduciary duties that now encompass the legitimacy of deployed systems. As established in the analysis above, corporate rules increasingly recognized that directors must take an active role in supervising complex operational systems, particularly where their failure may entail legal liability or ethical harm. In the AI context, this expectation encompasses systems that may embed bias, generate opacity, or undermine procedural fairness.

Although the duty of AI due care has not yet been codified as a distinct legal obligation, its contours are increasingly discernible through the regulatory logic of the AI Act, the trajectory of fiduciary jurisprudence, and the growing weight of soft law instruments. The AI Act does not operate merely as a technical instrument—it functions as a governance regime, one that implicitly requires both boards and corporate roles at large to calibrate internal oversight structures around algorithmic risk.[55] It extends the concept of care beyond financial prudence into the domains of information architecture, procedural legitimacy, and socio-technical design.

This evolution is reinforced by a proliferating corpus of soft law frameworks—notably the OECD AI Principles, ISO/IEC 42001, and national ethical guidelines—that offer practical models for risk assessment, transparency, and accountability. While formally non-binding, these instruments increasingly shape regulatory expectations and judicial reasoning, much like the role of the UK Corporate Governance Code in ESG contexts.[56] They serve as interpretive benchmarks for what constitutes responsible AI oversight, enabling firms to pre-empt legal uncertainty and demonstrate good-faith governance in a rapidly evolving regulatory landscape.

---

[55] On the latter topic, see Maria Lucia Passador, *The AI Act's Silent Impact on Corporate Roles* (Mar. 7, 2025) (unpublished manuscript), https://ssrn.com/abstract=5189809.

[56] Andrew Johnston, *ESG and Corporate Sustainability: A View from the UK*, 52 Ga. J. Int'l & Comp. L. 576, 601 (2024); Constantin Jung, *Sustainable Corporate Governance in the United Kingdom: Environmental Sustainability in Directors' Decision-Making*, EUI RSC Working Paper No. 2023/25, at 9 (2023), https://hdl.handle.net/1814/76113.

Boards that fail to engage with these instruments may find themselves ill-prepared to meet both legal scrutiny and public accountability. Conversely, proactive incorporation of international standards can serve as both shield and compass: insulating firms from compliance failures while guiding ethical orientation and strategic alignment. The fiduciary dimension of soft law lies in this anticipatory function—allowing directors to translate emerging norms into operational governance practices before hard law intervenes.

In such settings, formal compliance is no longer sufficient. Directors must ensure that AI systems are deployed with attention to fairness metrics, demographic impacts, and data integrity. This requires translating normative principles into concrete safeguards: audit trails, explainability thresholds, accountability mapping, and effective recourse mechanisms.

This evolving duty also implies structural adaptations at board level because those who do engage—actively, critically, and ethically—position their organisations to lead in an environment where technological integrity is rapidly becoming a cornerstone of corporate legitimacy. In particular, firms could consider mandating technological competence within the board or its delegated committees, ensuring directors can exercise meaningful oversight of AI systems.

Establishing dedicated 'AI governance committees'—distinct from generalist technology committees[57]—may further institutionalise responsibility for monitoring algorithmic risk, supervising third-party providers, and aligning AI deployment with corporate strategy. Crucially, AI oversight must be integrated into the board's informational flows, embedding periodic reporting and risk assessments into regular deliberations. These adjustments move the duty of AI due care from abstract expectation to embedded governance practice.

For analytical clarity, this article focuses on a single paradigmatic setting: the use of AI systems in employment- and credit-related decision-making.

---

[57] Maria Lillà Montagnani & Maria Lucia Passador, *Toward an Enhanced Level of Corporate Governance: Tech Committees as a Game Changer for the Board of Directors*, 15 J. Bus. Entrepreneurship & L. 1 (2022).

These domains concentrate the fiduciary tensions raised by algorithmic governance, as they combine high-stakes discretion, fundamental rights exposure, and board-level accountability In employment-related decision-making, the fiduciary implications of algorithmic reliance are particularly acute. AI systems used for recruitment screening, performance evaluation, or termination decisions operate at the intersection of strategic discretion, legal compliance, and fundamental rights protection. Under the AI Act, such systems are classified as high-risk, triggering heightened requirements of transparency, human oversight, and ex ante risk assessment.

For directors, reliance on these systems therefore cannot be treated as a neutral operational choice: it becomes a governance decision that directly conditions the legitimacy of fiduciary deference.

In each case, the duty of AI due care transforms abstract governance into concrete oversight practices—anchoring fiduciary responsibility in the technical, legal, and ethical scrutiny of algorithmic systems.

D. From Non-Compliance to Implementation: Operationalising the Duty of AI Due Care

Having articulated AI due care as an enhanced fiduciary standard, this Part turns from justification to breach: what, concretely, does failure look like in an AI-mediated governance environment? The point is not to construct a new liability regime, but to specify the failure modes through which AI due care becomes legally legible—particularly under doctrines of deference such as the BJR.

The analysis proceeds in three steps. First, it offers a functional taxonomy of AI governance failures that maps onto the AI Act's procedural architecture (notably Articles 9–12 and 14). Second, it isolates omission as a transversal mode of fiduciary breach—one that renders oversight failure actionable even absent affirmative misconduct. Third, it briefly traces the downstream institutional reverberations of these failures in adjacent domains such as insurance underwriting, disclosure practice, and internal controls.

The taxonomy distinguishes three analytically distinct, though often overlapping, categories: procedural neglect, epistemic abdication, and structural complicity.

Procedural neglect concerns the absence (or inadequacy) of the governance structures and documentation expressly mandated under the AI Act. These include failures to establish and maintain effective risk management systems, neglecting to ensure the quality and relevance of training, validation, and testing data, or omitting to keep complete and up-to-date technical documentation and logs (Arts 9–12).[58] While sometimes perceived as mere compliance checklists, these provisions function as codified proxies for meaningful board oversight. Their breach exposes not only compliance vulnerabilities but also erodes the procedural integrity upon which courts may base deference under standards such as the BJR.

Epistemic abdication, by contrast, reflects a breakdown not of form, but of substance. It arises when directors fail to interrogate the design assumptions, limits, and operational consequences of the AI systems deployed. The AI Act repeatedly signals the importance of informed human involvement—nowhere more so than in its requirement for human oversight mechanisms that are both "effective" and "commensurate" to the system's intended purpose and risk (Article 14). A board that fails to engage with this standard—whether by uncritically accepting vendor assurances, misunderstanding the system's learning capacity, or ignoring probabilistic uncertainty—ceases to exercise judgment in any meaningful sense. Recent literature underscores this form of abdication as a normative failure, where deference becomes unwarranted because discretion was never truly exercised.[59]

[58] Maria Lucia Passador, *AI in the Vault: AI Act's Impact on Financial Regulation* 14–15 (Bocconi Legal Stud. Rsch. Paper No. 4898828, July 1, 2024), forthcoming in 56 Loy. U. Chi. L.J. (2025), https://ssrn.com/abstract=4898828 (accessed June 7, 2025).

[59] Passador, *Governance Gambits*, supra note 46, at 3-4, 6, 18-19. (discussing how proceduralisation may hollow out substantive board discretion, leading to performative "theatre boards" and arguing that judicial deference under the BJR may become unwarranted when directors follow form without true deliberation, effectively amounting to a normative abdication of judgment.)

Most severe, however, is structural complicity: the deliberate tolerance or strategic deployment of systems whose risks are not only foreseeable, but also underregulated or unmitigated. This category includes decisions to adopt high-risk AI systems without conducting the required conformity assessment, to ignore post-market monitoring and incident reporting obligations, or to rely on providers whose governance and quality systems are manifestly insufficient. It also extends to tacit acceptance of opacity and bias where clearer alternatives exist—thus frustrating the AI Act's broader objective of ensuring that AI does not undermine fundamental rights.[60] This tripartite taxonomy is not intended merely as an ex post analytical device. Rather, it serves as a forward-facing grammar through which boards may structure their internal oversight cultures. It affirms that fiduciary governance in the algorithmic age cannot be reduced to technical fluency or formal compliance; it requires a deliberate engagement with design, risk, and accountability. Just as financial regulation once redefined directors' duties in response to systemic crises, so too must the AI Act reshape governance practices to ensure that discretion remains both legitimate and reviewable. A taxonomy of failure, in this sense, is a blueprint not for blame, but for fidelity—to law, to process, and to the human values that corporate governance is meant to uphold.

In addition to the individual categories of governance failure, it is essential to foreground a dimension that is often underestimated in doctrinal and practical discourse: the liability-bearing potential of omissions. In the context of AI governance, and particularly under the evolving contours of the duty of AI due care, omission is not a residual category, but a primary modality through which fiduciary duties may be breached. Its juridical significance lies not merely in what is done poorly, but in what is *not done at all*—the silence, the delay, the failure to inquire, to escalate, to record, or to intervene.

Contemporary fiduciary law has progressively eroded the historical bias in favour of act-based wrongdoing, especially in domains marked by structural complexity, systemic risk, and asymmetrical

[60] *See* AI Act, Arts 17–18, 43–51, 61–62; recitals 5, 38, and 47; and Title II.

information flows. In these settings, a failure to act with the vigilance, procedural foresight, or curiosity that the situation reasonably demands may be construed as a breach of the director's oversight duty, even absent malice, recklessness, or personal gain. This principle is not novel. It draws legitimacy from traditional doctrines of corporate law, such as the *Caremark* line of jurisprudence in Delaware, which explicitly recognises that a board's failure to institute and monitor appropriate information systems may trigger liability even in the absence of affirmative misconduct.[61] What the algorithmic age requires is the extension of this logic into technologically-mediated domains, where oversight is no longer grounded solely in visible conduct, but also in *what is neglected, omitted, or passively delegated.*

Within the functional taxonomy proposed—procedural neglect, epistemic abdication, and structural complicity—omission operates as the connective tissue. It is the mechanism through which neglect becomes actionable, abdication becomes reviewable, and complicity becomes attributable. In this sense, omission is not a fourth category, but a transversal mode of failure that underpins all three.

(i) In cases of procedural neglect, omission may manifest as the absence of mandated documentation, the failure to trigger internal reporting protocols, or the non-establishment of internal audit trails—each of which may appear bureaucratically minor, but together create a governance vacuum that undermines accountability. It is not the document that is missing, but the systemic infrastructure of responsibility that its absence reveals.

(ii) In cases of epistemic abdication, omission may take the form of board members failing to ask whether a model is explainable, whether its training data reflects unacceptable biases, or whether human override is technically feasible and organisationally embedded. Here, omission is not procedural, but intellectual—a refusal to engage with the structures that mediate decision-making.

[61] See above n. 58. *See also* Pierluigi Matera, *From Red Flags to Black Box: Caremark Duties in the Age of Artificial Intelligence* (Feb. 1, 2026) (unpublished manuscript), https://ssrn.com/abstract=6161886.

(iii) In cases of structural complicity, omission can be deliberate or strategically convenient. It encompasses the conscious choice not to challenge vendor claims, not to commission an external audit, or not to escalate anomalies in post-market performance—often because doing so might impede operational velocity or reveal liabilities that the firm prefers to leave latent. Here, omission merges into tacit endorsement: by failing to act, the board participates in the very governance failure it declines to confront.

Importantly, omission in this context must be interpreted through a positive fiduciary lens: directors are not simply required to refrain from harmful conduct, but to affirmatively engage with complex systems whose risks are known, foreseeable, or institutionally salient. The fiduciary standard is thus no longer limited to a reactive posture of damage control. It is a proactive mandate for critical inquiry, institutional design, and anticipatory risk governance. Silence, in this framework, is not neutrality; it is exposure.

The AI Act reinforces this interpretation through its pervasive reliance on documentation, procedural traceability, and ongoing human oversight. The requirement under Article 14 that human oversight be "effective" and "commensurate" presupposes not merely the presence of human actors, but their active engagement in monitoring, questioning, and, where necessary, countermanding AI decisions. Likewise, Article 29's emphasis on pre-deployment fundamental rights impact assessments presumes a governance structure that does not await harm before intervening, but instead acts in anticipation of risk—requiring boards to establish procedural and ethical safeguards *ex ante*. Where such interventions do not occur, the omission is neither innocent nor invisible; it is institutionally legible as a breach of the duty of care.

This reconceptualisation aligns with broader trends in corporate jurisprudence. In the wake of *Marchand v. Barnhill*, Courts have indicated that omissions—such as the failure to put a "mission-

critical" issue on the board's agenda—may suffice to establish liability.[62] Similarly, under German law, the *Sorgfaltspflicht* and *Aufsichtspflicht* doctrines impose positive organisational duties that, when unfulfilled, generate liability through institutional omission. These frameworks converge around a shared premise: in high-risk, high-complexity domains, failure to act is functionally indistinguishable from failure to govern.

For directors navigating AI governance, the implications are profound. It is no longer defensible to claim ignorance of how an AI system operates, nor is it sufficient to delegate oversight to technical teams without establishing reporting channels, review mechanisms, or escalation protocols. Fiduciary silence may no longer operate as a reliable shield against scrutiny, particularly where oversight omissions concern mission-critical algorithmic systems. The boardroom must evolve from a forum of post-hoc rationalisation into a site of active, pre-emptive vigilance.

Omission, in the final analysis, is not a lapse. It is a structural feature of governance regimes that fail to embed fiduciary duties into the procedural, epistemic, and ethical architectures of algorithmic decision-making. To mitigate this, boards must shift from passive receipt of information to active engagement with systems.

The taxonomy of governance failure outlined herein is therefore not only a typology of error, but a lens for responsibility. It asks boards not simply what they did, but what they allowed to occur through omission, silence, or strategic inaction.

The governance failures described above reverberate beyond adjudication. Procedural neglect, epistemic abdication, and structural complicity increasingly shape how fiduciary conduct is assessed across adjacent institutional domains, including insurance underwriting, disclosure practices, and internal control systems. In these settings, the question is not whether boards complied with the AI

[62] Roy Shapira, *Mission Critical ESG and the Scope of Director Oversight Duties*, 2022 Colum. Bus. L. Rev. 732, 745–47.

Act as such, but whether they can demonstrate substantive oversight over algorithmic decision-making infrastructures.

From a fiduciary perspective, AI due care increasingly operates as a condition of corporate defensibility. Directors and Officers (D&O) insurers, auditors, and market regulators are likely to scrutinise whether boards have established credible oversight mechanisms for AI systems—particularly where such systems mediate mission-critical decisions. Failures of documentation, monitoring, or escalation may be interpreted not merely as compliance gaps, but as evidence that fiduciary judgment was never meaningfully exercised.

These institutional effects are indicative rather than constitutive. They do not drive fiduciary doctrine, but reflect how shifts in the justificatory conditions of judicial deference are already being operationalised and internalised by non-judicial actors.

IV. Expanding the Duty of Loyalty under Algorithmic Governance

Much of the fiduciary debate in the context of AI has focused on the recalibration of the duty of care—emphasising the need for technological literacy, procedural vigilance, and informed oversight.[63] Yet, the duty of loyalty, traditionally framed around the avoidance of conflicts of interest and the preservation of undivided commitment, has received comparatively little scrutiny.[64] This is a mistake.[65] As directors increasingly delegate discretionary authority to AI systems, the very conditions that safeguard loyal judgment—transparency, independence, and integrity—are refracted through data-driven architectures that may embed partialities or advance misaligned interests. The opacity of algorithmic decision-making introduces new risks of conflicted influence: not necessarily because

[63] *See also* Langenbucher, supra note 1.
[64] Gerner-Beuerle et al., *Study on Directors' Duties and Liability*, supra note 38, at 118-162.
[65] Claire Boine, *Fiduciary Principles in AI: Utilizing the Duty of Loyalty to Align Artificial Intelligence Systems with Human Goals* (We Robot Conf., Univ. of Colo. Boulder 2023), https://www.bu.edu/law/files/2023/09/Fiduciary-paper.pdf.

directors act in bad faith, but because the systems through which they act may themselves be structurally compromised. This section argues that loyalty, no less than care, requires reconceptualisation under algorithmic governance—and that this reconceptualisation must encompass both latent conflicts and institutional design.

As directors delegate key discretionary powers to AI systems, the normative foundations of loyalty—transparency, undivided commitment, and integrity of judgment—are displaced by opaque, data-driven processes that may encode partialities unbeknownst even to those deploying them. A recommendation engine for acquisitions, sourced from a third-party provider in which a director holds an undisclosed financial interest, may be trained on data that subtly privileges targets aligned with that interest. Even where the director abstains from overt manipulation, the outcome may reproduce the effects of self-dealing in a technical guise.[66] The principle that directors may not profit from their fiduciary position—regardless of good faith or benefit to the company—has long been affirmed in English law, most notably in *Regal (Hastings) Ltd v Gulliver*, where the House of Lords imposed strict liability on directors who failed to obtain proper authorisation despite acting in the company's interests.[67] The relevance of Regal lies precisely in its formalism. Liability attached irrespective of good faith, benefit to the company, or absence of improper motive. What mattered was not moral blameworthiness, but the objective fact that the fiduciary position had been used in circumstances giving rise to a conflict. Similarly, in *Bhullar v Bhullar*, the Court of Appeal held that fiduciaries breached their duty by acquiring property of the type in which the company might conceivably have an interest, even though the company was not actively pursuing it at the time.[68] Such case law illustrates that disloyalty may lie not in malintent but in the mere fact of acting within a conflicted terrain. Bhullar is particularly instructive for algorithmic governance. The case confirms

---

[66] Luca Enriques & Dirk Andreas Zetzsche, *Corporate Technologies and the Tech Nirvana Fallacy*, 72 Hastings L.J. 55 (2020).
[67] *Regal (Hastings) Ltd. v. Gulliver*, [1967] 2 A.C. 134 (H.L.).
[68] *Bhullar v. Bhullar*, [2003] EWCA (Civ) 424.

that loyalty is breached not only by active self-dealing, but by occupying a decision-making space in which fiduciary judgment is structurally misaligned, even absent contemporaneous competition or intention to exploit. Algorithmic systems that embed optimisation criteria aligned with directors' external interests reproduce precisely this form of structural conflict.

In an AI-mediated environment, that terrain is rendered all the more difficult to detect. Directors may exploit algorithmic systems to serve concealed preferences—for instance, tuning supplier evaluation models to favour entities indirectly linked to their personal holdings, while attributing the result to neutral optimisation. These structural conflicts echo the concern raised in *In re El Paso Corp. Shareholder Litigation*, where a director's divided loyalties in the context of a major transaction gave rise to a finding that the board's process had been undermined by impermissible influence.[69] In such cases, the duty of loyalty is not simply infringed by private enrichment, but by the erosion of independent corporate judgment. The analogy with El Paso is doctrinal rather than rhetorical. Just as the presence of a conflicted human advisor compromised the integrity of the board's process, reliance on algorithmic systems whose design, data sources, or incentives are misaligned with the company's interests undermines the presumption that fiduciary judgment has been exercised.

Accordingly, the traditional emphasis on disclosing personal interests should now expand to include algorithmic dependencies. Directors ought to disclose not only their stake in counterparties, but any material influence over—or benefit from—the AI systems used to support corporate decisions. Moreover, where such systems are acquired from related parties or structured in ways that obscure contested objectives, loyalty obligations should extend to vetting their neutrality and provenance. This reframing finds support in the emerging regulatory landscape: the EU Artificial Intelligence Act imposes duties of transparency, traceability, and human oversight for high-risk systems, thereby articulating a governance model premised on systemic accountability. Loyalty, in this view, is no

[69] *In re El Paso Corp. S'holder Litig.*, C.A. No. 6949-CS (Del. Ch. Feb. 29, 2012).

longer exhausted by the absence of subjective conflict—it must be actively instantiated through institutional safeguards and technical design.

Finally, loyalty may be constructively breached where AI systems, though formally neutral, consistently marginalise stakeholder considerations that directors are statutorily required to weigh. Under section 172 of the Companies Act 2006, directors must have regard to long-term success and broader stakeholder impact. If an AI system optimises exclusively for shareholder return, filtering out environmental, social, or governance data, then reliance on its outputs may amount to a dereliction of fiduciary duty—particularly where such exclusion benefits directors indirectly. In this context, a recalibration of loyalty emerges as necessary: one that frames it as a standard of **AI loyalty oversight**—a positive obligation to interrogate the configuration, independence, and embedded assumptions of the systems through which board decisions are executed. In an age of socio-technical complexity, loyalty can no longer be presumed—it must be designed, disclosed, and defended.

In this light, loyalty demands more than abstention from self-dealing or disclosure of personal interest. It requires directors to actively interrogate the provenance, logic, and institutional alignment of the algorithmic systems through which judgment is exercised. Delegation does not diminish fiduciary responsibility—it transforms it. Where AI systems embed optimisation logics that subtly displace corporate objectives, or reflect external interests with insufficient scrutiny, the board's passive acceptance of such tools may amount to a breach not merely of procedural prudence, but of conflict-sensitive governance. The duty of loyalty, properly understood, must evolve into a standard of AI loyalty oversight: a positive, proactive obligation to ensure that technological proxies do not erode the independence and integrity of corporate judgment. In an algorithmic environment, loyalty is no longer presumed. It must be continuously secured—through disclosure, through institutional design, and through an unwavering commitment to fiduciary impartiality in the age of machine-mediated governance.

While it might be argued that the concerns raised here—particularly those relating to the design, selection, and oversight of algorithmic systems—fall more naturally within the remit of the duty of care, as matters of procedural propriety and sound governance process (on the logic that directors must own not the decision, but the decision-making process), we contend that these issues engage the fiduciary duty of loyalty in its most fundamental form. This is not merely a question of diligence or competence in process execution, but of preserving the moral and institutional integrity (i.e. independence of the decision-making process) of judgment within the corporate sphere. Where directors knowingly or negligently permit the use of AI systems whose design parameters, data inputs, or optimisation criteria embed interests aligned—directly or indirectly—with their own, or fail to disclose material influence over the systems deployed, the breach is not one of carelessness alone. It is a failure of fiduciary fidelity: a breakdown of the expectation that directors act with undivided commitment to the company's best interests, free from conflicted influence. In this light, loyalty entails more than the avoidance of self-dealing—it demands the active assurance that algorithmic proxies reflect, rather than displace, the director's obligation to exercise impartial judgment. The delegation of discretion to technical systems does not dilute this duty; it sharpens it.

The final dimension to be considered, discernible only in delicate filigree, concerns the emergence of a new conflict of interest: the one between the corporation and the algorithmic systems it employs. Still, this is not a matter of overt self-dealing, but rather a deeper issue of systemic misalignment. AI tools—particularly those developed and supplied by third-party vendors—may be configured or trained in ways that prioritise objectives extrinsic to the firm, or that reflect optimisation logics inconsistent with its statutory and fiduciary aims. In such instances, the system ceases to function as a neutral instrument and begins to exert directional influence over corporate judgment, operating with an opacity that renders scrutiny difficult and responsibility diffuse. Especially troubling are adaptive systems that recalibrate based on internal feedback, thereby opening the possibility of subtle

manipulation by insiders or commercial counterparties. These dynamics remain insufficiently explored, yet they raise fundamental questions about the adequacy of existing fiduciary frameworks. What begins as technical delegation may, under certain conditions, give rise to algorithmically embedded conflicts—forms of misalignment in which the duty of loyalty is not deliberately breached, but is nonetheless progressively undermined by the architecture of the delegation itself.

V. The BJR in the Age of AI

This Part does not argue for the reform or erosion of the BJR; rather, it contends that the conditions under which judicial deference has traditionally been justified have already been altered by the structural integration of algorithmic systems into board decision-making.

The BJR has not failed; rather, this Part argues that the justificatory premises traditionally invoked in its defence are increasingly strained under conditions of algorithmic mediation. The integration of AI systems into board-level decision-making does not render the BJR obsolete; rather, it exposes a growing tension between its traditional justificatory premises and the contemporary conditions under which corporate judgment is exercised.[70]

Long regarded as a doctrinal cornerstone of Anglo-American corporate law, the BJR functions as a liability shield: so long as directors act in good faith, with reasonable care, and in the honest belief that their decisions promote the company's best interests,[71] courts are institutionally constrained from substituting their judgment for that of the board. Judicial deference is thereby grounded in the recognition that corporate decision-making is intrinsically uncertain, often undertaken amid incomplete information, and susceptible to failure despite ex ante prudence. The metaphor

[70] Geneviève Helleringer & Florian Möslein, *AI & the Business Judgment Rule: Heightened Information Duty*, U. Chi. L. Rev. Online (2024), https://lawreview.uchicago.edu/online-archive/ai-business-judgment-rule-heightened-information-duty.

[71] See the leading case on the matter, *Aronson v. Lewis*, 473 A.2d 805, 812 (Del. 1984).

frequently invoked—that of a pilot navigating uncertain skies—aptly captures the normative ideal: directors should not be held to account for turbulence they cannot foresee.[72]

Yet this logic presupposes a central premise: that directors, even when wrong, are the primary architects of the decisions for which they are held accountable. This presumption begins to falter in the context of opaque algorithmic processes. Historically, judicial deference has been justified by reference to directors' epistemic position—their access to information, their capacity for contextual judgment, and their institutional role as the locus of discretion—but where judgment is increasingly mediated or partially displaced by opaque algorithmic systems, these assumptions no longer operate in the same way. Deference persists, but its underlying rationale is no longer anchored in the same conception of human judgment. Contemporary boards routinely employ AI systems for strategic planning, risk modelling, and operational optimisation. These systems are not merely analytical tools but often operate with a degree of semi-autonomy, producing outputs that may elude the intuitive grasp of their human overseers. In such contexts, directors may no longer occupy the same position traditionally presupposed by fiduciary deference, insofar as key decisional inputs are generated by systems whose internal logics are only partially accessible to human oversight.

The resulting strain on the BJR is conceptual as well as functional, insofar as the doctrine's traditional reliance on directors' informational superiority becomes difficult to sustain where judgment is materially mediated by opaque algorithmic systems. This tension is sharpened by the increasing proceduralisation of directors' duties evident in recent case law. Decisions such as *In re Citigroup Inc. Shareholder Derivative Litigation*[73] and *In re Rural/Metro Corp. Stockholders Litigation*[74] evince a judicial

[72] See also David Kershaw, *Business Judgment and the Idea of Honesty in the Exercise of Delegated Power*, in *The Foundations of Anglo-American Corporate Fiduciary Law* 21 (Andrew S. Gold & Paul B. Miller eds., Cambridge Univ. Press 2018); Bernard S. Sharfman, *The Importance of the Business Judgment Rule*, Faculty Scholarship No. 1586 (2017), https://digitalcommons.law.umaryland.edu/fac_pubs/1586; Stephen M. Bainbridge, *The Business Judgment Rule as Abstention Doctrine*, 57 Vand. L. Rev. 83 (2004); Lynn A. Stout, *On the Proper Motives of Corporate Directors (or, Why You Don't Want to Invite Homo Economicus to Join Your Board)*, 28 Del. J. Corp. L. 1 (2003).

[73] *In re Citigroup Inc. Shareholder Derivative Litigation*, 964 A.2d 106 (Del. Ch. 2009).

[74] *In re Rural/Metro Corp. Stockholders Litigation*, 88 A.3d 54 (Del. Ch. 2014).

willingness to examine not merely the substance of board decisions but the quality of the underlying governance processes. Courts now expect boards to establish and maintain verifiable systems of compliance and risk oversight; mere formalism does not suffice. Algorithmic systems further complicate these expectations. Their complexity and opacity introduce new forms of epistemic risk—the ones that directors may neither perceive nor control without meaningful scrutiny of the assumptions, data sources, and objectives embedded in the technology itself.

In this context, the analogy to conflicted human advisors is instructive. In *In re Del Monte Foods Company Shareholders Litigation*, the court faulted directors for relying on conflicted financial advisors without appropriate interrogation. It is plausible, indeed increasingly likely, that future courts will demand a similar degree of diligence in relation to AI systems, especially where such tools are developed by third parties with commercial interests that may diverge from those of the corporation.[75]

Moreover, as AI becomes embedded in industry standards for risk analysis and compliance, failure to implement such technologies may itself constitute a breach of fiduciary duty. This represents a normative inflection point. No longer is the question simply whether directors have used AI prudently; rather, the inquiry shifts to whether their inaction reflects a failure to meet the threshold of care expected in a technologically sophisticated governance environment.

The growing proceduralisation of board oversight demands more than just internal vigilance; it now calls for external normative anchoring. In this light, the AI Act assumes a key role not because of its territorial authority, but because of the type of governance problem it addresses. The AI Act does not operate as a source of directly applicable corporate obligations in the US or the UK. Rather, it crystallises a procedural response to a structural dilemma that fiduciary law in all advanced corporate systems is already confronting: how to justify deference when judgment is exercised through opaque,

[75] *In re Del Monte Foods Co. Shareholders Litigation*, 25 A.3d 813 (Del. Ch. 2011)

system-mediated processes: it furnishes the statutory scaffolding within the EU, and an analytically instructive reference point elsewhere, upon which the evolution of the BJR may be constructively reimagined. Articles 6 (risk classification), 14 (human oversight), and 29 (impact assessments) articulate regulatory expectations that increasingly bear upon board-level governance. Directors must therefore ensure that AI deployment is neither casual nor unmoored from compliance architecture. Engaging independent expertise, establishing internal protocols for oversight, and maintaining audit trails are no longer best practices—they increasingly function as conditions of fiduciary deference. Indeed, failure to implement such measures may not merely offend sound governance; it may increasingly be treated as inconsistent with the standard of fiduciary oversight to which courts are prepared to defer.

Such an approach does not bifurcate corporate decision-making into 'AI' and 'non-AI' silos. Rather, it demands that directors exercise the same agency with respect to algorithmic tools as they do with any other strategic input: to interrogate assumptions, weigh consequences, and remain accountable for outcomes. The analogical resonance with *In re Rural/Metro*[76] and *In re El Paso*[77] is instructive. Just as blind reliance on conflicted human advisors undermines the BJR, so too might uncritical deference to AI systems vitiate the presumption of reasoned judgment.[78]

Judicial doctrine may evolve accordingly. Courts, while recognising that directors are not expected to be engineers, may increasingly treat reliance on AI as compatible with deference only where it is deliberate, informed, and procedurally substantiated. The core tenet of the BJR—respect for managerial discretion—becomes difficult to sustain where decisions are effectively abdicated to systems whose operation directors neither understand nor supervise.[79] Here, the fiduciary obligation is one of stewardship, not surrender.

---

[76] *In re Rural Metro Corp. S'holders Litig.*, 88 A.3d 54 (Del. Ch. 2014).

[77] *In re El Paso Corp. S'holder Litig.*, 41 A.3d 432 (Del. Ch. 2012).

[78] Bainbridge & Henderson, *Boards-R-Us*, supra note 8; Bainbridge & Henderson, *Outsourcing the Board*, supra note 8; Gilson & Gordon, *Board 3.0*, supra note 8.

[79] Posner & Saran, *Judge AI*, supra note 9.

Moreover, the liability landscape is shifting. Articles 64–70 of the AI Act endow regulatory authorities with investigative and sanctioning powers, expanding the contours of both institutional and individual accountability. The AI Act's extraterritorial reach further intensifies the compliance burdens for transnational corporations, rendering harmonised governance frameworks essential.[80] In such circumstances, failures of documentation or transparency may acquire doctrinal salience, informing both regulatory assessments and private law claims concerning the adequacy of fiduciary oversight.

Ultimately, the notion of AI due care may be understood as an inflection in fiduciary reasoning, marking the point at which traditional accountability logics confront the conditions of technological governance. Directors who engage seriously with this duty preserve the presumption of judgment that the BJR affords. Those who do not, risk seeing that presumption unravel.

In light of this, the legitimacy of the BJR increasingly depends on assumptions that no longer hold unchanged once algorithmic mediation becomes structurally embedded in board decision-making. This will entail refining the criteria for "due care" and "loyalty oversight" to encompass both technological literacy and procedural vigilance. Absent such adaptation, the BJR risks offering protection in circumstances where directors no longer fulfil the role for which deference was historically justified.

Crucially, such demonstrability is not exhausted by ex post rationalisations of board decisions but inheres in ex ante procedural architecture. Judicial deference under the BJR becomes defensible only where boards can point to institutionalised forms of epistemic engagement, which may include, for example, traceable deliberation records, escalation protocols for anomalous outputs, or structured board-level review of AI governance assessments. The locus of "reasoned judgment" once again

[80] David Krause, The EU AI Act and the Future of AI Governance: Implications for U.S. Firms and Policymakers (Mar. 17, 2025) (unpublished manuscript), https://ssrn.com/abstract=5181797.

migrates from the *content* of board decisions to the integrity of the oversight process—this time, the latter just involves one more element.

Finally, the integration of algorithmic reasoning into board deliberations repositions the BJR within a broader systemic economy of trust. Where directors fail to document dissent from algorithmic recommendations, or to embed governance structures capable of resisting automation bias, they risk transforming the exercise of discretion into a form of de facto reliance on algorithmic outputs, thereby weakening the normative basis for judicial abstention. If the BJR is to remain a bastion of fiduciary legitimacy in the algorithmic age, it may need to be understood not only as a shield for managerial experimentation, but also as a framework within which demonstrable stewardship over digital systems becomes relevant. Only then can the BJR preserve its justificatory integrity in an era where the locus of decision-making is increasingly synthetic.

Seen in this light, the BJR continues to operate, but under altered epistemic conditions. Judicial deference, in this setting, becomes less a function of abstract trust in human judgment and more a function of whether directors can demonstrate meaningful stewardship over the algorithmic systems through which judgment is exercised.

VI. Conclusion: Fiduciary Deference under Algorithmic Conditions

This Article has argued that the growing integration of algorithmic systems into corporate decision-making cannot be adequately understood as raising questions of compliance, technical competence, or risk management alone. At stake is something more fundamental: the justificatory foundations of fiduciary deference itself. The BJR has long rested on an implicit premise—that corporate decisions, however fallible, remain attributable to human judgment exercised under conditions of uncertainty. Algorithmic mediation unsettles that premise by altering not only how decisions are informed, but how judgment is constituted, externalised, and, in some cases, partially displaced.

The European Union Artificial Intelligence Act brings this shift into sharp relief. Although the Act does not regulate corporate boards, nor does it purport to redefine directors' duties, it proceduralises oversight, formalises human supervision, and institutionalises documentation as a proxy for accountability. In doing so, it alters the background conditions under which fiduciary discretion claims legitimacy. What emerges is not a new category of liability, but a transformation in the assumptions that have historically justified judicial abstention. Where corporate judgment is increasingly mediated through opaque systems whose logic directors neither fully grasp nor meaningfully interrogate, the traditional grounds for deference become more fragile.

This insight matters well beyond the European regulatory context. Courts in the United States and other common-law jurisdictions routinely confront claims that turn on the legitimacy of board reliance—on experts, advisors, models, and increasingly, algorithmic systems. Yet fiduciary doctrine has largely treated such reliance as a neutral fact, rather than as a variable that reshapes the posture of decision-making itself. This Article suggests that algorithmic reliance is different in kind. It does not merely supplement judgment; it can reconfigure the locus of judgment in ways that strain the justificatory logic of doctrines premised on human discretion.

The argument advanced here is not that fiduciary law should become more demanding, nor that courts should abandon the BJR in the face of technological complexity. To the contrary, this Article defends fiduciary deference—but only on recalibrated terms. The duties of care and loyalty do not expand quantitatively; they change qualitatively. Care becomes intelligible only by reference to engagement with the systems through which decisions are made. Loyalty can no longer be reduced to the avoidance of personal conflict alone, but implicates stewardship over algorithmic infrastructures that may embed misaligned incentives, opaque optimisation logics, or normative distortions. Fiduciary responsibility, in this setting, turns less on the correctness of outcomes than on the integrity of the processes through which discretion is exercised.

For courts, this reframing clarifies what is—and is not—at stake in cases involving algorithmic governance. Judicial scrutiny need not focus on the technical merits of AI systems, nor substitute judicial judgment for managerial choice. The relevant inquiry is instead whether reliance on algorithmic tools remains compatible with the justificatory premises of deference. Where boards can demonstrate that algorithmic systems mediate, rather than substitute for, judgment—where oversight is substantive rather than ritualised, and responsibility remains institutionally traceable—judicial abstention remains normatively defensible. Where they cannot, deference risks insulating not discretion, but abdication.

For boards, the analysis underscores that algorithmic governance is not merely an operational or compliance issue, but a fiduciary one. The legitimacy of discretion increasingly depends on demonstrable stewardship over the conditions of decision-making itself—over how systems are selected, supervised, interrogated, and constrained. For regulators, the Article highlights the limits of proceduralisation as a substitute for judgment, and the importance of fiduciary law as a residual site of normative accountability in algorithmically mediated governance environments.

As algorithmic systems become embedded in boardroom decision architectures, questions that once appeared peripheral—about reliance, oversight, and delegation—are increasingly likely to surface at the centre of fiduciary litigation, regulatory enforcement, and judicial reasoning about the scope and limits of deference.

Once fiduciary deference is decoupled from its traditional assumptions about human judgment, its boundaries must be reconsidered. Not expanded, but re-anchored. In an algorithmic environment, the BJR endures not as a blanket shield for technologically mediated decisions, but as a doctrine whose legitimacy depends on the continued presence of judgment worthy of deference. The rule survives—but only insofar as it continues to protect judgment, rather than its systematic displacement.

The rule survives—but only insofar as it continues to protect judgment, rather than its systematic displacement.