



2026

## **The Algorithmic Fiduciary: Governing The Corporate & The Systemic Risk of Obscurities in AI-Driven ESG**

**Ansh Parashar**

**Mahee Mishra**

Follow this and additional works at: [www.beyondbriefslawreview.in](http://www.beyondbriefslawreview.in)

**Published by Paperleaf International Media and Publications under the aegis of Dexon Global LLP**

---

### Recommended Citation

Parashar Ansh, Mishra Mahee (2026) "The Algorithmic Fiduciary: Governing the Corporate & The Systemic Risk of Obscurities in AI-Driven ESG," *Beyond Briefs Law Review: Vol. 03, Issue 02, Article 2.*

This article has been brought to you for free and open access by Beyond Briefs Law Review. It has been accepted for inclusion in the Beyond Briefs Law Review by the Editorial Office of Beyond Briefs Law Review.



# **THE ALGORITHMIC FIDUCIARY: GOVERNING THE CORPORATE & THE SYSTEMIC RISK OF OBSCURITIES IN AI-DRIVEN ESG**

**Ansh Parashar<sup>1</sup>**

**Mahee Mishra<sup>2</sup>**

## **ABSTRACT**

This article critically examines the integration of Artificial Intelligence (AI) into Environmental, Social, and Governance (ESG) investing, a development promising a new era of objectivity but simultaneously introducing profound risks. While AI is positioned to elevate ESG metrics from voluntary disclosures to key determinants of market value, the prevalent use of opaque black box models creates significant governance and legal challenges that existing corporate law frameworks are ill-equipped to address. Employing a cross-disciplinary legal and doctrinal methodology, the paper integrates foundational corporate law doctrines, such as the fiduciary duty of care and the Business Judgment Rule, with securities regulation and antitrust theory, substantiating its analysis with evolving case law and pioneering regulatory paradigms like the EU's AI Act. The central argument is that the uncritical delegation of ESG oversight to unaudited algorithms subverts core tenets of corporate governance, forges new frontiers of liability for securities fraud, and introduces systemic market risks like tacit algorithmic collusion. The analysis reveals three primary conclusions: first, algorithmic opacity creates a 'responsibility gap,' placing corporate boards in direct conflict with their fiduciary duty to remain 'reasonably informed'; second, flawed AI-generated ESG scores can constitute materially false statements, linking internal governance failures to external securities liability; and third, the concentration of AI-ESG systems can amplify investment herding and facilitate anti-competitive coordination.

Ultimately, this paper challenges the prevailing narrative of 'algorithmic authenticity' by proposing a comprehensive framework of Auditable Algorithmic Governance (AAG) centred on mandated explainability, formal data diligence, and independent third-party auditing. This perspective reframes AI not as a replacement for human judgment but as a tool demanding robust human oversight, offering a forward-looking agenda for corporate boards and regulators to ensure technological deployment reinforces market integrity and authentic accountability.

---

<sup>1</sup> UG Student, Hidayatullah National Law University. You may contact the author at this email address: [ansh.232892@hnlu.ac.in](mailto:ansh.232892@hnlu.ac.in)

<sup>2</sup> UG Student, Hidayatullah National Law University. You may contact the author at this email address: [mahee.232921@hnlu.ac.in](mailto:mahee.232921@hnlu.ac.in)

**Keywords:** ESG-induced Corporate Liability, Black Box Issue, Fiduciary Duty, Explainable AI (XAI), Auditable Algorithmic Governance.

## Introduction

The ESG score changed from a nice-to-have voluntary disclosure metric to a central measure of market value, investor trust, and regulatory oversight because of the unparalleled rise of Social, Environmental, and Governance (ESG) Investing. Artificial intelligence (AI) has rapidly gained popularity as a tool to gather, examine and disseminate ESG data because of the promises of speed, precision and revelations. A so-called “single source of truth” is created through advanced methods like natural language processing, satellite imagery analysis and anomaly detection, reshaping ESG evaluation from a bias to a science. Proprietary AI tools are being used more by corporate boards and investors, especially with legislation like the EU’s Corporate Sustainability Reporting Directive (CSRD) and AI Act that call for detailed ESG reporting.<sup>3</sup> This, however, comes with governance, legal and systemic risks that a corporate law is not designed to handle. The interpretability gap in deep learning models, often called the “black box” issue, poses serious challenges in understanding a model’s logic, especially for directors basing strategic decisions on it. This kind of dependency on technology clashes with the duties of care, which expects directors to be “reasonably informed.” The Business Judgment Rule (BJR) offers protection, which courts have long extended to directors by accepting that they acted with an informed and rational process.<sup>4</sup> Director protections diminish when ESG responsibilities are delegated to unchecked algorithms, increasing the risk of liability for oversight failures. The failure in ESG communication is exemplified by the Brumadinho dam disaster in Brazil. The SEC lawsuit against Vale S.A. exemplifies how poorly managed disclosures can lead to enormous regulatory fines and market devaluation.<sup>5</sup> Hence, ESG disclosures must be handled with the same seriousness as any other legal obligation, they cannot be taken lightly. This article contends that the analysis of Environmental, Social, and Governance (ESG) metrics based on Artificial Intelligence (AI) analytics weakens the governance of a specific company, as well as the system it operates in. The use of AI creates an accountability gap in decision-making for technology users. On the other hand, the reliance on a handful of AI-based scoring institutions creates the imbalance of system-wide investing and implicit algorithmic collusion.<sup>6</sup> In an ESG context, AI serves not as an unbiased judge but more as a gatekeeper that misallocates capital and damages corporate reputation. To

---

<sup>3</sup> A.K. Gandhi, *Redefining ESG Compliance with Machine Learning and Predictive Analytics*, 6 *Int’l J. AI, BigData, Computational & Mgmt. Stud.* 66 (2025), <https://doi.org/10.63282/3050-9416.IJAIBDCMS-V6I2P108>.

<sup>4</sup> Furkan Koç, *AI in ESG Compliance: Legal Challenges Under CSRD and AI Act (July 2025)* (M.Sc. thesis, Tilburg Univ.), <https://doi.org/10.13140/RG.2.2.32043.76326>.

<sup>5</sup> *In re Caremark Int’l, Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996).

<sup>6</sup> *Sec. & Exch. Comm’n v. Vale S.A.*, No. 1:22-cv-02405 (E.D.N.Y. filed Mar. 28, 2022).

mitigate these dangers, the article advocates for a policy framework, Auditable Algorithmic Governance (AAG), which requires explainability, data due diligence, and external audits.<sup>7</sup> The case is presented as follows: Part I covers the ESG AI verification; Part II reviews governance problems through a fiduciary lens; Part III investigates securities regulation and disclosure liability; Part IV considers systemic risk; and Part V details the AAG framework. In the end, this article draws attention to AI as an instrument that requires human supervision, redirecting sustainability governance from algorithmic veracity to genuine accountability.

## **The Promise and Peril of AI In ESG Verification**

### **A. The Promise: AI as a Verification Engine**

AI promises new capabilities to assess ESG factors by verifying corporate disclosures and generating seemingly objective, data-driven insights for investors. Additionally, AI-driven monitoring tools, including drone footage, satellite imagery, and even IoT sensors, are able to identify environmental breaches in almost real time, eliminating the need for corporate self-reporting and bringing to light risks that were previously hidden. As a case in point, AI-powered image recognition can assist in monitoring deforestation and even illegal mining, which can help both investors and regulators to confirm sustainability assertions without having to rely on companies' disclosures.<sup>8</sup> In addition, computer vision models have been employed in the analysis of drone footage of manufacturing plants to quantify emissions, and machine learning is used to determine the legitimacy of emissions data in the context of climate agreements. Natural language processing (NLP) adds an important new dimension by analysing vast quantities of unstructured data. Through scrutinizing countless documents such as regulatory filings, NGO reports, academic papers, and media articles, NLP models can identify inconsistencies between a company's own communications and external, independent accounts.<sup>9</sup> This ability to check information on a large-scale usher a revision of ESG and compliance from being static, one-off annual checks to being perpetually evolving checks. Predictive analytics improves this function by estimating ESG risks such as labour conflicts, supply chain issues, or physical climate threats using real-time data and past data. The use of predictive analytics and NLP models fosters the reduction of human errors, decreases the information gap, and generates an aura of openness, which has led to the rapid use

---

<sup>7</sup> Furkan Koç, *AI in ESG Compliance: Legal Challenges Under CSRD and AI Act (July 2025)* (M.Sc. thesis, Tilburg Univ.), <https://doi.org/10.13140/RG.2.2.32043.76326>.

<sup>8</sup> N.L. Rane, S.P. Choudhary & J. Rane, *Artificial Intelligence–Driven Approaches to Strengthening ESG Criteria in Sustainable Business Practices: A Review* (n.d.).

<sup>9</sup> *The Impact of Artificial Intelligence on ESG: A Conceptual Framework for Practitioners and Policymakers*, J. Mgmt. for Glob. Sustainability (May 26, 2025).

of AI-driven ESG tools by investors, regulators, and asset managers. The technology that was promised to us has changed how the market views ESG reporting. The flow of funds into ESG-themed investments has set new demands from investors for more detailed, independently verifiable, and faster sustainability data. The ability of AI to bring in and process data is in line with these expectations and creates a sense of objectivity that earlier, ESG evaluations based on human judgment did not have. Far from complementing ESG reporting, AI starts to transform governance in its wake as board members and institutional investors adopt analytical results to strategic decisions and treat them as unquestionable facts.

### **B. The Peril: Black Box Models and Algorithmic Opacity**

The understanding of AI governance tools is hindered by complexity that marks both of its advantages and challenges. Deep learning systems more importantly advanced AI models function as “black boxes” where the decision-making is too complex to be understood even by the designers. This is particularly a problem for AI governance and risk management tools, where company executives and board members make critical decisions based on the ratings given, yet know nothing of the intricate details, underlying assumptions, and data of the systems. The issue is further compounded by ESG metrics, where AI governance models’ and risk management ratings’ methodologies remain undisclosed by citing intellectual property.<sup>10</sup> Opacity is a legal issue, not just a technical one. Duty of care statutes, for example, fiduciary duty statutes, require that before directors make any major decision, they must be “reasonably informed.” Inscrutable tools are a form of opacity and relying on them violates the standard since the tools are neither outcome nor process-oriented. The EU AI Act incorporates regulatory frameworks to explicitly address the risks associated with opacity. They enact responsibilities for explainability, risk controls, and human oversight for AI systems that are regarded as high-risk.<sup>11</sup> Furthermore, the risk of black-box ESG scoring systems is even more accentuated under the General Data Protection Regulation. Organizations using automated decision systems must be able to prove accountability and transparency under the regulation. This highlights legal tension. Risk matters like this emphasize the contradiction: AI offers the chance of insight free of bias, but its lack of clarity makes it difficult to trust, leaves the board open to litigation, and poses problems with meeting regulatory requirements.

---

<sup>10</sup> Mikael Minkkinen, Akseli Niukkanen & Matti Mäntymäki, What About Investors? ESG Analyses as Tools for Ethics-Based AI Auditing, 39 *AI & Soc’y* 329 (2024), <https://doi.org/10.1007/s00146-022-01415-0>.

<sup>11</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Artificial Intelligence (AI Act), arts. 6–15.

### **C. Gaming the Algorithm and Market Incentives**

Algorithm gaming has also been triggered by AI-ESG scoring models and is the newest form of regulatory arbitrage. Organisations knowledgeable of the key scoring metrics are able to adopt shallow measures to improve their ESG scores. Shallow measures such as publicising workplace diversity initiatives while ignoring deeper structural inequalities, or investing in carbon offsets without cutting down on carbon emissions are a few examples. These are signs of a wider problem: the proprietary scoring algorithms of ESG ratings challenge companies to meet the metrics and improve their image, rather than improve compliance and make impactful changes. Furthering this problem is the commercial interest of ESG rating companies. There is a clear conflict of interest as they offer advisory services to companies to improve their scores. The ESG metrics cannot be scrutinised by either regulators or investors due to the scoring models being kept under wraps and labelled as trade secrets. This leads to a scenario in which companies are more inclined to ‘feed the algorithm’ instead of working on improving real ESG metrics. Such strategic manipulation weakens the credibility of ESG reporting and explains why explainability and auditing mechanisms are essential. Even more critically, it sets the stage for the governance challenges explored in Part II in which the dependence on opaque, and possibly gamed, systems create a direct conflict with fiduciary duty and corporate oversight integrity.

### **D. Case Study: Algorithmic Governance Failure of Boohoo Group PLC**

The scandal surrounding The Boohoo Group PLC in 2020 shows the risks of relying on ESG algorithms to evaluate firms.<sup>12</sup> A series of investigative reports revealed that garment workers in Boohoo’s Leicester supply chain were paid a substandard wage of £3.50 per hour and were marred with squalid and abusive working conditions.<sup>13</sup> As a result of these reports, there was a government inquiry on modern slavery issues and an independent review led by Alison Levitt QC. The review, which was concluded months later in 2021, agreed with the allegations and deemed them “substantially true” and Boohoo’s modern slavery oversight was “inadequate.” The review also criticized Boohoo’s corporate governance, calling it “systemic corporate governance weaknesses”.<sup>14</sup> Scoffing at ethics and takfiri tactics, Boohoo managed to enjoy an enviable rating, called ESG, before the scandal. Boohoo was granted an AA ranking by MSCI, a notable ESG

---

<sup>12</sup> The Guardian, *Boohoo Accused of Failing to Improve Working Conditions in Its Supply Chain* (June 18, 2021); The Guardian, *The Billionaire Boohoo Family Who Started with a Market Stall in Manchester* (July 10, 2020).

<sup>13</sup> Corporate Justice Coalition, *Leicester’s Sweatshops: Abuses in Boohoo’s Value Chains* (Nov. 2023).

<sup>14</sup> *Ibid.*

ranking organization, which allowed many ESG investors to hold on to their shares. This evaluation rested a lot on the public-facing information that Boohoo provided and overlooked localised accounts of the mistreatment of workers. Not unlike other unethical “fast fashion” companies, Boohoo refused to make public its list of suppliers. This omission led to the ESG model presuming low-risk supply chains, which is a great example of how algorithm-based analyses that rely on partial information can bolster a deceptive image of sustainable business practices. Even after the scandal emerged, Boohoo’s rating was only slightly lowered, which is yet another proof of the gap between ground-level reality and machine-based scoring. The scandal revealed a serious failure in governance. Boohoo’s board was first informed about the supply chain issues in 2019 but did not take significant action.<sup>15</sup> The company’s investors and proxy advisors later criticized the directors for dragging their feet and relying on high ESG scores and internal reports, labelling the conduct as a breach of fiduciary duty. Although Boohoo started its ‘Agenda for Change’ initiative to improve supply chain issues, the efforts were described by labour advocacy groups as window dressing, further deepening the view that Boohoo was more interested in reputation risk than real change.<sup>16</sup> The Boohoo case study demonstrates the real-world consequences of AI-ESG algorithms, particularly in terms of risk management and board accountability. Directors who approve of algorithmic scoring without adequate diligence may be deemed reckless under fiduciary and statutory obligations, particularly in jurisdictions such as the UK, where the corporate governance code requires active risk management and open disclosure. This case shows why human oversight of AI algorithms, auditing of AI algorithms, and verified supply chain transparency must be incorporated into the ESG evaluation models, keeping in line with the rising expectations of the regulators for the accuracy of ESG information.

## **The Governance Crisis: Fiduciary Duty In The Age of the Algorithm**

### **A. Framing the Crisis: Board Oversight Meets Algorithmic Opacity**

Artificial intelligence (AI) has come to automate the collection and verification of data and, more importantly, bolster decision-making across various sectors, including Environmental, Social, and Governance (ESG) scoring.<sup>17</sup> Still, this dependence leads to significant governance issues: AI systems, especially those utilizing machine learning and neural networks, are “black boxes” that

---

<sup>15</sup> The Guardian, *Boohoo Accused of Failing to Improve Working Conditions in Its Supply Chain* (June 18, 2021); The Guardian, *The Billionaire Boohoo Family Who Started with a Market Stall in Manchester* (July 10, 2020).

<sup>16</sup> Corporate Justice Coalition, *Leicester’s Sweatshops: Abuses in Boohoo’s Value Chains* (Nov. 2023).

<sup>17</sup> Vedrana Božić, *Environmental, Social, and Governance Approach (ESG) & Artificial Intelligence (AI)* (May 2023), <https://www.researchgate.net/publication/370844694>.

hide their internal logic even from those who create them.<sup>18</sup> ESG scoring has long been under fire for inconsistent methodologies. Now directors' inability, due to algorithmic opacity, to check data and verify material risks further complicates the issue.<sup>19</sup> This issue conflicts with core principles in corporate law and the fiduciary responsibilities of the board, examining risk and making decisions in good faith, flowing directly from foundational corporate laws: the duty of care, the duty of loyalty, and the duty to make a rational business decision.<sup>20</sup> In this scenario, ESG scoring platforms function as critical 'mission systems' that influence investor reporting, shape investor views, and guide regulatory adherence, rather than as simple analytical tools.<sup>21</sup> Delaware case law stresses that directors need to be sufficiently knowledgeable of such systems so as to meet fiduciary responsibilities.<sup>22</sup> If directors are unable to interrogate or supervise the outputs of algorithms, then relying on these algorithms endangers accountability by producing what is termed as a "responsibility gap".<sup>23</sup> This continuation provides us with a context to analyze how fiduciary principles, which address financial and structural challenges, take on the issues posed by AI systems that lack transparency.<sup>24</sup>

## **B. The Business Judgment Rule & Duty of Care: Baseline Doctrines**

The Business Judgment Rule (BJR) serves as the gateway of protections available to directors; it protects them from liability provided that they act in good faith, with due care, and in the best interest of the corporation.<sup>25</sup> Further, as set out by Delaware law, directors' decisions are presumed to be reasonable in relation to the information available at the time of the decision.<sup>26</sup> Nonetheless, this protection is limited: an absence to monitor or an absence to implement oversight systems constitutes a breach of the BJR's protection.<sup>27</sup> The Delaware Chancery Court, in *In re Caremark International Inc. Derivative Litigation* (1996), signalled the need for directors to make sure the firm has a proper system in place for information and reporting that would notify of significant

---

<sup>18</sup> Furkan Koç, *AI in ESG Compliance: Legal Challenges Under CSRD and AI Act* (July 2025) (M.Sc. thesis, Tilburg Univ.), <https://doi.org/10.13140/RG.2.2.32043.76326>.

<sup>19</sup> Mikael Minkkinen, Akseli Niukkanen & Matti Mäntymäki, *What About Investors? ESG Analyses as Tools for Ethics-Based AI Auditing*, 39 *AI & Soc'y* 329 (2024), <https://doi.org/10.1007/s00146-022-01415-0>.

<sup>20</sup> William M. Lafferty, Lisa A. Schmidt & Donald J. Wolfe, Jr., *A Brief Introduction to the Fiduciary Duties of Directors Under Delaware Law*, 116 *Penn St. L. Rev.* 837 (2012).

<sup>21</sup> *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019).

<sup>22</sup> *In re Caremark Int'l, Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996).

<sup>23</sup> Paolo Balboni & Kimberly E. Francis, *Data Ethics and Digital Sustainability: Bridging Legal Data Protection Compliance and ESG for a Responsible Data-Driven Future*, 22 *J. Responsible Tech.* 100099 (2025), <https://doi.org/10.1016/j.jrt.2024.100099>.

<sup>24</sup> Seung U. Lee et al., *Integrating ESG and AI: A Comprehensive Responsible AI Assessment Framework*, *AI & Ethics* (2025).

<sup>25</sup> *Aronson v. Lewis*, 473 A.2d 805 (Del. 1984).

<sup>26</sup> *Smith v. Van Gorkom*, 488 A.2d 858 (Del. 1985).

<sup>27</sup> *Stone v. Ritter*, 911 A.2d 362 (Del. 2006) (en banc).

risks.<sup>28</sup> While Caremark defined a strict standard for liability, *Stone v. Ritter* (2006) named oversight failures breaches of the duty of loyalty when the failures are “conscious disregard” to the tasks.<sup>29</sup> This idea was applied in *Marchand v. Barnhill* (2019), where the court deemed directors responsible for not monitoring the food safety of a business whose core “monoline” risk necessitated such oversight.<sup>30</sup> Likewise, the Boeing Co. derivative litigation (2021) revealed the willingness of the courts to investigate directors that overlook warning signs relating to technical and operating risks.<sup>31</sup> Together, these cases clarify that oversight requires an active engagement of the board of directors.<sup>32</sup> The board must actively inquire into systems that impact the company’s compliance, safety, or reputation.<sup>33</sup> ESG metrics, having regulatory implications under frameworks such as the CSRD, must be considered a mission-critical area that requires due diligence.<sup>34</sup> In other words, the BJR cannot be invoked in defence of directors who accept AI-driven ESG metrics at face value without setting strict oversight mechanisms.<sup>35</sup>

### **C. From Oversight to Liability: ESG-AI As a Caremark Analogue**

Caremark, Marchand, and Boeing cases help us understand that ESG AI systems are not simply neutral tools but increasingly pivotal compliance infrastructure, particularly as ESG metrics grow significant to investors and regulatory reports. AI-generated ESG scores are the basis for marketing sustainability credentials, shaping shareholder perception and valuation, as well as sponsoring the direction for corporate disclosures.<sup>36</sup> As with safety and data-security systems, neglecting the oversight of such models’ imperfections and biases invites the same kind of liability.<sup>37</sup> Imagine a case in which a company’s ESG scores get inflated because of biased data, resulting in a regulatory investigation or damage to its image. If the AI’s method goes unchallenged by the board, with mission-critical ESG disclosures, such failure could be deemed a “conscious disregard” of oversight duties per Marchand. Likewise, the Boeing case indicates that courts are ready to charge directors with liability if they “lent oversight authority” to management and closed their eyes to system-level risks after red flags emerged. Similarly, not requesting transparency, a third-party audit, or periodic performance reviews for an opaque AI system that carries out ESG

---

<sup>28</sup> *Ibid.*

<sup>29</sup> *In re Walt Disney Co. Derivative Litig.*, 906 A.2d 27 (Del. 2006).

<sup>30</sup> *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019).

<sup>31</sup> *In re Boeing Co. Derivative Litig.*, 2021 WL 4059934 (Del. Ch. Sept. 7, 2021).

<sup>32</sup> *Ibid.*

<sup>33</sup> *Ibid.*

<sup>34</sup> Furkan Koç, AI in ESG Compliance: Legal Challenges Under CSRD and AI Act (July 2025) (M.Sc. thesis, Tilburg Univ.), <https://doi.org/10.13140/RG.2.2.32043.76326>.

<sup>35</sup> *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019).

<sup>36</sup> *In re Caremark Int’l, Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996)

<sup>37</sup> *Ibid.*

risk assessments is just as risky as the above. A similar precedent can be seen in cybersecurity litigation: boards have been charged with Caremark violations for neglecting to oversee technical risk environments. Even in instances where cases were dismissed, courts still highlighted the importance of directors overseeing technological threats and how they evolve. The ESG-AI issue presents equivalent risk, further exacerbated by regulations like the EU AI Act and CSRD. These regulations institutionalize the notion that AI systems, particularly those classified as high risk, have to be human-supervised and explainable. The intersection of regulations and case law increases in gravity and signifies an emerging norm: boards are expected to analyse the AI models used to generate ESG scores, perform risk assessments, and keep documented proof of deliberate engagement or face fiduciary accountability.

#### **D. The AI Judgment Rule and Explainability as a Fiduciary Imperative**

Recently, scholars have begun to anticipate that the notion of the fiduciary duties will expand to include technological literacy and algorithmic oversight of their operations. The duties of the board of directors and other executive officers have been interpreted and computed to evolve over time with the expanding industries.<sup>38</sup> In the case of ESG-AI, the AI Judgment Rule properly focuses fiduciary duties on directors' understanding, or responsible delegation to experts, of the black-boxed algorithmic systems before the outputs are used.<sup>39</sup> Explainable AI (XAI) techniques provide the board a practical method to meet these expectations.<sup>40</sup> They cannot in good faith assert that their judgment is informed when they are not familiar with how the AI-generated ESG scores came to be, and more so when such scores influence regulatory disclosures or capital allocation.<sup>41</sup> As noted by the researchers, the scope of the board's engagement should include the validation of the training data, check on the bias mitigation tactics, and the insistence on third-party audits to confirm dependability.<sup>42</sup> Without doing so, the opacity of AI would defeat the modern day oversight justifications under the Business Judgment Rule.<sup>43</sup> The EU AI Act clearly reflects this notion: Article 14 requires appropriate human oversight of high-risk systems, and Article 13

---

<sup>38</sup> William M. Lafferty, Lisa A. Schmidt & Donald J. Wolfe, Jr., A Brief Introduction to the Fiduciary Duties of Directors Under Delaware Law, 116 Penn St. L. Rev. 837 (2012).

<sup>39</sup> Vedrana Božić, Environmental, Social, and Governance Approach (ESG) & Artificial Intelligence (AI) (May 2023), <https://www.researchgate.net/publication/370844694>.

<sup>40</sup> The Impact of Artificial Intelligence on ESG: A Conceptual Framework for Practitioners and Policymakers, J. Mgmt. for Glob. Sustainability (May 26, 2025).

<sup>41</sup> VB v. Natsionalna Agentsia za Prihodite, Case C-340/21, EU:C:2023:XXX (CJEU).

<sup>42</sup> Shiyu Feng, Leveraging Artificial Intelligence to Enhance ESG Models: Transformative Impacts and Implementation Challenges, in Proc. 6th Int'l Conf. on Computing & Data Sci. (2024), <https://doi.org/10.54254/2755-2721/69/20241473>.

<sup>43</sup> Dinesh Kumar, Rajinder Kaur & Puneetish Kaur, Legal Analysis of Artificial Intelligence in Corporate Board Rooms, 12 Turk. J. Comput. & Math. Educ. 1514 (2021).

highlights the need for explainability and transparency.<sup>44</sup> These obligations also reflect a wider regulatory approach, which is also found in GDPR Article 22, that offers users human oversight in critical automated decision-making.<sup>45</sup> These legal instruments complement each other in setting a statutory minimum that fiduciary duty would probably incorporate, thus compelling directors' technological oversight. This translates to boards being obligated to request regular audits, KPIs, and "model cards" or equivalent deliverables from ESG-AI vendors.<sup>46</sup> Stipulating the oversight procedure goes beyond merely complying with regulations; it also ensures directors are shielded from accusations of deliberate ignorance.<sup>47</sup> Integrating AI systems into the business processes necessitates directors' active engagement or active questioning. Directors can no longer delegate accountability to vendors or AI teams. In an AI-fueled environment, courts are making it clear that responsible governance calls for oversight structures and a proof of technical knowledge so that boards do not abandon their responsibilities or permit ungoverned AI risks.<sup>48</sup>

### **E. Comparative and Policy Perspective**

The fiduciary expectations of AI systems that focus on environmental, social, and governance (ESG) efforts increase in complexity in the global regulatory context. For a detailed analysis of regulatory environments in different areas. The Corporate Sustainability Reporting Directive (CSRD) in the European Union explicitly covers governance of sustainability disclosures and places the liability of misstatements or omissions on the corporate boards and supervisory bodies. As the AI systems pose a risk to societal and economic well-being, the EU AI Act applies a risk-based regulatory framework to such systems. According to this framework, AI systems may be unacceptable-, high-, or limited-risk, with high-risk systems requiring strict oversight. Under such a framework, failure to oversee algorithmic decision-making exposes the directors to legal risk.<sup>49</sup> The AI Act further requires human oversight, record-keeping, and explainability, which complements fiduciary duties with statutory minimums. Additionally, Article 22 of GDPR restricts fully automated decisions with major legal or financial impacts, which strongly suggests that boards continue to exercise decisive governance over AI systems.<sup>50</sup> The United Kingdom focuses on

---

<sup>44</sup> Regulation (EU) 2024/1689 (AI Act), arts. 13–14.

<sup>45</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), art. 22.

<sup>46</sup> Seung U. Lee et al., *Integrating ESG and AI: A Comprehensive Responsible AI Assessment Framework*, AI & Ethics (2025).

<sup>47</sup> William M. Lafferty, Lisa A. Schmidt & Donald J. Wolfe, Jr., *A Brief Introduction to the Fiduciary Duties of Directors Under Delaware Law*, 116 Penn St. L. Rev. 837 (2012).

<sup>48</sup> *In re Capital One Consumer Data Sec. Breach Litig.*, 488 F. Supp. 3d 374 (E.D. Va. 2020).

<sup>49</sup> Regulation (EU) 2024/1689 (AI Act), arts. 6, 13, 14.

<sup>50</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), art. 22.

directors' duties in the same way through Section 172 of the Companies Act 2006, which requires the directors to work towards the success of the company in the light of its long-term impacts, including environmental and stakeholder considerations.<sup>51</sup> The UK Corporate Governance Code deals with the inner controls and audit committee obligations, which is in tune with EU's approach of ESG scoring, particularly that generated by AI, requiring active supervision by directors.<sup>52</sup> Indian legislation reflects this trend: Section 166 of the Companies Act institutes statutory responsibilities for care, skill, and diligence; concurrently, SEBI's Business Responsibility and Sustainability Reporting (BRSR) framework places stricter requirements on ESG disclosure.<sup>53</sup>

Particularly in the case of AI-generated investor relations that provide misleading or incorrect information, using AI systems of which the companies have no insight into their functioning, boards could incur liability for failing to adhere to their fiduciary responsibilities.<sup>54</sup> Across these jurisdictions, a converging policy trend is evident: ESG data and its precision are no longer issues of reputation to be managed but rather issues of statutory compliance, fiduciary duty, and market integrity to be managed. There is an expectation that directors will examine, document, and audit algorithmic systems rather than pass off supervision to vendors. The developing case law therefore indicates that boards of directors are not freed from responsibility for the inscrutability of algorithms; on the contrary, it increases their duties to enact governance structures that guarantee explainability, accountability, and regulatory compliance. This view in comparison to others highlights a global movement toward a duty of technological competence, which will in turn inform the analysis in Part III on the ways that defective AI ESG systems can convert governance failures into securities fraud or disclosure liability.

## **The New Frontier of Corporate Liability: From Misstatement to Maleficence**

### **A. The Algorithmic Materiality Doctrine: ESG Scores Turning to Securities Fraud**

The anti-fraud provision of the Section 10(b) of the Securities Exchange Act of 1934 supports the federal securities law, prohibiting making any false statement of a material fact or omitting a material fact necessary to make the statements clear, true and not misleading.<sup>55</sup> The fundamental question for the AI-generated ESG scores is whether they institute 'material' statements. An AI-generated ESG score is often promoted by a company in order to attract capital, being a material

---

<sup>51</sup> Companies Act 2006, c. 46, § 172 (UK).

<sup>52</sup> Fin. Reporting Council, UK Corporate Governance Code, Principles M & O (2018).

<sup>53</sup> Companies Act, No. 18 of 2013, § 166 (India).

<sup>54</sup> Satyam Computer Servs. Ltd. v. Union of India, (2009) 1 SCC 47.

<sup>55</sup> Securities Exchange Act of 1934, 15 U.S.C. § 10(b) (1934).

statement of fact, and if such score is being predicted on flawed data or rather a biased algorithm, the dissemination of such score can constitute a securities fraud.<sup>56</sup>

The legal standard for materiality pivots on whether a rational investor would be considering the information crucial in making an investment choice, however such standard is always subject to change with change in market priorities.<sup>57</sup> The Voluntary ESG disclosures have empirically shown to reduce the idiosyncratic risk of a firm, particularly in Initial Public Offerings (IPOs), where the asymmetry of information is higher.<sup>58</sup> Subsequently, a higher ESG score is a commanding signal to the market regarding the quality of risk management by a firm and the long-term viability.<sup>59</sup> Misrepresentation of such influential signal can deceive the investors about the firm-specific risks which they will be assuming. Such dual-lens recognize that external impacts of a company boomerang back to create financial risks. Thus, an AI-generated score that purports to measure a company's ESG standing is material under the classifications: it speaks to the financial resilience of a company and its broader societal impact, both of which are vital to the modern investor's calculus.

The recent enforcements by the U.S. Securities and Exchange Commission (SEC) provide clear evidence that the regulator is prepared to treat ESG-related disclosures as material and actionable. The landmark suit against the Brazilian mining giant Vale S.A. is a stark warning.<sup>60</sup> The SEC had alleged that Vale made false and misleading statements in its ESG disclosures concerning the safety of its dams prior to the catastrophic breakdown of the Brumadinho dam in 2019, which had killed 270 people, causing market capitalization of the company to plummet by over \$4 billion. The complaint explicitly targets the ESG disclosures, contending the company manipulated dam safety audits and acquired fraudulent stability certificates to assure investors, thereby downplaying immense risks.<sup>61</sup> The Vale case shatters the notion that sustainability reports are mere marketing

---

<sup>56</sup> Berg, F., Kölbel, J. F., & Rigobon, R., "Aggregate Confusion: The Divergence of ESG Ratings," *Review of Finance*, vol. 26, no. 1, 2022, pp. 1315-1344.

<sup>57</sup> Investeurope.eu. (2024). What is materiality? | ESG reporting standards guide | Invest Europe. [online] Available at: <https://www.investeurope.eu/invest-europe-esg-reporting-guidelines/assessing-materiality/what-is-materiality/>.

<sup>58</sup> Zou, H. (2024). 'Green or greed? Examining the realities of the ESG investment bubble', *Advances in Economics, Management and Political Sciences*, 72, pp. 296-301.

<sup>59</sup> *Ibid.*

<sup>60</sup> Securities and Exchange Commission of the United States v. Vale, S.A., 1:22-cv-02405, (E.D.N.Y).

<sup>61</sup> Duffy, M.P., Hesse, L.A., III, H.K., Kaushik, S., Levine, S.L., Pandit, A.I., Yadava, N. and Ryan, C.N. (2022). SEC Brings Landmark Suit Over Alleged Misrepresentations in ESG-Related Documents. [online] Jonesday.com. Available at: <https://www.jonesday.com/en/insights/2022/05/sec-brings-landmark-suit-over-alleged-misrepresentations-in-esgrelated-documents> [Accessed 24 Aug. 2025].

puffery, establishing them as legally significant documents that can form the basis of a securities fraud action.<sup>62</sup>

While the Vale case involved egregious and tragic facts, the SEC's action against BNY Mellon Investment Adviser, Inc. (BNYMIA) demonstrates that liability can also arise from process-related misrepresentations.<sup>63</sup> BNYMIA was charged and paid a \$1.5 million penalty for stating in various documents that all investments in certain mutual funds underwent an ESG quality review, when in fact, numerous investments did not have an ESG score at the time of purchase.<sup>64</sup> This precedent is directly applicable to the AI-ESG context. A company that promotes a high ESG score from a proprietary AI model is implicitly representing that a robust, comprehensive analytical process is in place. If that process is flawed, if the algorithm was not applied to all relevant business units or was trained on incomplete data, the resulting score is misleading, mirroring the process failure in the BNYMIA case. A subsequent action against Invesco Advisers, Inc., which resulted in a \$17.5 million penalty for similar misstatements about the extent of 'ESG integration,' reinforces this principle and signals escalating penalties.<sup>65</sup>

## **B. Redefining Materiality in Algorithmic Age: The Unexamined Algorithm & AI Judgment Rule**

Beyond the dominion of deliberate fraud, the deployment of the opaque AI models in ESG analysis creates a significant risk of negligence, specifically a breach of the board's fiduciary duty of care. This obligation being a cornerstone of the corporate law, requires the directors to act on a conversant basis, with meticulousness and skill that a reasonably prudent person would exercise in similar situations. The Business Judgment Rule (BJR) generally armors the directors from the liability for business decisions that turn out terribly, but this protection is predicated on the board having engaged in a reasonable decision-making process.<sup>66</sup> An uncritical reliance on an inscrutable

---

<sup>62</sup> *Ibid.*

<sup>63</sup> Fike, R., Carter, M. and Peloso, M.E. (2022). SEC Charges Investment Advisor with ESG Disclosure Violations in Latest Action by ESG Task Force. [online] Lexology. Available at: <https://www.lexology.com/library/detail.aspx?g=4ecc7c26-d971-48d1-be83-8e8245779843> [Accessed 24 Aug. 2025].

<sup>64</sup> *Ibid.*

<sup>65</sup> Sec.gov. (2024). SEC.gov | SEC Charges Invesco Advisers for Making Misleading Statements About Supposed Investment Considerations. [online] Available at: <https://www.sec.gov/newsroom/press-releases/2024-179>.

<sup>66</sup> Uchicago.edu. (2019). AI & the Business Judgment Rule: Heightened Information Duty | The University of Chicago Law Review. [online] Available at: <https://lawreview.uchicago.edu/online-archive/ai-business-judgment-rule-heightened-information-duty> [Accessed 25 Aug. 2025].

‘black box’ algorithm essentially undermines the predication, representing a potential abdication of the oversight of responsibilities of the board.<sup>67</sup>

A board of directors cannot prorogate to be ‘reasonably informed’ if the primary tool for assessing the risk and guiding the strategy is an algorithm whose entire internal logic, assumptions, and data dependencies remain unknown to them.<sup>68</sup> This leads towards a profound governance crisis, as the individuals charged with the human oversight are rendered incapable of scrutinizing the tools they are employing.<sup>69</sup> This parallels the consensus among the AI developers and legal practitioners that these systems should be used to support, but under no circumstances replace, the human experts, necessitating an expert ‘human-in-the-loop’ to analyze and verify the outputs.<sup>70</sup> This challenge is giving rise to a new legal doctrine scholars have termed the AI Judgment Rule.<sup>71</sup> This idea posits that the standard of care is not static but evolves with technology. It suggests that directors have an assenting duty to make the use of such advanced tools in order to meet their duty of care. A decision made without the available benefit of an AI-driven analysis might no longer be considered reasonably informed, thus stripping the board of BJR protection.<sup>72</sup> However, this heightened duty is a double-edged sword as it is now not merely a duty to use AI, but a duty to use it prudently.

This is where the concept of Explainable AI (XAI) becomes a fiduciary imperative.<sup>17</sup> XAI refers to a set of methods and technologies designed to make the decision-making processes of AI models transparent and understandable to humans.<sup>18</sup> It can highlight which data features most influenced a particular outcome or generate simplified models that approximate the complex logic of the AI, allowing for human-led plausibility checks.<sup>73</sup> The existence and increasing availability of XAI technologies make the continued reliance on opaque black box models legally untenable. XAI is the mechanism that allows directors to fulfill their evolved duty under the AI Judgment Rule. It enables them to move beyond blind deference and engage in genuine oversight: to question the assumptions of AI, probe for potential biases, and understand its limitations.<sup>74</sup> The failure to demand and utilize XAI can be seen as a conscious discarding of the duty to be informed, potentially rising to the level of a gross negligence and exposing directors to personal liability.

---

<sup>67</sup> Blouin, L. (2023). AI’s mysterious ‘black box’ problem, explained. [online] University of Michigan-Dearborn. Available at: <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained> [Accessed 25 Aug. 2025].

<sup>68</sup> Zhou, A. (2016). Analysis of the Volkswagen Scandal Possible Solutions for Recovery. [online] Available at: [https://gps.ucsd.edu/\\_files/faculty/gourevitch/gourevitch\\_cs\\_zhou.pdf](https://gps.ucsd.edu/_files/faculty/gourevitch/gourevitch_cs_zhou.pdf). [Accessed on 25 Aug. 2025].

<sup>69</sup> *Ibid.*

<sup>70</sup> *Ibid.*

<sup>71</sup> Langenbucher, K., 2024. AI Judgment Rule (s). U. Chi. L. Rev. Online, p.1.

<sup>72</sup> *Ibid.*

<sup>73</sup> Dubovitskaya, Elena & Buchholz, Annika. (2024). The Management and the Advice of (Un)Explainable AI. European Company and Financial Law Review. 20. 794-821. 10.1515/ecfr-2023-0033.

<sup>74</sup> *Ibid.*

This confluence of disclosure obligations and diligence duties creates a powerful feedback loop. A board's failure in its duty of care, for instance, by deploying an unexamined, opaque AI model for ESG scoring, is not merely an internal governance lapse.<sup>75</sup> It directly creates the conditions for a subsequent violation of securities law when the flawed, misleading output of that model is published and relied upon by investors.<sup>76</sup> This establishes a direct causal chain from a failure of internal process to externally actionable market harm. The initial breach of the duty of care becomes the root cause of the material misstatement, demonstrating a profound convergence of corporate governance and securities regulation in the algorithmic age.<sup>77</sup> Furthermore, AI acts as an amplifier of liability, as unlike an isolated misstatement by a single executive, a flawed AI-ESG model systematically and repeatedly produces erroneous outputs.<sup>78</sup> The algorithm, often positioned as the 'single source of truth' of an enterprise for sustainability metrics, can become a 'single point of failure'. The efficiency and scale that makes AI attractive to businesses also magnifies the legal and financial consequences when the deployment is not governed by human oversight.

## **The Market Contagion: AI-Driven ESG as a Vector for Systemic Risk**

### **A. Outcomes of Market Concentration: Aggregate Confusion and Herding Behaviour**

The market for ESG data and ratings is not a diverse ecosystem of independent analysts but an oligopoly dominated by a handful of major players, such as MSCI, Sustainalytics, and S&P Global.<sup>79</sup> A critical feature of this market structure is the significant consolidation that has occurred, with many pioneering independent raters being acquired by large, established financial data providers like Moody's and Morningstar.<sup>80</sup> This consolidation concentrates immense power over capital allocation in the hands of a few entities whose AI-driven products act as gatekeepers for trillions of dollars in ESG-mandated funds. A fundamental flaw in this market is what researchers have termed aggregate confusion: the ESG ratings produced by these dominant

---

<sup>75</sup> Feng, S. (2024). "Leveraging artificial intelligence to enhance ESG models: Transformative impacts and implementation challenges", in *Proceedings of the 6th International Conference on Computing and Data Science*.

<sup>76</sup> *Ibid.*

<sup>77</sup> Majekodunmi, T.G. (2025). "Embedding AI in ESG-Financial Reporting Systems to Enhance Non-Financial Disclosure Integrity, Assurance, and Investor Decision Relevance", *International Journal of Advance Research Publication and Reviews*, 2(6), pp. 464-484.

<sup>78</sup> *Ibid.*

<sup>79</sup> Kateryna Myrko (2025). ESG Scores Explained: MSCI vs Sustainalytics vs S&P Global Ratings. [online] FutureGreenWorld. Available at: <https://www.futuregreenworld.com/post/esg-scores-explained-msci-vs-sustainalytics-vs-s-p-global-ratings> [Accessed 25 Aug. 2025].

<sup>80</sup> *Ibid.*

agencies for the same company are often wildly divergent.<sup>81</sup> The correlation between ratings from different providers is remarkably low, ranging from just 0.38 to 0.71, a stark contrast to the 99% correlation observed among credit rating agencies.<sup>82</sup> This divergence stems from fundamental disagreements in three areas: scope (what attributes, like lobbying or water stress, are included), measurement (what indicators are used to quantify an attribute), and weight (how much importance is assigned to each attribute in the final score).<sup>83</sup> One study decomposed this divergence and found that differences in measurement accounted for 56% of the variation, with scope differences contributing 38% and weighting a mere 6%.<sup>84</sup> This variation undermines the credibility of disclosures and limits their utility for investors, who often face challenges comparing ESG performance due to unstructured data formats and opaque rating methodologies.<sup>85</sup>

Paradoxically, this landscape of noisy, conflicting, and opaque signals does not lead to cautious, diversified decision-making. Instead, it can fuel investment herding. Faced with aggregate confusion, asset managers may rationally abandon deep, fundamental ESG analysis, which is made difficult by the data chaos, and instead adopt simplifying heuristics. These can include momentum chasing (investing in companies whose ESG scores are improving), concentrating investments in a small number of “ESG darlings” that score well across multiple rating systems, or simply defaulting to the ratings provided by the largest and most influential agency.<sup>86</sup> These strategies, when adopted by numerous large asset managers simultaneously, lead to correlated capital flows that are detached from a company’s underlying fundamentals.

This dynamic is fertile ground for the formation of ‘green asset bubbles,’ where the market prices of assets deemed sustainable become inflated far beyond their intrinsic value.<sup>87</sup> Behavioural finance provides a powerful lens for understanding this phenomenon. The rush into ESG assets is often driven by a potent mix of cognitive biases and emotional drivers, including overconfidence in the future of green industries, confirmation bias that filters out negative information, and powerful emotional rewards such as the ‘warm glow’ of ethical investing and a pervasive Fear of Missing Out (FOMO) on the next big trend.<sup>88</sup> The market’s tendency to cluster firms into categories like

---

<sup>81</sup> Masulis, R. (2025). Does Common Ownership Raise Antitrust Concerns? [online] The Harvard Law School Forum on Corporate Governance. Available at: <https://corpgov.law.harvard.edu/2025/01/06/does-common-ownership-raise-antitrust-concerns/> [Accessed 25 Aug. 2025].

<sup>82</sup> *Ibid.*

<sup>83</sup> *Ibid.*

<sup>84</sup> *Ibid.*

<sup>85</sup> *Ibid.*

<sup>86</sup> Anton, M., Gine, M. and Schmalz, M.C. (2016). Common Ownership, Competition, and Top Management Incentives. SSRN Electronic Journal. [online] doi: <https://doi.org/10.2139/ssrn.2802332>.

<sup>87</sup> R. Copland, J. (2025). The ESG Bubble Is Bursting. [online] City Journal. Available at: <https://www.city-journal.org/article/esg-investing-funds-energy-economy> [Accessed 25 Aug. 2025].

<sup>88</sup> *Ibid.*

‘ESG Pioneers’ (early, consistent leaders) and ‘ESG Chasers’ (firms rapidly improving their scores) creates predictable narratives that investors can easily follow, further amplifying herd-like capital movements as money flows toward firms perceived to be on an upward ESG trajectory.<sup>89</sup> The result is a market susceptible to volatile swings based on shifts in sentiment and algorithmic outputs rather than stable, long-term value creation.

## **B. Common Ownership and Algorithmic Collusion**

The most pernicious systemic risk emerging from the confluence of AI-ESG and market structure is the potential for these systems to become a powerful mechanism for anti-competitive coordination. This threat is rooted in the phenomenon of ‘common ownership,’ a defining feature of modern capital markets where a few large, diversified institutional investors, notably BlackRock, Vanguard, and State Street (the ‘Big Three’), are simultaneously among the largest shareholders in nearly every major publicly traded company, including direct competitors within concentrated industries.<sup>90</sup> The common ownership thesis posits that when an investor owns significant stakes in, for example, all major airlines, its financial incentive shifts from encouraging one airline to aggressively compete against the others (which would lower prices and erode industry-wide profits) to promoting actions that maximize the value of its entire portfolio of airline stocks.<sup>91</sup> This creates an incentive for ‘soft competition,’ leading to higher prices, reduced output, and a form of tacit collusion that benefits all firms in the industry at the expense of consumers.<sup>92</sup> This is not a fringe theory; it is the subject of a vigorous debate among academics and has attracted significant attention from antitrust regulators globally.<sup>93</sup>

## **C. Antitrust in the Age of AI-ESG**

AI-ESG platforms introduce a novel and highly effective mechanism for common owners to execute on these anti-competitive incentives without engaging in the kind of explicit communication that would trigger antitrust liability. An AI-ESG model, ostensibly designed to promote sustainability, can be programmed, or simply evolve through machine learning, to reward behaviors that are simultaneously ‘green’ and anti-competitive. For example, an algorithm could assign a high positive score for capital discipline or resource efficiency, which in practice means rewarding a company for not building a new, more efficient factory. While framed as reducing

---

<sup>89</sup> *Ibid.*

<sup>90</sup> Perera, H., Lee, S.U., Liu, Y., Xia, B., Lu, Q., Zhu, L., Cairns, J. and Nottage, M., 2024. Achieving responsible AI through ESG: insights and recommendations from industry engagement. *arXiv preprint arXiv:2409.10520*.

<sup>91</sup> *Ibid.*

<sup>92</sup> *Ibid.*

<sup>93</sup> *Ibid.*

environmental impact, this action also serves to restrict industry supply, drive up prices, and increase profits for all incumbent firms in which the common owner is invested.<sup>94</sup> The ESG score provides a neutral, socially acceptable justification for what is, in effect, a collusive outcome. This risk is drawing increasing scrutiny from regulators. State attorneys general have launched investigations into ESG rating agencies like S&P Global, alleging that their methodologies politicize what should be a purely financial decision and may violate consumer protection laws.<sup>95</sup> More pointedly, the Federal Trade Commission (FTC) and the Department of Justice (DOJ) have filed a statement of interest in a multistate antitrust case against the Big Three, alleging a conspiracy to drive down coal production to further ESG goals.<sup>96</sup> In their filing, the agencies explicitly affirmed that asset managers can be held liable under federal antitrust laws when they use their stock holdings in multiple competitors to achieve anti-competitive goals, even when such actions are purportedly justified by social concerns.<sup>97</sup> This signals a clear regulatory intent to pierce the veil of ESG and scrutinize the competitive effects of coordinated investor action, regardless of its motivation.

The structure of modern capital markets, characterized by common ownership, creates an ‘illusion of diversification’ for the end investor. An individual contributing to a 401(k) may believe they are diversified by holding a broad market index fund. However, if the fund is managed by one of the Big Three common owners,<sup>98</sup> and that manager uses a widely adopted AI-ESG platform to guide its voting and engagement across all its holdings,<sup>99</sup> the underlying corporate behavior becomes increasingly harmonized. The algorithm rewards similar strategies across competing firms, strategies that prioritize portfolio-level stability over firm-level competition.<sup>100</sup> The investor is thus left holding a portfolio of companies that are, in strategic terms, converging. This represents a hidden concentration of risk, not the risk of market volatility, but the systemic risk of market stagnation, reduced innovation, and artificially inflated prices for consumers. This dynamic facilitates a sophisticated form of regulatory arbitrage. Antitrust law is well-equipped to prosecute explicit collusion, such as price-fixing cartels or ‘hub-and-spoke’ conspiracies where a common

---

<sup>94</sup> Hendijani Zadeh, M. and Hammami, A., 2025. ESG greenwashing and stock price crash risk: a channel analysis. *Applied Economics*, pp.1-17.

<sup>95</sup> *Ibid.*

<sup>96</sup> *Ibid.*

<sup>97</sup> *Ibid.*

<sup>98</sup> Ken Paxton (2022). Paxton Launches Investigation into S&P Global’s Use of ESG Factors in Credit Ratings, Potentially Violating Consumer Protection Laws. [online] Texas Attorney General. Available at: <https://www.texasattorneygeneral.gov/news/releases/paxton-launches-investigation-sp-globals-use-esg-factors-credit-ratings-potentially-violating> [Accessed 26 Aug. 2025].

<sup>99</sup> *Ibid.*

<sup>100</sup> Reber, B., Gold, A. and Gold, S., 2022. ESG disclosure and idiosyncratic risk in initial public offerings. *Journal of Business Ethics*, 179(3), pp.867-886.

party coordinates agreements among competitors.<sup>101</sup> These cases, however, typically rely on finding a ‘smoking gun’, evidence of an actual agreement. Algorithmic collusion, mediated through a third-party AI-ESG platform, can achieve the same anti-competitive outcome without creating such evidence. Each corporate board can independently decide to follow the recommendations of the AI model to boost its ESG score, knowing full well that its competitors, also owned by the same institutional investors, are receiving the same signals and incentives from the same algorithm. This achieves a coordinated market outcome without a formal agreement, posing a profound challenge to the existing antitrust enforcement framework.

### **Principles of Auditable Algorithmic Governance (AAG): A Possible & Plausible Solution**

The foundational principle of AAG is a strict mandate for explainability. Any AI model used for decisions that have a material impact on a company’s strategy, risk management, or public disclosures, particularly in the ESG context, must be technically explainable.<sup>102</sup> This directly confronts the black box problem that currently enables both unintentional error and intentional deception. Explainable AI (XAI) encompasses a suite of techniques that render an algorithm’s decision-making process transparent to human users, showing, for instance, which data inputs were most influential in reaching a conclusion.<sup>103</sup>

This mandate is not merely a technical specification; it is a legal and fiduciary necessity. By requiring XAI, the AAG framework empowers the board of directors to fulfill its heightened duty of care under the emerging AI Judgment Rule.<sup>104</sup> It allows them to conduct the meaningful plausibility checks that are central to prudent oversight, moving beyond passive acceptance of an AI’s output to active interrogation of its logic.<sup>105</sup> This is particularly critical in the legal domain, where interpretability is considered a primary challenge and a highly desired feature for AI systems to explain the reasoning behind an answer.<sup>106</sup> This transforms the AI from an inscrutable authority into a transparent advisory tool, ensuring that ultimate responsibility for corporate decisions remains with the human fiduciaries.

---

<sup>101</sup> *Ibid.*

<sup>102</sup> Ontiveros, Angel. (2024). “Development and Evaluation of an IT Tool for ESG Compliance Consulting for Companies Using LLMs.” Master’s Thesis, Universität Hamburg.

<sup>103</sup> *Ibid.*

<sup>104</sup> Zhaoyi Li (2024), Artificial Fiduciaries, 81 Wash. & Lee L. Rev. 1299, available at <https://scholarlycommons.law.wlu.edu/wlulr/vol81/iss4/3>.

<sup>105</sup> *Ibid.*

<sup>106</sup> De, J. (2025). Algorithmic Modeling of ESG Contingencies Using Intelligent Analytics and Predictive Parameterization of Corporate Risk. Data-Driven ESG Strategy Implementation Through Business Intelligence, [online] pp.141–172. doi: <https://doi.org/10.4018/979-8-3373-5142-1.ch006>.

The second principle recognizes that even a perfectly transparent algorithm is useless if it operates on flawed data. AAG requires boards to establish and oversee a formal process of algorithmic data diligence, ensuring the quality, provenance, and integrity of the data fed into material AI models.<sup>107</sup> This principle is a direct response to the “Garbage In, Gospel Out” risk and the potential for strategic data curation to generate misleading ESG scores. Such diligence is critical to overcoming persistent data-related gaps in ESG reporting, including the lack of standardized metrics, data fragmentation across internal systems, and limited interoperability between different ESG frameworks.<sup>108</sup> This process would be analogous to the due diligence conducted in financial transactions. It would involve systematic checks for the accuracy, completeness, and potential biases of datasets used to train and operate ESG algorithms. Management would be required to document data sources, collection methodologies, and any preprocessing or cleaning procedures. This would enable the board to make an informed assessment of the risk that the AI model is being misled or is operating on a skewed view of reality. The implementation of interoperable ontologies and semantic fabrics would be crucial, creating a standardized data structure that facilitates consistency and enables effective auditing across diverse data sources, from IoT sensors to financial statements.<sup>109</sup>

The final pillar of the AAG framework is the requirement for periodic, independent audits of critical AI systems by qualified third parties. This provides external validation and assurance to the board, investors, and regulators, creating a crucial accountability mechanism that is currently absent. The lack of audit-grade assurance erodes trust in ESG statements, making it difficult to integrate them into risk models and valuation frameworks.<sup>110</sup> The need for such verification is underscored by research showing significant hallucination rates in LLMs used for legal tasks, reinforcing the consensus that AI tools must be supported by expert human oversight and cannot be trusted to operate without verification.<sup>11</sup> Algorithmic auditing is a nascent but developing field that differs significantly from traditional financial audits, requiring a multidisciplinary approach.<sup>111</sup> A comprehensive algorithmic audit under AAG would include three distinct components:

---

<sup>107</sup> *Ibid.*

<sup>108</sup> *Ibid.*

<sup>109</sup> *Ibid.*

<sup>110</sup> GOV.UK (2022). Auditing algorithms: the existing landscape, role of regulators and future outlook. [online] GOV.UK. Available at: <https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook>.

<sup>111</sup> *Ibid.*

1. **Governance Audit:** A review of the company's policies, procedures, documentation, and human oversight structures related to the AI system. This assesses whether the company has a robust framework for managing algorithmic risk.
2. **Empirical Audit:** Black-box testing of the model's outputs. Auditors would feed the system-controlled inputs to test its performance, accuracy, and stability, and to probe for discriminatory or biased outcomes without needing to inspect the source code. This must include specific tests for algorithmic bias, which can arise from unrepresentative training data and lead to systemic misallocation of capital or reputational damage.
3. **Technical Audit:** An inspection of the algorithm's inner workings. This involves reviewing the source code, model architecture, and underlying mathematical assumptions to ensure the system is technically sound and functions as intended.

The AAG framework can be operationalized through a coordinated set of actions by corporate boards and financial regulators, creating a new standard of practice for the responsible use of AI.

### **For Corporate Boards:**

1. **Charter Amendments:** Boards should amend the charters of their Audit and/or Risk Committees to explicitly grant them oversight responsibility for the firm's use of material algorithmic systems, including those related to ESG.
2. **Board-Level Expertise:** Fulfilling this new oversight duty requires competence. Boards must either recruit directors with expertise in AI and data science or retain qualified independent advisors to help them scrutinize management's use of these technologies and the reports from algorithmic auditors.
3. **Vendor Due Diligence:** When procuring third-party AI-ESG solutions, boards must insist on contractual rights that align with AAG principles. This includes demanding explainable models, full transparency into data sources and methodologies, and rights for independent auditing.

### **For Regulators (e.g., the SEC):**

1. **Interpretive Guidance and Safe Harbors:** The SEC should issue interpretive guidance clarifying that a board's reliance on opaque, unaudited AI models for material ESG disclosures may not satisfy the reasonably informed basis required to invoke the Business Judgment Rule in shareholder litigation. Conversely, the guidance could establish that the

good-faith adoption of a framework like AAG provides evidence of a robust oversight process, creating a potential safe harbor for directors against claims of breaching their duty of care.

2. **Enforcement Modernization:** The SEC's Climate and ESG Task Force should integrate algorithmic auditing techniques into its enforcement toolkit. This would enable investigators to look beyond company disclosures and directly test the underlying AI systems for evidence of algorithmic greenwashing or systemic bias, allowing for more sophisticated and effective enforcement. Regulators should also note that ESG assurance providers are increasingly requiring proof of AI traceability and decision-rationale transparency during third-party audits, creating a market-driven incentive for AAG adoption.

## Conclusion

In conclusion, the use of artificial intelligence to evaluate the ESG compliance is a substantial opportunity to advance, however, it presents a major governance concern. To the investor and regulatory communities, AI can enhance the sustainability disclosures through its speed, accuracy, and predictive capacity. To the unexamined algorithms and black-box models of oversight, the fiduciary responsibilities at the heart of corporate governance are directly menaced. The consequences of the improperly managed ESG disclosures are vividly demonstrated through the SEC's enforcement actions and the Brumadinho dam collapse. In the modern world, the deeply buried ESG risks may no longer be treated as mere reputational concerns; as they can carry enormous legal and financial liability. With insufficient transparency, AI at the firm level creates a responsibility gap, lowers accountability, and increases the risk of liability for directors. At the systemic level, the reliance on a handful of AI-enabled ESG assessment providers creates inequalities, correlated trading, and even implied collusion, weakening competition and market integrity. These behaviors confirm that AI is not the neutral arbiter; it is the powerful gatekeeper that distorts both the capital flows and reputations of companies. The AAG framework advances our current standards by addressing explainability, data management, and third-party audits, offering a clear direction towards an 'Auditable Algorithmic Governance.' At both organizational and market levels, it reintroduces liability by reconciling fiduciary responsibilities with modern technologies. The implication makes itself clear: AI is a tool, not a system; human judgment and oversight of AI is a quintessential aspect of a tool. Genuine sustainability management is not verified by a system's credibility but by real responsibility, making certain that innovation aids, not weakens, the tenets of corporate and market governance.