

Article

Corporate AI Adoption and ESG Decoupling Under China's Dual-Carbon Policy: A Fraud Triangle Analysis

Jincun Fu ¹, Wenqian Gao ², Beibei Zhang ¹ and Jiayao Ye ^{1,*}¹ School of Business, North Minzu University, Yinchuan 750021, China; fujincun86@126.com (J.F.); bbei0107@163.com (B.Z.)² School of Accounting, Lanzhou University of Finance and Economics, Lanzhou 730020, China; gaowq@lzu.edu.cn

* Correspondence: yejiayudy@163.com

Abstract

The credibility of corporate environmental claims is fundamental to achieving China's dual-carbon goals, which aim to peak emissions by 2030 and achieve neutrality by 2060, yet widespread ESG decoupling, a phenomenon where disclosed ESG performance deviates from actual actions, directly undermines policy effectiveness. Guided by the Fraud Triangle Theory, this paper examines whether and how AI adoption curbs decoupling practices. Using data from Chinese A-share listed firms from 2009 to 2024 and mediation analysis, we find that AI adoption significantly reduces ESG decoupling. Specifically, we identify three distinct mediating pathways through which AI exerts its inhibitory effect: it reduces environmental uncertainty and ambiguity, thereby compressing opportunities for managers to engage in decoupling; it heightens media scrutiny and analyst attention, increasing deterrence pressure on firms; and it improves disclosure quality and transparency, undermining the rationalization of decoupling. Heterogeneity analyses reveal stronger effects in firms with R&D-experienced executives, higher reputation, and stronger governance, as well as in highly competitive industries, suggesting that complementary capabilities amplify AI's governance. These findings provide novel evidence on how AI governs environmental information fraud and offer policy implications for using digital technologies to enhance ESG authenticity.

Keywords: artificial intelligence; ESG decoupling; Fraud Triangle Theory; digital governance; China's dual-carbon policy

Academic Editor: Maria Rosa
Trovato

Received: 30 June 2026

Revised: 23 July 2026

Accepted: 25 July 2026

Published: 28 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Under China's "Dual-Carbon" policy framework, which commits the nation to achieving carbon peaking by 2030 and carbon neutrality by 2060, the credibility of corporate environmental, social, and governance (ESG) reporting has become a central regulatory concern. This policy framework creates compliance pressure through punishment threats and reputational incentives through preferential access to green financing. Consequently, firms' motivation to obtain benefits through symbolic environmental claims is significantly strengthened, making ESG decoupling more prevalent and harder to detect. The phenomenon of ESG decoupling refers to the systematic divergence between firms' symbolic environmental claims and their substantive actions, a practice that has increasingly undermined the legitimacy of environmental governance [1,2]. Such decoupling not

only distorts green financial market signals and misallocates capital but also directly hinders the achievement of carbon neutrality objectives.

While ESG disclosure can lower capital costs and facilitate external financing for genuinely environmentally performing firms [3], information asymmetry and profit-driven motives enable some firms to adopt selective disclosure strategies, substituting substantive green investment with symbolic ESG claims [4]. Regulators, constrained by limited resources, struggle to verify the vast volume of self-reported ESG data, and investors face technical barriers in interpreting heterogeneous and often opaque disclosure materials. Traditional governance mechanisms, such as auditing, regulatory inspections, and third-party ratings, have shown limited efficacy in detecting and deterring ESG decoupling, calling for new technological solutions. Although the potential of artificial intelligence (AI) has been preliminarily recognized, prior work has focused on drivers of AI adoption in environmental governance [5], but the governance consequences of AI adoption, particularly its capacity to curb ESG decoupling, remain undertheorized. Moreover, recent studies suggest that AI does not automatically improve governance outcomes; rather, its effectiveness depends on organizational capabilities to absorb, adapt, and apply AI in decision-making [6]. As the world's largest developing economy and manufacturing exporter, China faces acute dual pressures of economic growth and environmental protection. The challenges of "symbol vs. substance" decoupling encountered in its ESG reporting offer important lessons for other emerging market economies. Therefore, examining AI's governance effects on ESG decoupling in the Chinese context is not only policy-relevant but also theoretically generalizable.

The rapid advancement of AI and its adoption in corporate governance offer a promising avenue to mitigate ESG decoupling. Recent studies have documented that AI adoption significantly curbs corporate ESG decoupling [7]. However, these studies largely focus on single mediating pathways and have not integrated AI's multi-faceted governance potential into a systematic theoretical framework. To solve this problem, frameworks such as institutional theory and signal theory are available, but while institutional theory can reveal firms' isomorphic responses to institutional pressures, it cannot explain the internal operational conditions that enable decoupling behaviors. Signal theory focuses on information transmission processes, but it fails to address the underlying drivers of firms' deliberate fabrication of false signals. By contrast, the Fraud Triangle Theory from criminology and organizational behavior decomposes decoupling into three distinct pathways: pressure, opportunity, and rationalization [8]. This framework offers a more refined theoretical lens for understanding how AI can simultaneously act on the multiple drivers of ESG decoupling. Building on this theory, we conjecture that AI may restrain ESG decoupling through three distinct pathways: strengthening internal control (opportunity dimension), enhancing external oversight (pressure dimension), and improving disclosure quality (rationalization dimension), which serves as a useful complement to existing research. The corresponding research hypotheses will be developed in the following section.

Our empirical analysis, using data from Chinese A-share listed firms from 2009 to 2024 and following the two-step causal mediation approach [9], yields several key findings. AI adoption significantly curbs ESG decoupling, a result that remains robust to a series of endogeneity and robustness checks, including instrumental variable estimation. All three mediating pathways within the fraud triangle theory framework are statistically significant, with particularly pronounced effects for external oversight and disclosure quality. Heterogeneity analyses show that the inhibitory effect of AI is stronger in firms with R&D-experienced executives, in more competitive industries, and in regions with more developed factor markets and stricter environmental regulations.

The remainder of this paper is structured as follows. Section 2 presents the institutional background and theoretical analysis. Section 3 describes the research design. Section 4 reports the analysis of empirical results. Section 5 reports the mechanism test. Section 6 reports the heterogeneity analysis. Section 7 offers conclusions, and Section 8 offers policy suggestions and discusses study limitations.

2. Institutional Background and Theoretical Analysis

2.1. Institutional Background

After decades of rapid growth, China's economy faces mounting environmental pressures. Total carbon emissions have continued to rise, and the traditional extensive development model has become unsustainable. In response, the Chinese government proposed the "dual-carbon" targets in 2020—carbon peaking by 2030 and carbon neutrality by 2060. In 2021, the State Council issued the Action Plan for Carbon Peaking Before 2030, mandating key polluting industries to establish emission monitoring systems and imposing penalties for non-compliance. Concurrently, the Measures for the Administration of Corporate Environmental Information Disclosure require listed firms to disclose detailed environmental performance indicators and undergo third-party verification [10,11]. These regulations increase firms' incentives to obtain compliance benefits through symbolic claims, creating institutional drivers for ESG decoupling. However, the gap between emission reduction targets and firms' actual abatement capacity has incentivized "symbolic compliance." When regulation intensity rises but enforcement resources remain limited, some firms meet requirements through selective disclosure rather than substantive action [12].

Under this policy framework, regulators have strengthened ESG disclosure guidance. In 2018, the China Securities Regulatory Commission revised the Corporate Governance Code, establishing a basic ESG disclosure framework [13]. In 2022, the State-owned Assets Supervision and Administration Commission mandated full ESG reporting for all central-government-controlled listed firms by 2023 [14]. These measures have spurred ESG rating agencies (e.g., SynTao Green Finance, Huazheng), generating standardized datasets covering the A-share market [15]. Approximately 18% of listed firms exhibit measurable decoupling, with higher rates in regions with weaker regulation [16,17]. Notably, ESG ratings from different agencies often diverge significantly [18]. This divergence provides room for "rating arbitrage"—firms may selectively disclose favorable information or cooperate with higher-rated agencies [19].

Meanwhile, AI policies have aligned with ESG. Since 2017, the State Council has established multiple "Artificial Intelligence Innovation and Development Pilot Zones," and the 14th Five-Year Plan promotes "AI+" initiatives [12]. In 2024, the "Data Elements×" Three-Year Action Plan encouraged using AI for environmental monitoring, carbon accounting, and ESG verification [20]. These policies provide incentives for firm-level AI adoption and create quasi-natural experimental conditions for examining AI's governance effects. As of 2024, over 60% of Chinese A-share listed firms mentioned AI-related applications in their annual reports [16,17].

However, traditional environmental regulation tools—such as periodic inspections, ex-post penalties, and third-party audits—struggle to cope with the massive volume of unstructured ESG information [4]. Regulatory resources are limited, and firms enjoy considerable discretion in disclosure, allowing symbolic compliance to persist. The institutional environment described above—rigorous emission reduction pressure, information divergence in ESG ratings, and the limitations of traditional governance tools—sets the stage for studying AI's governance role. A central question emerges: Given that the dual-

carbon policy demands genuine emission reduction yet enforcement constraints encourage symbolic compliance, can AI help resolve this contradiction, and if so, through what mechanisms?

2.2. Theoretical Framework

To answer the above questions, we introduce the fraud triangle theory [8]. Unlike perspectives that explain greenwashing solely through information asymmetry or institutional pressure, the fraud triangle theory simultaneously encompasses internal motivation (pressure), external conditions (opportunity), and psychological factors (rationalization). Thus, it is a more suitable unified framework for analyzing AI's governance mechanisms [21,22].

ESG decoupling—the systematic divergence between firms' symbolic environmental claims and substantive actions—is essentially selective disclosure or exaggeration driven by information asymmetry [2]. This behavior exhibits high structural isomorphism with classic financial fraud: both involve information manipulation, both aim to mislead stakeholders, and both seek to evade punishment or obtain undue benefits. Therefore, applying the fraud triangle theory to analyze ESG decoupling has a solid logical foundation.

The theory posits that fraud occurs when three conditions converge: opportunity (institutional or procedural loopholes), pressure (external incentives such as performance targets or financing needs), and rationalization (subjective justification that alleviates moral guilt). In the ESG decoupling context, information asymmetry and managerial discretion constitute “opportunity”; capital market pressure, regulatory scrutiny, or internal performance targets constitute “pressure”; and managers' belief that “greenwashing will not be detected” constitutes “rationalization.”

The theory has been successfully applied to explain corporate greenwashing. For example, Kurpierz and Smith (2020) built a “greenwashing triangle” framework [21]; Yu and Hwang (2026) mapped board attributes to pressure, opportunity, and rationalization [23]; Ren et al. (2025) found that ESG rating disagreement induces greenwashing through financing constraints (pressure), information asymmetry (opportunity), and managerial myopia (rationalization) [18].

Recent work further suggests that the governance effects of AI are not automatic but rather contingent on firms' organizational capabilities. Han et al. (2026) demonstrate that AI adoption strengthens firms' absorptive, adaptive, and innovative capabilities—collectively referred to as dynamic capabilities—which in turn enable more effective strategic decision-making [6]. This perspective implies that AI's impact on ESG decoupling is unlikely to be direct or mechanical; instead, it operates through the enhancement of organizational processes that allow firms to better absorb, interpret, and act upon environmental information.

Based on this framework, we infer that AI may curb ESG decoupling through three distinct pathways, corresponding to the three dimensions of the fraud triangle. First, compressing opportunity: AI strengthens internal control systems, reducing the scope for selective disclosure. Second, increasing pressure: AI enhances the information processing capabilities of analysts and media, raising the probability of detecting greenwashing and the associated reputational costs. Third, undermining rationalization: AI improves disclosure transparency and verifiability, weakening managers' belief that greenwashing will go undetected. We now derive the research hypotheses.

2.3. Research Hypotheses

2.3.1. Direct Effect of AI Adoption on ESG Decoupling

AI offers a powerful direct pathway to mitigate ESG decoupling through its data sensing, processing, and analytical capabilities. According to signaling theory [24], in information-asymmetric markets, firms have incentives to release embellished ESG signals to influence stakeholder perceptions. Traditional governance mechanisms struggle to cope with the massive volume of unstructured corporate disclosures, whereas AI can simultaneously intervene at the production, internal governance, and external governance levels. Importantly, the effectiveness of AI in this regard is likely to depend on firms' internal capabilities to absorb and integrate AI-driven insights into decision-making processes [6]. Firms with stronger dynamic capabilities, such as those with R&D-experienced leadership or robust governance structures, may be better positioned to translate AI adoption into tangible ESG improvements.

At the production level, deep AI integration into manufacturing enables real-time optimization of energy consumption, material flow, and emissions. For instance, machine-learning-based intelligent control systems can adaptively adjust production parameters to minimize carbon emissions while maintaining output [25]. Such efficiency gains reduce the inherent need for firms to exaggerate their environmental claims.

At the internal governance level, AI-driven decision support systems integrate real-time operational data, external regulations, and internal strategic goals, embedding ESG considerations into core decision variables [26]. Intelligent process automation standardizes environmental data collection, accounting, and reporting, substantially reducing human errors and manipulation opportunities. This technological approach raises the technical barriers to internal information manipulation.

At the external governance level, through natural language processing and complex network analysis, AI can automatically identify vague promises, inconsistencies, and contradictions in ESG reports and cross-validate them with multiple data sources [27]. This capability empowers analysts, media, and regulators: AI-enhanced quantitative models enable more accurate assessment of environmental risk; data mining tools make investigative reporting more penetrating; and AI-based compliance platforms enable real-time anomaly detection. Decoupling behaviors become more easily identifiable, verifiable, and disseminable within technology-enhanced oversight networks, imposing substantial potential costs—reputational loss, deteriorating financing conditions, and increased regulatory penalties—that create a strong deterrent effect on management.

Moreover, AI-driven intelligent decision systems are reshaping the underlying logic of ESG. By optimizing operational decisions based on real-time environmental data and establishing dynamic risk assessment models, AI shifts firms from passive compliance to active governance [23]. In summary, AI curbs ESG decoupling through three sub-pathways: production-source optimization, internal governance solidification, and external oversight intensification. Based on the above analysis, we propose:

Hypothesis 1 (H1). *AI adoption curbs corporate ESG decoupling.*

2.3.2. Opportunity Dimension: The Mediating Role of Internal Control

Internal control quality is a key institutional arrangement constraining information manipulation. Effective internal control—through segregation of duties, authorization and approval, and regular reconciliations—reduces the scope for selective disclosure [28]. High-quality internal control significantly reduces ESG decoupling by institutionalizing constraints on managerial discretion [29]. However, traditional internal control often relies on manual judgment and supervision, which suffer from implementation frictions and

institutional inertia. In this context, AI offers a promising way forward, yet its effectiveness depends on the quality of human–AI collaboration. In other words, algorithmic detection must be integrated with managerial judgment and corrective action [5].

AI can strengthen internal control systems in three significant ways. First, process automation. AI-driven automated approval and process control reduce manual intervention, standardizing environmental data collection, accounting, and reporting, thereby lowering the possibility of deliberate manipulation [30]. Second, real-time monitoring and data triangulation. AI's real-time monitoring capabilities can automatically compare information from different sources (e.g., carbon emission data versus production records, energy purchases versus emission factors) to detect anomalies or inconsistencies [30]. This automated cross-validation makes any “window-dressing” difficult to conceal. Third, traceability and accountability. AI's record-keeping mechanisms fully track every data modification and approval, enhancing the accountability function of internal controls and making any managerial changes fully auditable [24]. These technological constraints significantly raise the expected cost of engaging in decoupling.

When internal control is strengthened, firms' opportunities to engage in ESG decoupling are significantly compressed. Management can no longer exaggerate environmental performance or conceal unfavorable information without being detected [31]. Therefore, AI may curb ESG decoupling by improving internal control effectiveness. Accordingly, we propose:

Hypothesis 2 (H2). *AI adoption curbs corporate ESG decoupling by improving internal control effectiveness.*

2.3.3. Pressure Dimension: The Mediating Role of External Oversight

External oversight is an important governance mechanism that constrains corporate behavior. Chen et al. (2026) confirm that external monitors can discipline corporate greenwashing when they acquire better information about firms [32]. Analysts and the media, as information intermediaries, exert pressure through information discovery and reputational sanctions [33,34]. Empirical research shows that analyst coverage significantly curbs greenwashing, especially in non-heavily polluting industries [35,36]. Social media attention also influences greenwashing by exacerbating managerial myopia [37], indicating that external attention substantively affects disclosure behavior. However, traditional external oversight is limited by information processing capacity and struggles to handle the ever-growing volume of ESG data.

AI can enhance external oversight in three dimensions [37]. First, sentiment monitoring and signal identification. AI-assisted sentiment monitoring systems can track media coverage and social media discussions in real time, quickly identifying potential greenwashing signals. NLP-based sentiment analysis can automatically assess public opinion trends, providing early warnings for investors. Second, data analytics and consistency assessment. AI-driven data analytics tools enable analysts to process massive amounts of ESG data and more accurately assess the consistency between corporate disclosures and actual performance. This technological capability makes previously undetectable “talk-action” gaps quantifiable and comparable. Third, standardization and anomaly detection. AI-automated report generation improves the standardization of information disclosure, facilitating comparative analysis and anomaly detection by external monitors. When external oversight intensity increases, the reputational risk and expected regulatory punishment associated with greenwashing rise correspondingly [10]. This pressure dampens management's motivation to engage in greenwashing and encourages substantive environmental actions rather than symbolic claims [38]. Therefore, AI may curb ESG decoupling by enhancing external oversight intensity. Accordingly, we propose:

Hypothesis 3 (H3). *AI adoption curbs corporate ESG decoupling by enhancing external oversight intensity.*

2.3.4. Rationalization Dimension: The Mediating Role of Disclosure Quality

Disclosure quality directly affects whether managers believe that greenwashing will go undetected. When the disclosure regime is robust and transparent, management finds it difficult to mask true behavior through vague language or selective disclosure, thereby compressing the psychological space for “rationalization” [39]. High-quality disclosure exposes the gap between symbolic claims and substantive actions, thereby curbing decoupling. However, traditional disclosure relies on firms’ voluntary reporting and limited third-party verification, which cannot guarantee information authenticity and completeness. Rationalization in the fraud triangle framework operates through self-deception—managers justify decoupling with narratives such as “this is just a temporary discrepancy” or “our practices are no worse than industry norms.” AI-enhanced disclosure quality disrupts this process by eliminating the ambiguity that sustains such self-deception. When ESG data becomes precise, real-time, and traceable, the gap between claims and actual performance becomes visibly unambiguous, leaving managers with no room to convince themselves that the discrepancy is justified.

AI can improve disclosure quality in three ways, thereby undermining management’s “rationalization” that greenwashing will go unnoticed. First, semantic analysis and consistency checking. Natural language processing can automatically identify ambiguous expressions, exaggerated claims, or logical inconsistencies in ESG reports [40]. For example, AI can detect a contradiction between “We are committed to reducing carbon emissions” and an actual increase in emissions, flagging it as a potential greenwashing signal. Second, multi-source data cross-validation. AI can cross-validate data from different sources (e.g., self-reported data versus government monitoring data, industry benchmarks), enhancing information reliability and comparability [40,41]. This validation mechanism makes it difficult for firms to maintain inconsistent claims across different channels. Third, automated disclosure and auditing. AI-assisted automated disclosure systems reduce the scope for manual manipulation and enhance information consistency and traceability [9,42]. AI audit tools can continuously monitor disclosure trends and promptly detect abnormal fluctuations.

When disclosure quality improves, management’s rationalization that “greenwashing will not be discovered” is undermined. Managers become aware that any false statement can be identified by AI systems and exposed to public and regulatory scrutiny. Therefore, AI may curb ESG decoupling by improving disclosure quality. Accordingly, we propose:

Hypothesis 4 (H4). *AI adoption curbs corporate ESG decoupling by improving disclosure quality.*

3. Research Design

3.1. Data Sources and Sample Selection

This paper uses Chinese A-share listed firms from 2009 to 2024 as the initial sample. This time window adequately captures the evolution of AI policies and the development of AI technologies. In 2015, China first identified AI as a priority development area; in 2017, AI was formally elevated to a national strategy with a clear “three-step” roadmap; starting in 2018, a series of policies promoting the deep integration of the digital and real economies were launched; and in 2022, “digital–real integration” was written into the report of the 20th National Congress of the Communist Party of China.

ESG actual performance ratings come from the Huazheng ESG rating; ESG disclosure scores come from the Bloomberg database; other firm-level data are from the CSMAR database. Sample screening excludes (1) financial firms (to maintain homogeneity), (2) ST, *ST, and delisted firms, and (3) observations with missing data. The final sample comprises 14,033 firm-year observations. All continuous variables are winsorized at the 1st and 99th percentiles.

3.2. Variable Definitions

Dependent variable: ESG decoupling (*Decou*). Following Hu et al. (2023) and Yu et al. (2020), we measure the divergence between a firm's ESG disclosure and its actual performance using the standardized difference between the two scores [10,43]. For the “saying” dimension (*disclosure*), we use the Bloomberg ESG disclosure score, which has covered global listed firms since 2010 and achieved full A-share coverage after 2015, focusing on the completeness and standardization of disclosure [44]. For the “doing” dimension (actual performance), we use the Huazheng ESG rating, which has covered all A-shares since 2009 and balances disclosure, performance, and compliance, offering a continuous and reliable measure of firms' true ESG outcomes.

To address the scale difference between the Bloomberg and Huazheng ESG scores, we follow Walker and Wan (2012) and standardize both measures using Z-scores before taking their difference [45]. A positive *Decou* indicates that the firm's ESG disclosure is better than its actual performance, implying decoupling or greenwashing (the larger the value, the more severe). A negative *Decou* indicates that actual performance exceeds disclosure, characterizing a “substantive but low-profile” firm.

Independent Variable: AI application level. The annual report is the official channel through which firms communicate strategic orientation and operational priorities. The frequency of AI-related terms in the annual report directly reflects the firm's strategic emphasis on and tendency to adopt AI. Following the measurement approach of Acemoglu et al. (2022) for firm-level AI adoption [46], we construct an AI-related term dictionary using Python 3.9 natural language processing. We count the number of AI-related words in the annual reports of Shanghai and Shenzhen A-share listed firms from 2010 to 2024, add one, and take the natural logarithm (*Lnwords*) as the measure of firm-level AI application.

While text-based measures are widely adopted, a key concern is that keyword frequency may capture firms' rhetorical emphasis on AI rather than actual AI implementation—a limitation inherent to such measures and widely recognized in the literature [47]. Recent studies have distinguished between “nominal AI adoption” (disclosure about AI) and “substantive AI adoption” (verifiable AI activities), noting that textual measures may reflect both [48]. Nevertheless, textual analysis remains the most widely used approach for measuring firm-level AI adoption in large-sample studies due to its feasibility, scalability, and consistency [49]. To mitigate this concern, we include AI patent applications (*AI_patent*) as an alternative measure in our robustness checks.

Building on this robustness strategy, we further calculated the Pearson correlation coefficient between *Lnwords* and *AI_patent*. The correlation is positive and statistically significant ($r = 0.2928$, $p < 0.01$), suggesting that firms with greater AI disclosure also tend to engage in more substantive AI innovation, at least in part.

Control Variables: Drawing on existing literature, we include the following control variables: financing constraint index (*SA*), firm age (*Age*), firm size (*Size*), revenue growth rate (*Growth*), total asset turnover (*ATO*), return on assets (*ROA*), leverage ratio (*Lev*), ownership concentration (*Top5*), board size (*Board_Scale*), and CEO duality (*Dual*).

Definitions and measurement methods are detailed in Table 1.

Table 1. Variable definitions.

Variable	Meaning	Measurement Method
<i>Decou</i>	ESG Decoupling	Standardized ESG disclosure score minus standardized ESG actual performance score (Z-score standardized before subtraction)
<i>Lnwords</i>	AI application level	$\ln(\text{number of AI-related keywords in the annual report} + 1)$
<i>SA</i>	Financing constraint index	$SA = -0.737 \times \text{firm size} + 0.043 \times \text{firm size}^2 - 0.040 \times \text{firm age}$
<i>Age</i>	Firm age	Observation year minus founding year
<i>Size</i>	Firm size	Natural logarithm of year-end total assets
<i>Growth</i>	Revenue growth rate	$(\text{Current year revenue} / \text{Prior year revenue} - 1) \times 100\%$, measuring firm growth
<i>ATO</i>	Total asset turnover	Sales revenue/average total assets, reflecting asset operation efficiency
<i>ROA</i>	Return on assets	Net profit/average total assets, measuring profitability
<i>Lev</i>	Leverage ratio	Year-end total liabilities/total assets, reflecting debt risk
<i>Top5</i>	Ownership concentration	Shareholding ratio of the top five largest shareholders, supplementing ownership structure
<i>Board_Scale</i>	Board size	Natural logarithm of total number of board members, reflecting corporate governance level
<i>Dual</i>	CEO duality	Equals 1 if the chairman and CEO are the same person, otherwise 0

Note: All continuous variables are winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

3.3. Descriptive Statistics

Table 2 reports the descriptive statistics for the main variables used in our analysis. The final sample consists of 14,033 firm-year observations from 2009 to 2024. The mean value of ESG decoupling (*Decou*) is -0.2859 , with a standard deviation of 1.2022, indicating substantial variation in the degree of decoupling across firms. The negative mean suggests that, on average, firms in our sample tend to have actual ESG performance slightly exceeding their disclosure levels, though a positive maximum (5.7874) shows that severe greenwashing exists for some firms. The AI application level (*Lnwords*) has a mean of 0.7614 and a median of 0.0000, implying that a considerable portion of firms have not yet adopted AI-related terms in their annual reports, while the maximum of 5.9839 reflects that some firms have a strong strategic emphasis on AI.

Regarding firm characteristics, the mean of the financing constraint index (*SA*) is -3.8517 , with a relatively small standard deviation (0.2807), suggesting that most sample firms face similar financing conditions. Firm size (*Size*) averages 23.2357, consistent with the typical scale of Chinese A-share listed firms. Profitability (*ROA*) averages 0.0425, and leverage (*Lev*) averages 0.4772, indicating moderate debt levels and reasonable profitability. Ownership concentration (*Top5*) has a mean of 54.5149, showing that the largest five shareholders hold a majority of shares on average. CEO duality (*Dual*) has a mean of 0.2114, meaning that about 21% of sample firms have the chairman and CEO positions held by the same person. All continuous variables have been winsorized at the 1st and 99th percentiles to mitigate the influence of outliers.

Table 2. Descriptive statistics.

Variable	N	Mean	SD	Min	Median	Max
<i>Decou</i>	14,033	-0.2859	1.2022	-3.8703	-0.3617	5.7874
<i>Lnwords</i>	14,033	0.7614	1.1150	0.0000	0.0000	5.9839
<i>SA</i>	14,033	-3.8517	0.2807	-4.6977	-3.8556	-2.9739
<i>Age</i>	14,033	2.5223	0.6597	0.6931	2.7081	3.5553

<i>Size</i>	14,033	23.2357	1.2868	19.4149	23.1543	26.4523
<i>Growth</i>	14,033	0.1562	0.4120	−0.6735	0.0972	5.0755
<i>ATO</i>	14,033	0.6341	0.4376	0.0467	0.5360	2.6445
<i>ROA</i>	14,033	0.0425	0.0613	−0.5562	0.0378	0.2223
<i>Lev</i>	14,033	0.4772	0.1989	0.0298	0.4884	0.9104
<i>Top5</i>	14,033	54.5149	16.6264	10.6459	54.2955	99.2300
<i>Board_Scale</i>	14,033	2.1740	0.2007	1.6094	2.1972	2.7081
<i>Dual</i>	14,033	0.2114	0.4083	0.0000	0.0000	1.0000

Note: This table reports the descriptive statistics of the main variables. *Decou* is the ESG decoupling measure (standardized disclosure minus standardized performance). *Lnwords* is the natural logarithm of one plus the number of AI-related keywords in the annual report. *SA* is the financing constraint index. All continuous variables are winsorized at the 1st and 99th percentiles. Definitions of all variables are provided in Table 1.

3.4. Model Specification

To test Hypothesis H1, we construct the following baseline regression model:

$$Decou_{i,t} = \alpha_0 + \alpha_1 Lnwords_{i,t} + \alpha_2 Controls_{i,t} + \sigma_i + \delta_t \quad (1)$$

where i indexes firms and t indexes years. $Decou_{i,t}$ is the ESG inconsistency index of firm i in year t . $Lnwords_{i,t}$ is the natural logarithm of (the number of AI-related keywords in the annual report of firm i in year t plus 1); a larger value indicates a higher level of AI application. $Controls_{i,t}$ a vector of $Controls_{i,t}$ variables. σ_i and δ_t represent industry fixed effects and year fixed effects, respectively. $\varepsilon_{i,t}$ is the idiosyncratic error term.

4. Analysis of Empirical Results

4.1. Baseline Regression Results

Table 3 reports the baseline regression results examining the impact of AI adoption on ESG decoupling. Column (1) includes only year fixed effects; Column (2) adds industry fixed effects. The coefficient of *Lnwords* in Column (2) is −0.0723, significant at the 1% level. Column (3) further adds all control variables. The coefficient remains significantly negative at the 1% level (−0.0613), and the adjusted R^2 increases from 0.3588 in Column (1) to 0.4371 in Column (3), indicating that the inclusion of controls improves the model's explanatory power.

In terms of economic significance, based on Column (3), a one-standard-deviation increase in *Lnwords* (1.1150) reduces ESG decoupling by approximately 0.0613, which corresponds to about 5.1% of the standard deviation of the dependent variable. This suggests a modest but statistically detectable governance effect of AI adoption, consistent with the view that AI serves as one of multiple factors influencing ESG decoupling rather than a dominant driver. Among the control variables, profitability (*ROE*) and board size (*Board_Scale*) are negatively associated with ESG decoupling, consistent with the view that better financial performance and stronger governance structures help firms substantiate their ESG commitments. In contrast, ownership concentration (*Top5*) becomes positively significant in the full model, suggesting that large shareholders may, under certain circumstances, over-promise to maintain market image while lacking corresponding execution discipline.

Taken together, these results provide strong support for Hypothesis H1 that AI adoption significantly curbs corporate ESG decoupling.

Table 3. Baseline regression results.

Variable	(1) Decou	(2) Decou	(3) Decou
<i>Lnwords</i>	−0.1251 *** (0.0141)	−0.0723 *** (0.0149)	−0.0613 *** (0.0143)
<i>SA</i>			0.0201 (0.0703)
<i>Age</i>			0.1043 *** (0.0284)
<i>Size</i>			−0.0402 ** (0.0176)
<i>Growth</i>			0.1290 *** (0.0215)
<i>ATO</i>			0.0075 (0.0402)
<i>Roe</i>			−0.3507 *** (0.1239)
<i>Roa</i>			−0.5039 (0.3726)
<i>Lev</i>			0.7272 *** (0.0949)
<i>Top5</i>			0.0028 *** (0.0010)
<i>Board_Scale</i>			0.1468 ** (0.0698)
<i>Dual</i>			0.1101 *** (0.0320)
Constant	−0.1907 *** (0.0206)	−0.2322 *** (0.0184)	−0.3127 (0.5265)
Observations	14,018	14,018	14,018
Adjusted R ²	0.3588	0.4147	0.4371
Industry FE	YES	YES	YES
Year FE	YES	YES	YES

Note: Robust t-statistics are in parentheses. *** and ** denote significance at the 1% and 5% levels, respectively. Column (1) controls for year fixed effects only; column (2) adds industry fixed effects; column (3) further includes all control variables.

4.2. Endogenous Test

To address potential endogeneity concerns—including reverse causality, omitted variables, measurement error, and self-selection—we employ three complementary approaches: instrumental variable (IV) estimation, the Heckman two-step procedure, and lagged regressions. Tables 4 and 5 summarize the results.

Table 4. Endogeneity tests: Instrumental variable estimation.

Variable	(1) IV1 1st	(2) IV1 2st	(3) IV2 1st	(4) IV2 2st	(5) IV3 1st	(6) IV3 2st
<i>IV_IndustryMean</i>	0.3379 *** −0.0459					
<i>IV_Broadband × L1_AI</i>					0.1342 *** −0.0239	
<i>Lnwords (fitted)</i>		−0.5255 *** −0.1159		−0.0873 ** −0.0382		−0.2998 *** −0.0968

Observations	13,911	13,911	14,018	14,018	11,183	11,183
First-stage F	181.12		214.93		309.8	
Cragg-Donald Wald F	182.31					
Sargan <i>p</i> -value			0.531			
Controls	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Note: Standard errors are in parentheses. *** and ** denote significance at the 1% and 5% levels, respectively. All continuous variables are winsorized at the 1st and 99th percentiles.

Table 5. Endogeneity tests: Heckman selection model and lagged variables.

Variable	(1) Heckman	(2) Lag1	(3) Lag2
<i>Lnwords</i>	−0.0616 *** −0.01		
<i>IMR</i>	1.0934 ** −0.4316		
<i>L. Lnwords</i>		−0.0747 *** −0.0172	
<i>L2. Lnwords</i>			−0.0730 *** −0.0199
Observations	14,020	11,320	9405
Controls	YES	YES	YES
Industry FE	YES	YES	YES
Year FE	YES	YES	YES

Note: Standard errors are in parentheses. *** and ** denote significance at the 1% and 5% levels, respectively. All continuous variables are winsorized at the 1st and 99th percentiles.

4.2.1. Instrumental Variable Estimation

The relationship between AI adoption and ESG decoupling could be confounded by reverse causality and omitted variables. We therefore construct three instrumental variables to account for this endogeneity.

IV1: Industry-year average of AI adoption (excluding the focal firm). The peer imitation effect implies that the overall industry AI transformation trend positively drives individual firms' AI investment. The leave-one-out mean remains highly correlated with firm-level AI adoption, satisfying the relevance condition theoretically. Meanwhile, the industry-average AI intensity is determined by the behavior of other firms within the industry. An individual firm's ESG decoupling behavior is unlikely to alter the industry-wide technological landscape. Industry-level technology shocks operate through AI adoption channels and have no direct pathway to affect ESG decoupling, thus satisfying the exogeneity condition.

Following the literature on technology adoption spillovers [50], we use the leave-one-out industry-year mean of *Lnwords* as an instrument, denoted *IV_IndustryMean*. A firm's AI adoption is influenced by the average AI adoption level in its industry through peer effects, competitive pressures, and common technology shocks. This industry-level instrument is unlikely to be directly correlated with an individual firm's ESG decoupling, satisfying the exclusion restriction. Column (1) of Table 4 presents the first-stage result: the instrument has a positive and highly significant coefficient (0.3379, $p < 0.01$). The first-stage F-statistic is 181.12, and the Cragg-Donald Wald F-statistic is 182.31, far exceeding the Stock-Yogo critical value of 16.38 for 10% maximal IV bias, rejecting weak identification. The sample size drops to 13,911 because observations from industry-year cells with only

one firm cannot generate a leave-one-out mean. The second-stage coefficient of *Lnwords* in Column (2) is -0.5255 ($p < 0.01$), confirming the negative effect of AI on ESG decoupling.

IV2: Lewbel heteroskedasticity-based instruments [51]. The heteroskedasticity in the disturbance term, constructed by interacting exogenous control variables with residuals from the endogenous variable regression, exhibits a natural statistical association with firm-level AI adoption, theoretically satisfying the relevance condition. Meanwhile, these instruments are generated internally from model variables without directly accessing external determinants of ESG decoupling. They remain immune to interference from micro-level decisions such as ESG disclosure manipulation or performance bias, thus satisfying the exogeneity condition.

To validate this approach, we first conduct the Breusch-Pagan test for heteroskedasticity, which rejects the null of homoskedasticity at the 1% level, confirming the presence of heteroskedasticity required for the Lewbel procedure. Column (3) of Table 4 reports the first-stage results; because Lewbel generates multiple internal instruments, we do not present individual coefficient estimates but report the joint F-statistic, which is 214.93 ($p < 0.001$), well above the Stock-Yogo critical value, rejecting weak instrument concerns. The second-stage coefficient of *Lnwords* in Column (4) is -0.0873 ($p < 0.05$). The Sargan test of overidentifying restrictions yields a p -value of 0.531, confirming the validity of the instruments.

IV3: Lagged city-level broadband penetration $\times$ lagged AI adoption. Lagged broadband scale reflects regional digital infrastructure levels, which jointly enable firms' AI development foundations and reduce digital transformation costs, theoretically satisfying the relevance condition. Meanwhile, broadband construction is guided by local public planning, and the lagged specification mitigates reverse causality concerns. Regional infrastructure cannot directly alter firm-level ESG decoupling, thus satisfying the exogeneity condition.

Drawing on the literature that uses broadband infrastructure as an instrument for digital technology adoption [52], we construct a instrument by interacting the one-period lag of city-level broadband subscribers (per 100 persons) with the one-period lag of *Lnwords*, both standardized. This instrument is denoted *IV_Broadband* $\times$ *L1_AI*. The intuition is that regional broadband infrastructure reflects the local digital environment that facilitates firm-level AI adoption, while being plausibly exogenous to individual firms' ESG decoupling. Using lagged values further mitigates reverse causality concerns. Column (5) of Table 4 shows the first-stage coefficient of this instrument is 0.1342 ($p < 0.01$), with an F-statistic of 309.80. The second-stage coefficient of *Lnwords* in Column (6) is -0.2998 ($p < 0.01$), reinforcing the robustness of our main finding.

Across the three IV specifications, the coefficient of *Lnwords* remains significantly negative. The variation in magnitude—larger for IV1 (-0.5255) and smaller for IV2 (-0.0873) and IV3 (-0.2998)—reflects different local average treatment effects (LATE) for different complier populations, which is common in IV estimation. Importantly, the IV2 and IV3 estimates are closer to the baseline coefficient (-0.0613), suggesting that endogeneity does not drive our main result.

4.2.2. Heckman Two-Step Selection Model

A firm's decision to adopt AI may be non-random, influenced by factors such as financial conditions, managerial choices, and external competition, potentially leading to sample selection bias. Ignoring this self-selection process could bias OLS estimates. We therefore employ the Heckman two-step procedure to correct for this potential bias [50].

We first construct a treatment indicator *Treat* equal to 1 if a firm has positive *Lnwords* (i.e., mentions AI in its annual report) and 0 otherwise. In the first stage, we estimate a probit model of AI adoption using all control variables from the baseline regression plus

the industry-year mean instrument (IV1) as an exclusion restriction. This exclusion restriction affects ESG decoupling only through its impact on AI adoption, satisfying the identification condition. From this model, we compute the inverse Mills ratio (IMR), which captures selection bias arising from unobservable factors. In the second stage, we include the IMR as an additional regressor in the baseline model.

Column (1) of Table 5 shows that the coefficient of IMR is 1.0934, significant at the 5% level, confirming the presence of sample selection bias and justifying the use of the Heckman model. More importantly, after controlling for IMR, the coefficient of *Lnwords* remains -0.0616 ($p < 0.01$), which is highly consistent with the baseline result. This indicates that even after accounting for self-selection in AI adoption, the inhibitory effect of AI on ESG decoupling remains robust, ruling out sample selection bias as a driver of our main finding.

4.2.3. Lagged Independent Variables

Reverse causality is a common endogeneity concern in empirical analysis: it is possible that rather than AI adoption curbing ESG decoupling, firms with poor ESG performance are more likely to adopt AI to mask their behavior, or that the two interact bidirectionally. Moreover, the governance effect of AI may not be instantaneous; it could take time for AI adoption to translate into tangible improvements in ESG authenticity. To mitigate reverse causality concerns and capture dynamic effects, we replace the contemporaneous *Lnwords* with its one-period and two-period lags (L. *Lnwords* and L2. *Lnwords*) in the baseline model.

Columns (2) and (3) of Table 5 show that the coefficients of L. *Lnwords* and L2. *Lnwords* are -0.0747 and -0.0730 , respectively, both significant at the 1% level. This indicates that AI adoption in previous one to two years still exerts a significant inhibitory effect on current ESG decoupling, with magnitudes comparable to the baseline result. This confirms the persistence of AI's governance effect and further rules out reverse causality—because past AI adoption cannot be determined by future ESG decoupling.

4.3. Robustness Tests

To ensure the reliability of our main findings, we conduct a series of robustness checks following established practices in the literature. These tests include replacing the independent variable, adding province fixed effects, and excluding specific samples. Table 6 summarizes the results.

Table 6. Robustness checks.

Variable	(1) Decou	(2) GW	(3) Decou	(4) Decou	(5) Decou	(6) Decou
<i>AI_patent</i>	-0.0022 ** -0.0009					
<i>Lnwords</i>		-0.0517 *** -0.0181	-0.0596 *** -0.0142	-0.0584 *** -0.0148	-0.0620 *** -0.0164	-0.0606 *** -0.0158
Constant	-0.3076 -0.5253	-1.1961 * -0.682	-0.3673 -0.5228	-0.3235 -0.5657	-0.0497 -0.5652	-0.0269 -0.5695
Observations	14,018	9378	14,018	12,741	10,583	11,254
Adjusted R ²	0.4361	0.019	0.4426	0.4504	0.4496	0.4336
Controls	YES	YES	YES	YES	YES	YES
Province	NO	NO	YES	NO	NO	NO
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Note: Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Column (1) replaces the independent variable with *AI_patent* (natural

logarithm of one plus the number of AI patent applications). Column (2) uses an alternative measure of ESG decoupling following [10]. Column (3) adds province fixed effects. Column (4) excludes AI-related industries (computer, communication, and other electronic equipment manufacturing; instrumentation manufacturing). Column (5) excludes firms headquartered in the four municipalities (Beijing, Shanghai, Tianjin, Chongqing). Column (6) excludes the COVID-19 pandemic period (2020–2022). All continuous variables are winsorized at the 1st and 99th percentiles.

4.3.1. Alternative Measure of AI Adoption

One concern about the baseline specification is that the measure of AI adoption might be subject to measurement bias. We therefore employ the number of AI patent applications as an alternative proxy for a firm's AI adoption level.

Following Luo et al. (2026), Acemoglu et al. (2022), and Khan et al. (2024) [17,46,53], we construct *AI_patent* as the natural logarithm of one plus the number of AI patent applications by the firm in the given year. Theoretically, AI patents more directly capture firms' substantive technological innovation activities and tangible investments in AI [43]. Compared with keyword frequency extracted from annual reports—which may include symbolic statements or strategic disclosures—patent counts better reflect a firm's actual capability to embed AI technologies into production, operations, or governance processes [16,17]. Patents are widely used in the literature to measure corporate technological innovation, and their validity in the AI domain has been well established [30,53]. Column (1) of Table 6 shows that the coefficient of *AI_Patent* is -0.0022 , significant at the 5% level, consistent with the baseline conclusion.

4.3.2. Alternative Measure of ESG Decoupling

To further verify the robustness of our baseline findings, we adopt an alternative measure of ESG decoupling. In the baseline specification, we capture the “talk-versus-done” nature of decoupling from the firm's own perspective by taking the standardized difference between the Bloomberg ESG disclosure score (talk) and the Huazheng ESG performance rating (done). However, the relative position of a firm compared to its industry peers provides another important dimension for assessing decoupling behavior. Following Hu et al. (2023) [10], we construct a peer-relative decoupling measure (*GW*), defined as the deviation of a firm's ESG disclosure from the industry-year mean of ESG performance, scaled by the industry-year standard deviation. This measure captures whether a firm's ESG claims are excessive relative to its peers' actual performance, offering a complementary perspective on decoupling.

We re-estimate model (1) using *GW* as the dependent variable. Column (2) of Table 6 shows that the coefficient of *Lnwords* is -0.0517 , significant at the 1% level. The magnitude and significance are consistent with the baseline results, confirming that our main findings are not sensitive to the specific measurement of ESG decoupling.

4.3.3. Adding Province Fixed Effects

Differences across provinces in economic development, environmental regulation intensity, and local policies could lead to omitted variable bias. To alleviate this concern, we add province fixed effects to the baseline model, which already includes industry and year fixed effects.

A common approach to absorb time-invariant provincial characteristics—such as regional development policies, institutional quality, and environmental enforcement intensity—is to include province fixed effects [54]. Column (3) of Table 6 shows that after adding province fixed effects, the coefficient of *Lnwords* remains -0.0596 , significant at the 1% level, indicating that the baseline results are not confounded by omitted provincial-level variables.

4.3.4. Excluding Specific Samples

Particular subsets of the sample might unduly influence the estimated relationship. We therefore perform three exclusion tests.

First, excluding AI-related industries. Firms in high-tech industries such as computer, communication, and other electronic equipment manufacturing, as well as instrumentation manufacturing, are inherently more engaged with AI and typically possess stronger digital R&D capabilities, which could mechanically strengthen the observed AI-ESG relationship. After excluding these two industries, Column (4) of Table 6 shows that the coefficient of *Lnwords* is -0.0584 , still significant at the 1% level.

Second, excluding municipalities. Firms headquartered in the four municipalities directly under the central government (Beijing, Shanghai, Tianjin, and Chongqing) may benefit from unique policy privileges, superior infrastructure, and greater knowledge spillovers. Column (5) of Table 6 shows that the coefficient of *Lnwords* is -0.0620 ($p < 0.01$), consistent with the baseline result.

Third, excluding the COVID-19 pandemic period (2020–2022). The pandemic caused unprecedented disruptions to economic activities, potentially biasing the estimated relationship. We drop observations from 2020 to 2022. Column (6) of Table 6 shows that the coefficient of *Lnwords* is -0.0606 ($p < 0.01$), further confirming the robustness of our baseline findings.

Taken together, across all specifications—alternative independent variable, additional province fixed effects, and various sample exclusions—the coefficient of the core independent variable remains negative and statistically significant, indicating that our findings are robust.

5. Mechanism Test

Having established that AI adoption significantly curbs ESG decoupling (Table 3), we now turn to examining the underlying mechanisms. Specifically, we test the three mediating pathways derived from the fraud triangle theory: opportunity (internal control quality), pressure (media attention), and rationalization (disclosure quality).

The three mediators correspond to the three independent conditions specified by the fraud triangle theory: opportunity, pressure, and rationalization [8]. By theoretical construction, these conditions are conceptually distinct and mutually irreducible—none is assumed to be a prerequisite for another. This theoretical independence provides the primary justification for treating them as parallel mediators [55,56]. To further assess whether this theoretical distinction is empirically supported, we conduct two diagnostic checks: pairwise correlation analysis and variance inflation factor (VIF) estimation.

As shown in Appendix A Table A1, the pairwise Pearson correlation coefficients among environmental uncertainty, media attention, and disclosure quality are all below 0.30, indicating that no strong linear association exists between any pair of mediators. In addition, we calculated the VIFs in a regression including all three mediators simultaneously, along with the full set of control variables, all well below the conventional threshold of 5. These diagnostics suggest that multicollinearity is unlikely to be a serious concern, and that each mediator captures a sufficiently distinct mechanism to be examined as a parallel pathway.

To test these hypotheses (H2–H4), we follow the causal mediation approach widely adopted in the recent literature [9] and directly identify the causal effect of AI adoption on the proposed mediators. The following model is specified:

$$M_{i,t} = \beta_0 + \beta_1 Lnwords_{i,t} + \beta_2 Control_{i,t} + \sigma_i + \delta_t + \varepsilon_{i,t} \quad (2)$$

where $M_{i,t}$ represents each mediating variable (environmental uncertainty, media attention, and disclosure quality). All other variables are defined identically to those in

model (1) Table 7 reports the regression results. The effects of these mediators on ESG decoupling ($M \rightarrow Y$) have been established in the theoretical analysis and existing literature, and thus are not re-estimated here.

Table 7. Mechanism test results: Testing the Fraud Triangle framework.

Variable	Opportunity	Pressure		Rationalization
	(1) Environmental Uncertainty	(2) Media Attention	(3) Analyst Attention	(4) Disclosure Quality
<i>Lnwords</i>	−0.0407 **	0.0974 ***	0.3603 **	0.0149 ***
	−0.017	−0.0207	−0.165	−0.0042
Observations	12,779	13,983	14,018	10,698
Adjusted R ²	0.2417	0.3886	0.4243	0.2129
Controls	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Note: Robust standard errors are in parentheses. *** and ** denote significance at the 1% and 5% levels, respectively. Environmental uncertainty (opportunity dimension) is measured as industry-adjusted sales volatility. Media attention (pressure dimension) is the natural logarithm of one plus the number of media reports. Analyst attention (pressure dimension) is defined as the natural logarithm of one plus the number of analysts following the firm in that year. Disclosure quality (rationalization dimension) is the inverse of the CSMAR disclosure rating. All continuous variables are winsorized at the 1st and 99th percentiles.

5.1. Compressing the Opportunity for ESG Decoupling

AI reduces environmental uncertainty and ambiguity in markets, technology, and regulation by processing massive heterogeneous data in real time, generating high-precision dynamic forecasts, and continuously tracking external risk signals. This compression of uncertainty narrows the opportunity window for managers to engage in selective disclosure or symbolic compliance through information asymmetry, thereby strengthening the constraining effectiveness of internal control and ultimately lowering the risk of internal control failure due to external shocks. Specifically, environmental uncertainty refers to the volatility and unpredictability of a firm's external environment, including market conditions, technological changes, and regulatory shifts. High environmental uncertainty can create “opportunity” for ESG decoupling by blurring information, increasing the difficulty of external monitoring, and expanding managerial discretion [57]. In such an opaque environment, firms are more likely to engage in selective disclosure or exaggeration without being detected. Conversely, if AI helps reduce environmental uncertainty, it should compress the opportunity for decoupling. Indeed, AI can process massive amounts of data to generate more accurate forecasts and real-time risk assessments, thereby reducing uncertainty [58].

Column (1) of Table 7 presents the regression result of AI adoption on environmental uncertainty. The coefficient of *Lnwords* is −0.0407, significant at the 5% level. This indicates that AI adoption significantly reduces the level of environmental uncertainty faced by firms. Combined with the evidence in the literature that high uncertainty provides greater leeway for symbolic compliance and decoupling [59], this finding supports the mediating role of environmental uncertainty as an opportunity-diminishing channel.

5.2. Increasing the Pressure on ESG Decoupling

The adoption of AI standardizes and verifies the process of information generation and disclosure within firms, which enables external media to access, mine, and cross-val-

idate corporate information more conveniently, thereby significantly increasing the intensity of media attention on firms. This intensified external pressure ultimately helps to curb ESG decoupling. External oversight is a key governance mechanism that constrains corporate behavior. Media, as an information intermediary, exerts pressure on firms through information discovery and reputational sanctions [37]. Greater media attention raises the expected costs of greenwashing, thereby increasing the “pressure” dimension of the fraud triangle. AI can enhance media monitoring by enabling real-time sentiment analysis, automated news aggregation, and anomaly detection [28]. Therefore, if AI curbs ESG decoupling by intensifying external pressure, we should observe a positive effect of AI adoption on media attention.

To test the “pressure-enhancing” effect of media attention, we follow Gao and Chen [33] and measure it as the natural logarithm of one plus the number of media reports. Column (2) of Table 7 reports the regression result of AI adoption on media attention. The coefficient of *Lnwords* is 0.0974, significant at the 1% level. This indicates that AI significantly amplifies media scrutiny. A one-standard-deviation increase in *Lnwords* (1.1150) increases media attention by approximately 0.1086, or about 8.7% of its standard deviation. Combined with the well-documented evidence that media attention effectively deters greenwashing [36], this result supports the mediating role of media attention as a pressure-enhancing channel.

To further test the robustness of the pressure dimension, we employ analyst coverage as an alternative proxy for external oversight. Unlike media attention, analysts convey professional and systematic firm-specific information through earnings forecasts, research reports, and investment ratings, with their monitoring mechanisms being more sustained and targeted [33]. Theoretically, AI adoption can attract greater analyst following by enhancing the standardization and verifiability of corporate disclosures, thereby reducing analysts’ information processing costs. We measure analyst coverage (*Analyst*) as the natural logarithm of one plus the number of analysts following the firm, and re-estimate model (2). Column (3) of Table 7 shows that the coefficient of *Lnwords* is 0.3603 ($p < 0.05$), indicating that AI significantly increases the intensity of analyst coverage. Combined with the well-documented evidence that analyst coverage effectively curbs greenwashing [35,36], this result supports the mediating role of the pressure channel from an additional perspective, further reinforcing the robustness of the pressure-dimension findings.

5.3. Undermining the Rationalization of ESG Decoupling

AI improves the transparency, verifiability, and timeliness of ESG disclosure through specific mechanisms including automated textual validation, multi-source data cross-verification, and standardized reporting procedures. This high-quality disclosure environment prevents managers from using vague language or selective statements to convince themselves that their decoupling behaviors will remain undetected, thereby effectively undermining the rationalization space in the fraud triangle and ultimately curbing ESG decoupling. Disclosure quality directly affects whether managers believe that greenwashing will go undetected. When disclosure is transparent, verifiable, and timely, managers find it difficult to mask true behavior through vague language or selective disclosure, thereby compressing the psychological space for “rationalization” [39]. AI can improve disclosure quality by automating natural language processing to detect inconsistencies, cross-validating multi-source data, and standardizing reporting procedures [40,41]. Thus, if AI curbs ESG decoupling by undermining rationalization, we should observe a positive effect of AI adoption on disclosure quality.

To test the “rationalization-undermining” effect of disclosure quality, we follow Zhan et al. [39] and measure it as the inverse of the CSMAR disclosure rating. Column (4)

of Table 7 reports the regression result of AI adoption on disclosure quality. The coefficient of *Lnwords* is 0.0149, significant at the 1% level. This indicates that AI significantly improves the quality of ESG disclosure. Combined with the evidence that high-quality disclosure curbs ESG decoupling [37], this finding supports the mediating role of disclosure quality as a rationalization-undermining channel.

6. Heterogeneity Analysis

The technology effect is contingent upon organizational capabilities and environmental conditions [60]. On this basis, the inhibitory effect of AI on ESG decoupling may vary depending on firm characteristics, industry attributes, and regional institutional environments. In this section, we examine three levels of heterogeneity: firm-level (managerial characteristics and firm attributes), industry-level (competition and institutional environment), and regional-level (factor markets and environmental regulations). Table 8, Table 9 and Table 10 present the results for firm-level, industry-level, and regional-level heterogeneity, respectively.

Table 8. Heterogeneity analysis: Firm-level factors.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Executive R&D Background Yes	No	Corporate Reputation High	Low	Corporate Governance Quality High	Low
<i>Lnwords</i>	−0.0637 *** −0.0148	−0.0269 −0.0462	−0.0672 *** −0.0182	−0.0570 *** −0.0192	−0.0634 *** −0.0193	−0.0581 *** −0.0187
Observations	12,162	1445	7217	6800	7032	6985
Adjusted R ²	0.422	0.5395	0.4131	0.4765	0.4248	0.4592
Controls	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Coefficient diff. <i>p</i> -value	0.0000		0.0000		0.0000	

Note: Robust standard errors clustered at the firm level are in parentheses. *** indicates significance at the 1% level. Corporate reputation is based on media rankings and social media sentiment scores. Corporate governance quality is constructed from board independence, audit committee effectiveness, and ESG-related governance mechanisms. All continuous variables are winsorized at the 1st and 99th percentiles.

Table 9. Heterogeneity analysis: Industry-level factors.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Industry Competition High	Low	Digital Economy Industries Yes	No	Technologically Fast-Iterating Industries Yes	No
<i>Lnwords</i>	−0.0702 *** −0.0173	−0.0454 ** −0.0181	−0.0547 *** −0.0181	−0.0537 ** −0.0241	−0.0631 * −0.0341	−0.0541 *** −0.0152
Observations	7037	6980	6784	7160	2024	11,991
Adjusted R ²	0.4479	0.4285	0.4056	0.4644	0.3495	0.4538
Controls	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Coefficient diff. <i>p</i> -value	0.0000		0.0000		0.0000	

Note: Robust standard errors clustered at the firm level are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Digital economy industries are classified following the National Bureau of Statistics of China. Technologically fast-iterating industries are identified based on R&D intensity and the pace of technological change.

Table 10. Heterogeneity analysis: Regional-level factors.

Variable	(1)	(2)	(3)	(4)
	Industry Competition		Digital Economy Industries	
	High	Low	Yes	No
<i>Lnwords</i>	−0.0650 *** (0.0187)	−0.0587 *** (0.0165)	−0.0699 *** (0.0202)	−0.0464 *** (0.0175)
Observations	6174	7842	7200	6815
Adjusted R ²	0.4399	0.4340	0.4649	0.4176
Controls	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Coefficient diff. <i>p</i> -value	0.0000		0.0000	

Note: Robust standard errors are in parentheses. *** indicates significance at the 1% level. *p*-values for coefficient difference tests: 0.0000 (R&D vs. non-R&D), 0.0000 (high vs. low competition). All continuous variables are winsorized at the 1st and 99th percentiles.

6.1. Firm-Level Heterogeneity

6.1.1. Executive R&D Background

Executives with research and development (R&D) backgrounds possess stronger technical expertise and longer decision horizons, enabling them to promote the adoption of AI-driven governance tools at the organizational level [61,62]. Such executives are more sensitive to the potential of AI applications in environmental governance and are more inclined to leverage AI as a strategic tool for substantive ESG improvement rather than as a means of symbolic compliance. We construct a dummy variable based on executive resumes in the CSMAR database, equal to 1 if the firm has at least one executive with an R&D background, and 0 otherwise. The sample is split into two subgroups, and model (1) is re-estimated for each.

Columns (1) and (2) of Table 8 show that the coefficient of *Lnwords* is −0.0637 ($p < 0.01$) in firms with R&D executives, whereas it is −0.0269 and statistically insignificant in those without. The coefficient difference is highly significant ($p = 0.0000$). This confirms that AI's inhibitory effect is concentrated in firms with R&D-experienced leadership, consistent with the view that technical expertise facilitates the adoption and effectiveness of digital governance tools.

6.1.2. Corporate Reputation

Firms with higher reputation face greater reputational risk and stakeholder scrutiny, which may amplify the deterrent effect of AI-driven monitoring [34]. Once high-reputation firms are exposed for greenwashing, their reputational capital would suffer greater losses, making them more responsive to AI-based monitoring. In contrast, low-reputation firms have less to lose and may respond less to AI-driven scrutiny. We measure corporate reputation using rankings from annual surveys conducted by leading Chinese business media outlets, supplemented by social media sentiment scores. Firms are split at the sample median into high-reputation and low-reputation groups.

Columns (3) and (4) of Table 8 show that the coefficient of *Lnwords* is -0.0672 ($p < 0.01$) in high-reputation firms and -0.0570 ($p < 0.01$) in low-reputation firms, with a significant difference ($p = 0.0000$). This suggests that AI's inhibitory effect is more pronounced in firms with higher reputational capital, consistent with the theoretical expectation that reputational risk amplifies the deterrence of external oversight.

6.1.3. Corporate Governance Quality

Firms with stronger corporate governance structures are better positioned to implement AI-driven governance tools effectively [28]. Good governance provides the institutional foundation—transparent decision-making processes, effective monitoring mechanisms, and accountability systems—that enables AI to function as a credible governance device. Firms with weaker governance quality may face internal implementation barriers that undermine AI's governance effectiveness. We construct a governance quality index based on board independence, audit committee effectiveness, and the completeness of ESG-related governance mechanisms. Firms are split at the sample median into high-governance and low-governance groups.

Columns (5) and (6) of Table 8 show that the coefficient of *Lnwords* is -0.0634 ($p < 0.01$) in high-governance firms and -0.0581 ($p < 0.01$) in low-governance firms, with a significant difference ($p = 0.0000$). This confirms that AI's governance effect is reinforced by stronger internal governance quality, indicating that a sound institutional environment is an important prerequisite for AI to exert its governance function.

A general concern is that firms with certain favorable characteristics—such as R&D-experienced executives, high reputation, or strong governance—may already possess higher ESG standards irrespective of AI adoption. To systematically address this concern, we conduct coefficient difference tests for each dimension. The results show that the coefficient difference tests are statistically significant across all dimensions, suggesting that the observed heterogeneity reflects genuine moderating effects of firm characteristics rather than selection bias.

6.2. Industry-Level Heterogeneity

6.2.1. Industry Competition

Firms in highly competitive industries face greater pressure to differentiate themselves and maintain market standing. Under intense competition, firms are more likely to adopt substantive environmental improvements to gain legitimacy and competitive advantage, and AI can assist in such efforts [16,17,63]. Moreover, competitive pressure may prompt firms to respond more actively to AI-driven monitoring. We measure industry competition using the Herfindahl-Hirschman Index (HHI) based on firm sales within each industry-year. Firms are assigned to the high-competition group if the industry HHI is below the sample median (i.e., less concentrated), and to the low-competition group otherwise.

Columns (1) and (2) of Table 9 show that the coefficient of *Lnwords* is -0.0702 ($p < 0.01$) in high-competition industries and -0.0454 ($p < 0.05$) in low-competition industries. The difference is statistically significant ($p = 0.0000$), indicating that AI's governance role is more pronounced when firms face stronger market pressure.

6.2.2. Digital Economy Industries

Firms in digital economy industries possess superior data infrastructure, technical capabilities, and digital awareness, which may facilitate the adoption and effectiveness of AI-driven governance tools [16,17]. Moreover, firms in these industries are more accustomed to data-driven decision-making and have stronger internalization capacity for AI

technologies, potentially amplifying AI's governance effects on ESG decoupling. We classify industries as digital economy industries following the official classification by the National Bureau of Statistics of China, which encompasses computer, communication, and other electronic equipment manufacturing, as well as software and information technology services.

Columns (3) and (4) of Table 9 show that the coefficient of *Lnwords* is -0.0547 ($p < 0.01$) in digital economy industries and -0.0537 ($p < 0.05$) in non-digital economy industries. The coefficient difference is statistically significant ($p = 0.0000$), indicating that AI's inhibitory effect is more pronounced in digital economy industries, consistent with the view that digital readiness amplifies AI's governance potential.

6.2.3. Technologically Fast-Iterating Industries

Industries characterized by rapid technological iteration face shorter product life cycles and greater pressure to innovate. In such environments, firms are more likely to proactively adopt cutting-edge technologies for substantive ESG improvements to maintain competitive advantage. Meanwhile, rapid technological change also implies that firms' ESG disclosures are subject to more intense scrutiny from stakeholders, thereby strengthening the deterrent effect of AI-driven monitoring. Based on the techno-economic characteristics of different industries, we identify technologically fast-iterating industries based on R&D intensity and the pace of technological change.

Columns (5) and (6) of Table 9 show that the coefficient of *Lnwords* is -0.0631 ($p < 0.1$) in fast-iterating industries and -0.0541 ($p < 0.01$) in non-fast-iterating industries, with a significant difference ($p = 0.0000$). Although the significance level in fast-iterating industries is marginal, the coefficient magnitude is larger, and the difference test confirms a meaningful gap between the two groups.

6.3. Regional-Level Heterogeneity

6.3.1. Factor Market Development

Well-developed factor markets facilitate the efficient allocation of resources, including capital, labor, and technology. In regions with higher factor market development, firms are better able to access the financial and human resources needed to adopt and leverage AI for environmental governance [15,38]. Conversely, in less developed factor markets, resource constraints may hinder AI adoption or limit its effectiveness. We therefore expect AI's inhibitory effect to be larger in regions with higher factor market development.

Following the mainstream literature, we use the factor market sub-index of the Fan Gang Marketization Index to measure regional factor market development. Cities with an index above the sample median are classified as high-development regions, and those below as low-development regions. We then re-estimate model (1) for each group.

Columns (1) and (2) of Table 10 report the results. The coefficient of *Lnwords* is -0.0650 ($p < 0.01$) in high-development regions and -0.0587 ($p < 0.01$) in low-development regions. The difference is marginally significant ($p = 0.0654$), suggesting that AI's effect is somewhat stronger in regions with more developed factor markets, although the gap is modest. This is plausible because even in less developed markets, AI may still exert a significant effect, but better factor market conditions slightly enhance its efficiency.

6.3.2. Regional Environmental Regulation Intensity

Stricter environmental regulations increase the cost of greenwashing and enhance the credibility of environmental enforcement. In regions with stronger environmental regulation, firms face higher reputational and financial risks from being caught in ESG decoupling, which may amplify the deterrent effect of AI-driven monitoring [13]. Moreover,

AI technologies are often used by regulators in high-regulation areas to screen corporate environmental reports, creating a synergistic effect. Hence, we expect AI's inhibitory effect to be larger in regions with stronger environmental regulation intensity.

We measure regional environmental regulation intensity using the number of environmental administrative penalties per capita or the proportion of environmental protection expenditure in GDP, as collected from the China Environmental Statistics Yearbook. Cities are split at the sample median into high-regulation and low-regulation groups.

Columns (3) and (4) of Table 10 show that the coefficient of *Lnwords* is -0.0699 ($p < 0.01$) in high-regulation regions and -0.0464 ($p < 0.01$) in low-regulation regions. The difference is highly significant ($p = 0.0000$). This confirms that stricter environmental regulation reinforces AI's ability to curb ESG decoupling, consistent with the idea that regulatory pressure and technological monitoring work in tandem.

7. Discussion

7.1. Theoretical Framework and Contributions

Our findings demonstrate that the fraud triangle theory provides a coherent analytical framework for understanding AI's governance role in ESG decoupling. The three dimensions, opportunity, pressure, and rationalization, operate as distinct yet complementary channels through which AI exerts its inhibitory effect. This framework moves beyond single-mechanism approaches by capturing the multidimensional nature of corporate fraud, where all three conditions must converge for decoupling to persist. By showing that AI simultaneously compresses opportunity (via reducing environmental uncertainty), intensifies pressure (via enhancing media attention and analyst coverage), and undermines rationalization (via improving disclosure quality), our study provides empirical support for the theoretical proposition that effective governance requires addressing all three fraud triangle dimensions. This finding extends the applicability of the fraud triangle theory beyond its traditional domain of financial fraud to the emerging context of ESG and digital transformation.

A distinctive feature of our framework is the explicit distinction between the three fraud triangle dimensions and their empirical proxies. Unlike prior studies that have often conflated theoretical constructs with their measurements, we consistently treat environmental uncertainty, media attention, and disclosure quality as observable indicators of opportunity, pressure, and rationalization, respectively, rather than equating them with the dimensions themselves. This conceptual clarity strengthens the theoretical contribution and provides a replicable template for future research applying the fraud triangle theory to diverse governance contexts.

This study makes three theoretical contributions. First, it advances the AI governance literature by shifting the analytical focus from the direct question of whether AI curbs ESG decoupling to the more nuanced question of how AI exerts its governance effects through multiple distinct channels. While prior studies have documented the inhibitory effect of AI on ESG decoupling, they have largely treated AI as a monolithic technology without unpacking the mechanisms through which it operates. Our identification of three parallel mediating pathways provides a more granular understanding of AI's governance role and highlights that different mechanisms may be relevant under different conditions.

Second, it extends the fraud triangle theory to the ESG context, demonstrating that the three fraud conditions identified by Albrecht are equally applicable to understanding environmental information manipulation [8]. Prior applications of the theory to ESG contexts have primarily focused on identifying the presence of these conditions [18,21]. Our study goes further by showing how a technology, AI, can simultaneously intervene in all three conditions, thereby offering a proactive governance approach rather than merely

a diagnostic one. This extension enriches both the fraud triangle theory and the ESG literature by bridging these previously separate domains.

Third, it contributes to the digital governance literature by showing that AI's governance effects are not automatic; rather, they depend on organizational capabilities and institutional conditions. The heterogeneous effects we observe, which are stronger in firms with R&D-experienced executives, higher reputation, stronger governance quality, and in more competitive and digitally advanced industries, suggest that AI is not a technological panacea. This finding cautions against technological determinism and highlights the importance of complementary investments in governance infrastructure, echoing the dynamic capabilities perspective articulated by Han et al. (2026) [6]. In other words, AI becomes effective when firms possess the absorptive capacity to integrate AI-driven insights into decision-making processes.

7.2. Comparison with Existing Studies

Our findings both align with and extend previous research. Consistent with recent studies documenting AI's inhibitory effect on ESG decoupling [16,27], we confirm the direct negative relationship. However, our mediation analysis reveals that this effect operates through three distinct channels, which prior studies have not jointly examined. This multidimensional perspective offers a more complete account of AI's governance mechanisms than existing single-pathway analyses.

In relation to studies applying the fraud triangle theory to greenwashing [18,21], we extend the framework by showing that AI can simultaneously intervene in all three conditions rather than merely identifying their presence. Our findings also complement studies examining the role of external monitors in disciplining corporate behavior. Chen et al. (2026) demonstrate that institutional investors' site visits can curb greenwashing by providing external monitors with better information about firms [32]. Our study extends this logic by showing that AI serves as a technology that enhances the information environment for multiple types of monitors, including analysts, media, and regulators, thereby amplifying the disciplinary effect of external oversight. While Chen et al. (2026) [32] focus on a single external monitor (institutional investors), we show that AI enhances the information environment for multiple monitors simultaneously.

Our heterogeneity findings corroborate the Technology-Organization-Environment (TOE) framework by demonstrating that AI's effectiveness is contingent on organizational capabilities and environmental conditions [6]. The stronger effects observed in firms with R&D-experienced executives and strong governance quality align with the TOE emphasis on organizational readiness. Similarly, the more pronounced effects in highly competitive and digitally advanced industries are consistent with the TOE focus on environmental factors that shape technology adoption outcomes. This pattern, that technology effects are conditional on complementary assets, is well established in the innovation literature but is novel to the ESG literature, where AI has often been treated as a uniformly effective tool.

7.3. Boundary Conditions of AI's Governance Effects

Our heterogeneity analysis reveals that AI's governance effects are bounded by both internal and external conditions, providing a more nuanced understanding of where and for whom AI is most effective.

At the firm level, the effect is stronger in firms with R&D-experienced executives (Table 8, Columns 1–2), higher reputation (Table 8, Columns 3–4), and stronger governance quality (Table 8, Columns 5–6). This suggests that organizational absorptive capacity, the ability to recognize, assimilate, and apply new information, is critical for translating AI

adoption into governance outcomes. Firms lacking these capabilities may adopt AI symbolically rather than substantively, mirroring the very decoupling behavior that AI is intended to curb. This insight has important implications for practitioners: investments in AI should be accompanied by investments in human capital and governance infrastructure.

At the industry level, the effect is more pronounced in highly competitive (Table 9, Columns 1–2) and digitally advanced industries (Table 9, Columns 3–4). In competitive markets, firms face greater pressure to differentiate themselves, creating stronger incentives to use AI for genuine ESG improvement rather than symbolic compliance. In digital economy industries, firms possess superior data infrastructure and technical capabilities that facilitate AI adoption and effectiveness. However, the effect is significantly weakened in administratively monopolized sectors (Table 9, Columns 5–6), where firms face less competitive pressure and weaker incentives to respond to AI-driven governance. This suggests that market competition and technological readiness amplify AI's governance potential, while administrative protection undermines it.

At the regional level, stricter environmental regulations (Table 10, Columns 3–4) and more developed factor markets (Table 10, Columns 1–2) strengthen the effect. Stringent regulations increase the cost of greenwashing and the credibility of enforcement, creating a synergistic effect with AI-driven monitoring. Well-developed factor markets provide firms with better access to the financial and human resources needed to adopt and leverage AI for environmental governance. These findings highlight the role of institutional context in shaping AI's governance effectiveness.

Together, these boundary conditions suggest that AI is not a universal governance solution; its effectiveness depends on the enabling environment in which it is deployed. This insight has important implications for policymakers seeking to leverage AI for ESG: investment in AI should be accompanied by investments in complementary organizational capabilities, market competition, regulatory enforcement, and institutional infrastructure. Without such complementary conditions, AI adoption may become another form of symbolic compliance, precisely the problem it is meant to solve.

8. Conclusions and Policy Implications

8.1. Conclusions

This study investigates whether and how AI adoption affects corporate ESG decoupling under China's dual-carbon policy. Using a sample of Chinese A-share listed firms from 2009 to 2024 and a two-step causal mediation approach, we obtain several key findings.

First, AI adoption significantly curbs ESG decoupling. This result remains robust after a series of endogeneity tests, including instrumental variable estimation, the Heckman two-step selection model, and lagged regressions. Economically, a one-standard-deviation increase in AI adoption reduces ESG decoupling by about 5.1% of its standard deviation.

Second, drawing on the fraud triangle theory, we identify three distinct mediating pathways. AI compresses the opportunity for decoupling by reducing environmental uncertainty (coefficient -0.0407 , $p < 0.05$); it increases the pressure for decoupling by enhancing media attention (coefficient 0.0974 , $p < 0.001$) and analyst attention (coefficient 0.3063 , $p < 0.05$); and it undermines the rationalization of decoupling by improving disclosure quality (coefficient 0.0149 , $p < 0.001$). These results confirm that AI governs ESG decoupling through a multi-dimensional mechanism corresponding to the three elements of the fraud triangle.

Third, the inhibitory effect of AI varies systematically across firm characteristics, industry attributes, and regional institutional environments. At the firm level, the effect is

stronger in firms with R&D-experienced executives, higher reputation, and stronger governance quality. At the industry level, the effect is more pronounced in highly competitive industries, digital economy industries, and technologically fast-iterating industries. At the regional level, it is stronger in regions with higher factor market development and stricter environmental regulations. These findings suggest that internal governance capacity, market competition, and external institutional conditions jointly shape the boundary conditions of AI's governance effects.

Overall, this study provides a unified theoretical framework—the fraud triangle theory—to explain the mediating mechanisms, and offers granular evidence on the boundary conditions of AI's governance role, contributing to the literature on digital governance and corporate sustainability.

8.2. Policy Implications

Our findings carry several policy implications for leveraging digital technologies to enhance ESG authenticity under China's dual-carbon policy, and more broadly, the three identified mediating pathways reflect general properties of AI as a governance technology, with applicability extending beyond China to setting where ESG decoupling threatens policy credibility, thus providing a referential policy instrument for global environmental governance.

First, accelerate AI adoption in corporate environmental governance. Given the significant inhibitory effect of AI on ESG decoupling, regulators should encourage firms to integrate AI-driven monitoring systems into their internal control and ESG reporting processes. For example, listing requirements could include provisions for AI-based data verification, and tax incentives or green credit preferences could be offered to firms that adopt AI for environmental management.

Second, adopt differentiated strategies for AI governance promotion. Heterogeneity analyses show that AI's effect is stronger in firms with R&D-experienced executives, higher reputation, and stronger governance quality. For firms with weaker executive technical capabilities or less developed governance structures, regulators should provide targeted capacity-building programs and technical support to enhance their ability to adopt and implement AI governance tools.

Third, prioritize AI governance tools in highly competitive industries and digital economy industries. Industry-level heterogeneity reveals that AI's inhibitory effect is more pronounced in highly competitive industries, digital economy industries, and technologically fast-iterating industries. Policymakers should promote pilot programs and the diffusion of AI governance tools in these industries first, and then scale up successful experiences to other sectors.

Fourth, strengthen regional institutional environments to create enabling conditions for AI governance. Regional heterogeneity indicates that factor market development and environmental regulation intensity significantly shape AI's governance effectiveness. For regions with underdeveloped factor markets or lax environmental regulations, stronger support for data infrastructure and regulatory enforcement is needed to provide institutional guarantees for AI's governance function.

8.3. Limitations and Future Research

Despite its contributions, this study has several limitations that point to promising avenues for future research.

First, measurement of AI adoption. Our primary measure of AI adoption relies on AI-related keyword frequency in annual reports (*Lnwords*). While widely used, this approach may capture firms' rhetorical emphasis on AI rather than actual implementation. To mitigate this concern, we employed AI patent applications (*AI_patent*) as an alternative

measure. The consistent results across these measures suggest that our main conclusions are not merely an artifact of measurement choices. Future research could further validate our findings using more granular data, such as AI-related job postings or survey-based adoption indicators.

Second, generalizability. Our empirical analysis is based on Chinese A-share listed firms, which may limit generalizability to other institutional settings. However, the theoretical logic we advance—that AI curbs ESG decoupling by compressing opportunity, increasing pressure, and undermining rationalization—is not context-specific, as these mechanisms operate through fundamental governance channels that exist across institutional settings. Future research could extend the analysis to non-listed firms, other emerging economies, or cross-country settings.

Third, endogeneity. Although we employed multiple identification strategies—including instrumental variable estimation, the Heckman two-step procedure, and lagged regressions—no observational study can definitively rule out all forms of endogeneity. Future research could exploit quasi-natural experiments, such as policy shocks or regulatory changes, to provide additional causal evidence.

Fourth, potential interactions among mediators. While we treat the three mediators as parallel pathways based on theoretical reasoning [8] and empirical diagnostics (low pairwise correlations and VIFs), we cannot completely rule out all possible interactions among them. For instance, improved disclosure quality could attract greater media attention, creating a reinforcing loop between the rationalization and pressure channels. However, the low correlations and VIFs suggest that such interactions, if present, are unlikely to compromise the parallel-pathway interpretation. Future research could employ structural equation modeling or serial mediation analysis to further explore these potential interdependencies.

Fifth, measurement of ESG decoupling. Our measure, based on the standardized difference between ESG disclosure and ESG performance, provides a reasonable proxy for greenwashing. However, a negative gap may reflect under-disclosure rather than the absence of decoupling. We interpret the magnitude of the gap rather than its sign alone, but future research could develop more refined measures to better distinguish between strategic under-disclosure and genuine alignment.

Despite these limitations, the consistency of our findings across multiple identification strategies, alternative variable specifications, and heterogeneity dimensions lends confidence to our main conclusions.

Author Contributions: J.F.: Conceptualization, Methodology, Software, Validation, Investigation, Visualization, Funding acquisition. W.G.: Methodology, Validation, Formal analysis, Data curation, Writing—original draft. B.Z.: Theoretical framework construction, Mechanism analysis, Visualization optimization. J.Y.: Reference sorting, Format typesetting, Result double-checking, Validation, Resources, Writing—review and editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China (Western Project), managed by the National Office for Philosophy and Social Sciences (Grant No. 23XGL001). The project title is “Configurational Effects and Targeted Regulation of Greenwashing in High-Carbon Enterprises”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the corresponding author upon reasonable request.

Acknowledgments: We are grateful for the support from North Minzu University for providing us with the necessary research facilities and academic resources.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Collinearity Diagnostics for Mediating Variables of the Fraud Triangle.

Variable Name	VIF Value	Pearson Correlation Coefficient (with Highly Correlated Variables)	Collinearity Diagnostic Criteria	Judgment of Results
<i>environmental uncertainty</i>	1.25	Correlation with Media_6, $r = -0.0370^*$	$VIF < 5$, $ r < 0.3$, No collinearity	No collinearity
<i>media attention</i>	1.30	Correlation with disclosure quality, $r = -0.1941^*$	$VIF < 5$, $ r < 0.3$, No collinearity	No collinearity
<i>disclosure quality</i>	1.31	Correlation with environmental uncertainty, $r = -0.1653^*$	$VIF < 5$, $ r < 0.3$, No collinearity	No collinearity

Note: * denote statistical significance at the 10% level.

References

- Christensen, D.M.; Serafeim, G.; Sikochi, A. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *Account. Rev.* **2022**, *97*, 147–175.
- Tashman, P.; Marano, V.; Kostova, T. Walking the walk or talking the talk? Corporate social responsibility decoupling in emerging market multinationals. *J. Int. Bus. Stud.* **2019**, *50*, 153–171.
- Dhaliwal, D.S.; Li, O.Z.; Tsang, A.; Yang, Y.G. Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *Account. Rev.* **2011**, *86*, 59–100.
- Marquis, C.; Qian, C. Corporate social responsibility reporting in China: Symbol or substance? *Organ. Sci.* **2014**, *25*, 127–148.
- Zhou, C. Toward human-AI collaboration for sustainable development: unveiling the drives of artificial intelligence adoption in environmental governance. *J. Sustain. Dev.* **2026**. <https://doi.org/10.1002/sd.71277>
- Han, X.; Wu, Y.; Li, X. Artificial intelligence adoption, dynamic capabilities, and firm risk-taking. *J. Int. Financ. Manag. Account.* **2026**, *37*, 477–511.
- Bergougui, B.; Sulimany, H. G. H.; Alharbi, A. A. Combating ESG greenwashing through AI models: Evidence from disaggregated AI technologies, mechanisms, and thresholds. *Corp. Soc. Responsib. Environ. Manag.* **2026**. <https://doi.org/10.1002/csr.70689>
- Albrecht, W.S.; Wernz, G.W.; Williams, T.L. *Fraud: Bringing Light to the Dark Side of Business*; Irwin Professional Publishing: Burr Ridge, IL, 1995. pp. 1–15.
- Imai, K.; Keele, L.; Yamamoto, T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* **2010**, *25*, 51–71.
- Hu, S.; Wang, M.; Wu, M.X.; Wang, A.L. Voluntary environmental regulations, greenwashing and green innovation: Empirical study of China's ISO14001 certification. *Environ. Impact Assess. Rev.* **2023**, *102*, 107224.
- Tang, Y.H.; Miao, X.; Zang, H.Y.; Gao, Y.H. Information disclosure on hazards from industrial water pollution incidents: Latent resistance and countermeasures in China. *Sustainability* **2018**, *10*, 1475.
- Cheng, Z.; Yang, G. The impact of artificial intelligence on corporate ESG greenwashing. *Socio-Econ. Plan. Sci.* **2026**, *103*, 101832.
- Zhao, Y.N.; Lee, C.C. The impact of vertical environmental regulation mechanism on greenwashing. *J. Clean. Prod.* **2024**, *475*, 143637.
- Jiang, C.X.; Li, X.X.; Xu, Q.F.; Liu, J.H. Does environmental protection tax impact corporate ESG greenwashing? *Econ. Anal. Policy.* **2024**, *84*, 774–786.
- Liu, X.; Yang, Q.; Wei, K.; Dai, P.-F. ESG Rating disagreement and idiosyncratic return volatility: Evidence from China. *Res. Int. Bus. Financ.* **2024**, *70*, 102368.
- Li, D.; Zhang, Z.; Gao, X. Does artificial intelligence deter greenwashing? *Finance Res. Lett.* **2024**, *67*, 105954.
- Luo, T.; Li, Z. How environmental centralization curb corporate greenwashing behavior: Evidence from China. *Bus. Strateg. Environ.* **2026**, *35*, 2411–2428.

18. Ren, G.Q.; Jing, M.; Wang, Z.M. True green or fake green? The impact of ESG rating disagreement on corporate greenwashing behavior. *Bus. Ethics Environ. Responsib.* **2025**. <https://doi.org/10.1111/beer.70058>
19. Wang, W.J.; Sun, Z.Y.; Wang, L. Does ESG rating divergence exacerbate management tone manipulation? – Empirical evidence based on MD&A text. *J. Bus. Res.* **2025**, *197*, 115449.
20. Bai, C.; Yao, D.; Xue, Q. Does artificial intelligence suppress firms' greenwashing behavior? Evidence from robot adoption in China. *Energy Econ.* **2025**, *142*, 108168.
21. Kurpierz, J.R.; Smith, K. The greenwashing triangle: adapting tools from fraud to improve CSR reporting. *Sustain. Account. Manag. Policy J.* **2020**, *11*, 1075–1093.
22. Soltani, B. The anatomy of corporate fraud: A comparative analysis of high profile American and European corporate scandals. *J. Bus. Ethics.* **2014**, *120*, 251–274.
23. Yu, J.Z.; Hwang, Y.S. Behind the green veil: Board characteristics, greenwashing, and the fraud triangle. *Humanit. Soc. Sci. Commun.* **2026**, *13*, 632.
24. Spence, M. Job market signaling. *Q. J. Econ.* **1973**, *87*, 355–374.
25. Tian, X.; Shi, D. Facilitating or inhibiting: A study on the impact of artificial intelligence on corporate greenwashing. *Sustainability.* **2025**, *17*, 2154.
26. Yang, J.; Cai, J.; He, L.; Hu, J.; Ma, X. Artificial intelligence and corporate internal control quality: Evidence from Chinese listed firms. *J. Risk Financ. Manag.* **2026**, *19*, 502.
27. Lei, X.; Kocoglu, M.; Nikbakht, E. The announcement effect: How AI adoption reduces corporate ESG greenwashing. *Financ. Res. Lett.* **2026**, *99*, 109890.
28. Huang, R.B.; Huang, Y.B. Does internal control contribute to a firm's green information disclosure? Evidence from China. *Sustainability.* **2020**, *12*, 3197.
29. Chen, P.Y.; Chu, Z.Z. Mere facade? Is greenwashing behaviour lower in low-carbon corporates? *Bus. Strateg. Environ.* **2024**, *33*, 4162–4174.
30. Jiao, A.; Lu, J.; Ren, H.; Wei, J. The role of AI capabilities in environmental management: Evidence from USA firms. *Energy Econ.* **2024**, *134*, 107653.
31. Du, X.Q.; Jian, W.; Zeng, Q.; Chang, Y.Y. Do auditors applaud corporate environmental performance? Evidence from China. *J. Bus. Ethics.* **2018**, *151*, 1049–1080.
32. Chen, L.; Fang, H.; Xiong, J.; Yang, Z. Institutional investors' site visits and corporate greenwashing: Evidence from China. *J. Int. Financ. Manag. Account.* **2026**, *37*, 38–62.
33. Gao, Y.X.; Chen, Y. Watchdogs or enablers? Analyzing the role of analysts in ESG greenwashing in China. *Sustainability.* **2024**, *16*, 4339.
34. Long, L.; Wang, C.; Zhang, M. Does social media pressure induce corporate hypocrisy? Evidence of ESG greenwashing from China. *J. Bus. Ethics.* **2025**, *197*, 311–338.
35. Wang, Y.; Hu, F.; Wang, Y. Analyst coverage and greenwashing: Evidence from Chinese A-Share listed corporations. *Int. Rev. Econ. Financ.* **2024**, *94*, 103423.
36. Zhang, C.; Wu, X. Can analyst coverage reduce corporate ESG greenwashing? Evidence from China. *Sustainability.* **2025**, *17*, 123.
37. Ren, J.Y.; Wu, P.; Hou, L.Y. Social media attention and corporate greenwashing: Evidence from China. *Corp. Soc. Responsib. Environ. Manag.* **2024**, *31*, 5446–5465.
38. Yin, L.; Yang, Y.Y. How does digital finance influence corporate greenwashing behavior? *Int. Rev. Econ. Financ.* **2024**, *93*, 359–373.
39. Zhan, X.; Zhan, Y.; Li, G. The Cost of ESG Decoupling: How Misaligned Disclosures Undermine Corporate Sustainability? *SAGE Open.* **2025**, *15*, 21582440251369170.
40. Liu, J.; Yuan, Y.; Zhu, Z. The role of artificial intelligence in enhancing ESG disclosure quality in Accounting. *J. Risk Financ. Manag.* **2026**, *19*, 15.
41. Jiang, N.; Wu, G.; Zou, W.; Chen, X. Responsible use of artificial intelligence and the mitigation of corporate greenwashing: The role of executive oversight. *Socio-Econ. Plan. Sci.* **2026**, *105*, 102436.
42. Zhang, D.Y. The pathway to curb greenwashing in sustainable growth: The role of artificial intelligence. *Energy Econ.* **2024**, *133*, 107562.
43. Yu, E.P.Y.; Van Luu, B.; Chen, C.H. Greenwashing in environmental, social and governance disclosures. *Res. Int. Bus. Financ.* **2020**, *52*, 101192.
44. Yoon, B.; Lee, J.H.; Byun, R. Does ESG performance enhance firm value? Evidence from Korea. *Sustainability.* **2018**, *10*, 3635.

45. Walker, K.; Wan, F. The harm of symbolic actions and green-washing: Corporate actions and communications on environmental performance and their financial implications. *J. Bus. Ethics*. **2012**, *109*, 227–242.
46. Acemoglu, D.; Autor, D.; Hazell, J.; Restrepo, P. Artificial intelligence and jobs: Evidence from online vacancies. *J. Labor Econ.* **2022**, *40*, S293–S340.
47. Wang, H.; Qiu, F. AI adoption and labor cost stickiness: Based on natural language and machine learning. *Inf. Technol. Manag.* **2023**, *26*, 163–184.
48. Zhang, J.; Huang, W.; Weng, Z.; Ma, J.; Wang, A. Words and deeds in AI transformation: CEO IT background and the rhetoric–substance gap. *J. Technol. Soc.* **2026**, *86*, 103336.
49. He, L.Y.; Wang, L. Can artificial intelligence curb greenwashing? Firm-level evidence based on large language model. *Energy Econ.* **2025**, *152*, 108145.
50. Heckman, J.J. Sample selection bias as a specification error. *Econometrica*. **1979**, *47*, 153–161.
51. Lewbel, A. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econ. Stat.* **2012**, *30*, 67–80.
52. Nunn, N.; Qian, N. US food aid and civil conflict. *Am. Econ. Rev.* **2014**, *104*, 1630–1666.
53. Khan, M.K.; Huo, C.; Zahid, R.M.A.; Maqsood, U.S. The automated sustainability auditor: Does artificial intelligence curtail greenwashing behavior in Chinese firms? *Bus. Strateg. Environ.* **2024**, *33*, 9015–9039.
54. Zhao, P.J.; Zeng, L.; Li, P.; Lu, H.; Hu, H.; Li, C.; Zheng, M.; Li, H.; Yu, Z.; Yuan, D. China’s transportation sector carbon dioxide emissions efficiency and its influencing factors based on the EBM DEA model with undesirable outputs and spatial Durbin model. *Energy*. **2018**, *238*, 121934.
55. Preacher, K.J.; Hayes, A.F. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods*. **2008**, *40*, 879–891.
56. Hayes, A.F.; Little, T.D. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 3rd ed.; The Guilford Press: New York, NY, USA; pp. 73–83.
57. Yang, X.; Liu, B.; Liu, Z.; Zhang, J.; Sun, Q. Does environmental uncertainty increase the likelihood of greenwashing? The roles of government subsidies and media attention. *Corp. Soc. Responsib. Environ. Manag.* **2025**, *32*, 2616–2629.
58. Paliwal, G.; Donvir, A.; Gujar, P.; Panyam, S. Accelerating Time-to-Market: The role of generative AI in product development. In proceedings of the 2024 IEEE Colombian Conference on Communications and Computing (COLCOM), Barranquilla, Colombia, 21–23 August 2024; pp. 1–9.
59. Xue, X.; Wang, L.; Hu, N. Economic policy uncertainty and corporate social responsibility disclosure similarity: Evidence from China. *British Accounting Review*. **2024**, *56*, 101305.
60. Yan, H.; Li, D.; Cao, P. Beyond the single thread: How organisational, technological, and environmental factors jointly shape green patent persistence in textile firms. *Ind. Textila*. **2026**, *77*, 115.
61. Hao, X.L.; Miao, E.X.; Wen, S.F.; Wu, H.T.; Xue, Y. Executive green cognition on corporate greenwashing behavior: Evidence From A-Share listed companies in China. *Bus. Strateg. Environ.* **2025**, *34*, 2012–2034.
62. Yu, J.Y.; Qiu, Q.; Qiao, Y.Y. Managerial decision horizon and corporate greenwashing: Evidence from China. *Bus. Ethics Environ. Responsib.* **2025**, *34*, 1662–1683.
63. Su, Y.Y.; Zhu, C.H.; Albitar, K. Market competition and corporate ESG greenwashing: A perspective from new entrants. *Bus. Strateg. Environ.* **2026**, *35*, 2266–2294.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.