

AI washing in corporate finance: conceptual foundations, measurement, economic consequences and future directions

Yongzheng Cai

Shenzhen Audencia Financial Technology Institute, Shenzhen University, Shenzhen, China

yongzheng.cai@audencia.com

Abstract. Artificial Intelligence (AI) has become the most attractive label in corporate strategy narratives, yet the intensity with which firms talk about AI has begun to diverge systematically from what they actually do. This phenomenon is known as AI washing. Drawing on four core empirical studies, this paper reviews the corporate finance literature on AI washing. This study develops an operational definition—a positive gap between AI Talk (disclosure) and AI Walk (substantive investment)—and delimit it against greenwashing, digital washing and impression management. This study explains why the gap pays through signalling, information asymmetry and image management, and identify cognitive dissonance as a countervailing corrective force. This study catalogues how existing work measures the construct and assess the validity of each proxy. Three outcome chains are synthesized: AI washing increases access to bank loans but not necessarily the quality of the money obtained; it has an inverted U-shaped effect on the technological gap; and it inflates investor optimism, widens analyst disagreement and raises crash risk. This study examines who can detect AI washing and conclude with seven directions for future research, arguing that the field's next step is a systematic answer to why firms wash, who believes them, who detects them, and what the practice buys against what it costs.

Keywords: AI washing, information disclosure, signalling theory, cognitive dissonance, bank credit, debt maturity mismatch, stock price crash risk, technological gap

1. Introduction

Artificial Intelligence (AI) has evolved from a specialized research topic into a general-purpose technology with economy-wide reach. As it diffuses, firms have acquired a strong incentive to be seen as "AI companies". Annual reports, Management Discussion and Analysis (MD&A) sections, strategic plans and press releases are now saturated with the vocabulary of AI transformation, AI enablement and intelligent upgrading. Being perceived as an AI firm is associated with higher valuations, easier access to policy-directed credit and stronger stakeholder approval.

This enthusiasm rests on a fundamental difficulty: AI capability is largely unobservable from outside. It resides in algorithms, data pipelines, model deployments and organizational knowledge rather than in tangible assets that outsiders can inspect. Algorithmic black boxes, the proprietary character of AI systems, long and uncertain development cycles and the absence of industry assurance standards together produce a verifiability

gap. As a result, even sophisticated counterparties such as banks, analysts and institutional investors cannot reliably distinguish genuine AI capability from rhetorical claims. When verifiability fails, disclosure becomes cheap talk—and cheap talk can be used strategically.

This is the terrain on which AI washing emerges. It defines it as the positive gap between a firm's AI Talk—the extent to which it discloses AI capability, strategy and applications—and its AI Walk—the substantive AI investment and capability it actually possesses. Liu and Li [1] describe it as opportunistic behavior in which firms deliberately exaggerate or even fabricate their level of AI application in order to attract market attention and obtain commercial benefits. Liu et al. [2] define it as the strategic inflation of AI-related disclosure beyond actual investment in AI capability. Sun et al. [3] frame it as a gap between AI narrative and actual capability—misleading technology marketing. All three definitions share a common core: a deliberate decoupling of symbolic communication from substantive action in a specific and highly salient technological domain.

A small but rapidly growing empirical literature has begun to trace the consequences of this decoupling. Liu and Li [1] find that AI washing improves firms' access to bank loans, and that external governance mechanisms weaken the effect. Sun et al. [3] document an inverted U-shaped relationship between AI washing and the corporate technological gap, channeled through both firm-level and industry-level R&D investment, with organizational cognitive dissonance providing the corrective mechanism. Liu et al. [2] show that AI washing raises stock price crash risk by inflating investor optimism and degrading the information environment. A fourth study, although located in a different field, supplies the conceptual and methodological infrastructure for the question the existing literature cannot yet answer: Liu et al. [4] show that credible external certification increases long-term borrowing and extends debt maturity, thereby alleviating corporate maturity mismatch. That design provides a template for asking a more informative question—not whether AI washing brings more credit, but whether it brings better credit.

Three problems nonetheless leave the field fragmented. First, the construct is not settled: some studies treat AI washing as a continuous deviation, others as a binary act, and almost none theorize the boundary between strategic exaggeration and reasonable optimism about an emerging technology. Second, measurement is inconsistent along both dimensions: action-side proxies include AI capital expenditure, AI patent applications, AI patents granted and AI hiring, and it is unclear whether these are interchangeable. Third, results accumulate dependent variable by dependent variable—bank credit, technological gap, crash risk—without an integrating account that connects the mechanisms. The risk is a list of findings rather than an explanation of a phenomenon.

2. Definition of AI washing

2.1. An operational definition

AI washing is a strategically manufactured positive gap between the AI capability a firm claims and the AI capability it actually possesses. The construct has four defining elements.

First, a gap: The behavior requires a measurable deviation between disclosure and substance. The gap is not defined against an absolute standard but against a benchmark: Liu and Li [1] classify a firm as washing when its AI disclosure intensity is above, and its AI investment below, the industry-year mean, while Liu et al. [2] and Sun et al. [3] standardize the two components within year and take the difference. The construct is therefore relational—a firm washes relative to its industry peers in the same period, not relative to some absolute standard of AI capability.

Second, intent: The gap must originate in a strategic decision rather than in honest but unrealized intentions or in an exogenous shock that obstructed execution. This is the hardest element to verify empirically, and it is precisely why all existing measures are indirect: they identify observable symptoms of the gap rather than the intention behind it. In practice, benign explanations are excluded through benchmarks and reference groups.

Third, technological domain: The claim must refer specifically to AI. This matters because the salience of AI as a technology label is itself the source of the payoff, and because the verifiability of AI claims differs systematically from that of other disclosure domains.

Fourth, disclosure as the instrument: Firms act through language—annual reports, MD&A, sustainability and strategy documents, press releases—rather than through misstated accounting numbers. AI washing is a narrative strategy, which places it firmly in the disclosure literature rather than the earnings management literature.

The construct decomposes into two measurable components: AI Talk and AI Walk. AI Talk captures symbolic communication: the volume and prominence of a firm's stated AI capability, strategy and applications. AI Walk captures substantive action: what the firm actually invests, patents, hires and deploys. AI washing is the difference between them. This decomposition is the analytical backbone of the entire empirical literature, and all three core studies operationalize it the same way at the conceptual level, even though they do not agree at the measurement level (See Section 4).

2.2. AI washing is not AI optimism

Three boundaries need to be stated explicitly, because failing to draw them produces false positives in measurement and confusion in interpretation.

AI washing is not AI optimism. Many firms genuinely believe AI will be essential to their future and say so before substantive results appear. Ambition about an emerging general-purpose technology is not deception. The distinguishing criterion is the relationship between the claim and the firm's own investment behaviour. A firm that claims AI leadership while maintaining high investment relative to peers is expressing optimism; a firm that claims AI leadership while investing less than peers is washing. This is why all existing measures use contemporaneous, within-industry benchmarks rather than absolute thresholds.

AI washing is not execution failure. A firm may invest seriously in AI and still fail, because AI projects have long and uncertain payoffs [3]. Such a firm exhibits a gap between commitment and outcome, not between talk and action: the action occurred. Conversely, the absence of visible results does not prove washing, because AI capability may be genuine but not yet monetised. Outcome-based rather than input-based measurement therefore risks conflating execution failure with strategic misrepresentation.

AI washing is not a single act. It is a portfolio of claims running from the relatively defensible (describing a real AI-assisted process in inflated strategic language) to the plainly false (claiming a deployment that does not exist). Liu and Li [1] explicitly allow fabrication and exaggeration to coexist. The intensity of the behaviour is thus continuous, with a direct implication for measurement: binary classification discards the information that non-linear specifications require.

2.3. Adjacent constructs

AI washing belongs to the family of "decoupling" constructs. Table 1 compares it with the three closest constructs.

Table 1. AI washing and adjacent decoupling constructs

Dimension	AI washing	Greenwashing	Digital washing	Impression management
Object of the claim	AI capability, strategy and applications	Environmental performance and commitments	Digital transformation and digitalization	Any organizational attribute
Primary motive	Offensive: capture a technology growth premium	Defensive: preserve legitimacy and deflect regulation	Mixed: legitimacy and resource acquisition	Broad: manage audience perceptions
Verifiability of the claim	Very low: algorithmic black boxes, proprietary systems, no assurance standards	Low to moderate: third-party ESG ratings and disclosure rules exist	Low: broad and loosely defined scope	Varies with the attribute claimed
Detection benchmark	Industry-year peers; within-year standardized talk-walk difference	ESG ratings, emissions data, certification schemes	Digitalization indices, IT spending data	Analyst and media scrutiny
Documented consequences	Higher bank credit; non-monotonic innovation effects; higher crash risk	Capital misallocation, reputational and regulatory penalties	Distorted digital strategy, wasted resources	Mispricing, expectation gaps

The comparison shows that AI washing is analytically distinct rather than a new instance of greenwashing. Two differences are decisive. The first is motive. Liu et al. [2] argue that greenwashing is typically defensive, aimed at preserving legitimacy or deflecting regulatory risk, whereas AI washing is an offensive strategy through which firms signal technological leadership and capture a growth premium [5]. The payoff structures therefore differ: the payoff to greenwashing is avoiding penalties, whereas the payoff to AI washing is capturing a premium. The second, and more substantive, is verifiability. Environmental performance can at least be partially observed through emissions, energy use and third-party ratings, and a substantial assurance infrastructure has grown up around it [6]. AI capability cannot. Technical complexity, algorithmic black boxes, the proprietary character of AI systems and the absence of standardised AI disclosure or assurance frameworks mean that these claims are substantially harder to falsify [3]. The practical implication is that AI washing should be both more prevalent and harder to detect than greenwashing—an asymmetry that reappears in Section 8, because it predicts where external governance can and cannot constrain the behaviour.

2.4. A typology

Although the literature treats AI washing as a single construct, three variants can be distinguished, and they are likely to have different consequences. Capability washing exaggerates the AI capability a firm already has, for example describing routine automation as machine learning. Application washing exaggerates the number or importance of AI applications in products and processes, coming closest to fabricating a deployment that does not exist. Commitment washing involves forward-looking statements about AI strategy, which are inherently harder to falsify because the outcomes lie in the future. Commitment washing is probably the most common form in practice, and it is also the form that Sun et al.'s [3] cognitive dissonance mechanism is best suited to explain: a public commitment the firm cannot currently honour creates the internal tension that subsequently

drives genuine investment. Distinguishing these variants empirically is an unresolved measurement problem to which Section 10 returns.

3. The mechanism of AI washing

If AI washing did not pay, there would be nothing to explain. The literature identifies three mutually reinforcing mechanisms and one countervailing force—the force that keeps the behaviour from being merely a transfer of value from outsiders to the firm.

3.1. Signalling with an unverifiable label

The first mechanism is the AI label functioning as a signal. Standard signalling logic requires that a credible signal be costly to imitate, so that only firms with genuine quality can afford to send it, with cost serving as the mechanism that separates true from false types [7]. The AI signal violates this condition in two ways: it is extremely cheap to send, requiring only the use of particular vocabulary in a disclosure document, and it is extremely hard to verify, because the technology is opaque and no assurance standard exists. When false signals are cheap and verification is expensive, the separating equilibrium breaks down and the signal loses its information content while retaining its rhetorical value. Managers therefore disclose selectively, weighing the costs and benefits of release [8].

Why, then, does the label retain value when it carries no information? Because receivers do not condition on verifiability. Banks operating under a policy orientation that favors technology firms have lending preferences aligned with the AI narrative. In that environment, AI claims lower banks' subjective perception of credit risk whether or not the claims are defensible. Liu and Li [1] provide direct evidence: AI washing significantly increases the scale of new bank loans, with coefficients between 0.0061 and 0.0087 of lagged total assets across specifications, all significant at the 1% level. The signal works because it is consistent with the receiver's priors, not because it is informative.

The salience of AI as a technology also means that the growth premium attached to the label operates independently of the credit channel. Sun et al. [3] note that AI narratives can create a short-run appearance of technological progress, attract capital inflows and raise market expectations for an entire industry, so that firms which do not engage in AI narrative may lose investor sentiment. This creates a coordination problem: when peers all talk about AI, silence is costly, and the equilibrium level of AI Talk rises on its own with no change in AI Walk.

3.2. Information asymmetry and verification costs

Information asymmetry is the core mechanism, and the literature treats it as the channel through which the other two operate. Three layers of opacity reinforce one another.

The first layer is technological. The complexity of AI and the algorithmic black box make it difficult for external stakeholders, including ordinary investors and consumers, to verify accurately the AI capability a firm claims [3]. A firm can claim that a particular product or operating process uses AI, and the high technical barriers and proprietary character of AI systems make such claims effectively unfalsifiable for non-specialists [2].

The second layer is institutional. There is no AI disclosure standard comparable to accounting standards, and no widely accepted AI assurance framework. The field's own definitional ambiguity compounds the problem: what counts as AI is itself contested, which means firms retain considerable discretion over what to

report and how to frame it. Sun et al. [3] emphasise that current governance frameworks in the AI field remain vague, so the arbitrage space between legal risk and potential reward remains wide.

The third layer is data fragmentation. Even where verification technology exists, the data required to apply it may not be available. Liu and Li [1] make this argument explicitly: fintech has improved banks' information-processing capacity but cannot fundamentally identify AI washing, because large-scale information verification is difficult and data barriers persist. The problem is information siloing and fragmentation—either the information firms accumulate cannot be processed effectively by banks, or banks lack sufficient digital footprint to assess firms accurately. This residual leaves room for washing.

The consequences of this multi-layered opacity are observable in the information environment itself. Liu et al. [2] show that AI washing significantly widens analyst forecast dispersion, with a coefficient of 0.051 on their disclosure opacity measure. Analysts, the most sophisticated information intermediaries available, become worse at forecasting the earnings of washing firms. Opacity is therefore not only a precondition of the behavior but also a consequence of it, which constitutes the self-reinforcing dynamic formalized in Section 9.

3.3. The countervailing force: washing as a commitment device

These mechanisms imply that AI washing extracts value from the less informed party, but they do not imply that the behavior is costless to the firm or that its effects are monotonic. A distinctive contribution of the recent literature is the identification of a corrective force operating from inside the organization.

Sun et al. [3] argue that the inconsistency between a firm's AI claims and its actual capability generates the psychological tension described by cognitive dissonance theory [9]. At low levels of washing, the immediate benefits—market recognition, policy support, valuation gains—provide a psychological reward that allows the firm to rationalize the inconsistency, so no correction occurs. As washing intensifies, however, the conflict between what is claimed and what exists becomes harder to rationalize, particularly for employees and managers in direct contact with the technology. Once accumulated dissonance exceeds the organization's tolerance threshold, consistency-restoring mechanisms activate: the firm increases genuine R&D investment to bring its actions into line with its words. The resources the narrative has accumulated—reputational gains and improved capital access—then provide the financial basis for that investment. Sun et al. [3] illustrate the mechanism with the case of iFlytek, which promoted AI concepts intensively in its early years and saw its technological gap widen while its actual technical reserves were limited, before gradually closing the gap by increasing R&D investment, refining its technology strategy and recruiting technical talent.

This argument matters for how the phenomenon should be theorized. AI washing is not merely an ethical violation with costs; it is a dynamic organizational behavior whose consequences depend on the level of the behavior, the firm's absorptive capacity and the institutional environment. Whether washing makes a firm "forever false" or "forces it to become real" is an empirical question, and Section 6 reviews the relevant evidence.

4. The measurement of AI washing

All empirical measures of AI washing follow the same architectural logic: measure AI disclosure, measure AI substance, and compare the two. The literature differs in how each component is constructed and how the comparison is aggregated. Table 2 summarises the designs.

Table 2. Measurement of AI washing across four studies

Study	Sample and setting	AI Talk (disclosure proxy)	AI Walk (action proxy)	Aggregation rule
Liu et al. [2]	Chinese A-shares, AI-related word 2010–2024, 33,463 firm-years	AI-related word frequency in MD&A divided by total words	AI-related capital expenditure divided by total assets; identified from intangibles and fixed assets carrying AI keywords	Continuous: within-year standardised Talk minus standardised Walk; higher values indicate more severe washing
Liu and Li [1]	Chinese A-shares, 2010–2023, 20,140 observations	Full annual report keyword frequency, weighted by inverse document frequency within the industry-year; TM = 1 if above the industry-year mean	Level of AI investment; DL = 1 if below the industry-year mean	Binary conjunction: AI washing = 1 if and only if TM = 1 and DL = 1
Sun et al. [3]	Chinese A-shares, 2007–2022, 3,842 firms, 36,200 firm-years	AI-related keyword frequency in annual reports, using a self-constructed dictionary	Number of AI-related patents granted that year (CNIPA)	Continuous: within-year standardised Talk minus standardised Walk
Liu et al. [4] (methodological benchmark)	Chinese A-shares, 2014–2022, 1,757 firm-years; GSCM certification	Not applicable (treatment is externally verified certification)	Actual financing structure: long-term loans and debt maturity	Difference-in-differences on a propensity-score-matched sample; source of the maturity mismatch measure

4.1. Aggregation: continuous gaps and binary classification

The two aggregation rules in use have different properties, and the choice affects both estimation and interpretation. The continuous rule, used by Liu et al. [2] and Sun et al. [3], defines AI washing as the standardised difference between AI Talk and AI Walk, with both components standardised using their cross-sectional mean and standard deviation within each year. Higher values indicate greater divergence. This rule preserves the intensity of the behaviour, which is indispensable when the relationship of interest is non-linear—precisely what Sun et al. [3] demonstrate, since their central finding of an inverted U-shaped relationship could not have been identified with a binary measure. The rule also treats both sides symmetrically, so firms that talk more than they do and firms that do more than they say lie at opposite ends of the same scale, the latter being interpretable as understatement.

The binary rule, used by Liu and Li [1], defines AI washing as the conjunction of "talking more than peers" and "doing less than peers", with both thresholds set at the industry-year mean. This corresponds to how a bank or regulator would naturally classify a firm, and its coefficient can be interpreted directly as the effect of "being a washing firm". The cost is the loss of intensity information and reliance on an arbitrary threshold: a

firm slightly above the mean in talk and slightly below it in action is grouped with extreme cases, and changes in sample composition change who is classified as a washer.

5. AI washing and corporate financing

The most direct and best-evidenced consequence of AI washing is its effect on credit availability. Here the evidence is unusually clear along one dimension and conspicuously silent along another, and the gap between them is the most promising direction in the field.

The financing literature therefore establishes one clear result and leaves one equally clear gap. The result is that AI washing robustly raises the quantity of bank credit across settings, and that the signalling mechanism works because the signal is consistent with lenders' priors rather than because it has been verified. The gap is that nothing is known about whether this credit is good money. Evidence from certified financing suggests what the answer might look like: more credit coexisting with shorter maturities and more severe maturity mismatch. Until that question is tested directly for AI washing, the claim that AI washing eases financing constraints remains unsupported. It is equally possible that AI washing merely substitutes cheap short-term funding for expensive long-term funding, which would leave firms more exposed rather than less constrained.

6. AI washing and innovation

The second outcome chain concerns the effect of AI washing on firms' real technological trajectories. Here the literature reaches its theoretically most interesting conclusion, and it is the opposite of the intuition with which the research question began. Two opposing predictions follow from the same premises. The first is crowding out. If narrative strategies pay well while genuine innovation pays uncertainly and slowly, a firm that discovers it can obtain recognition and resources by talking will shift effort away from substantive activity. Symbolic communication substitutes for R&D, capability accumulation falls behind where it should be, and the firm falls further from the technological frontier. This is the mechanism familiar from the greenwashing literature, where symbolic environmental behaviour crowds out substantive environmental investment [5]. The second is a commitment or correction mechanism. If a firm has publicly committed to AI leadership, it has created an expectation it must eventually honour. To avoid exposure it must either elaborate the narrative further or begin building capability. As described in Section 3.4, the internal tension created by the gap between claim and reality provides a direct incentive to close it. On this logic, washing is a self-correcting deviation: exaggeration eventually forces the firm to become what it claimed to be. The evidence resolves this tension not by choosing between the two predictions but by showing that the direction of the effect depends on the level of the behaviour. Sun et al. [3] document a robust inverted U-shaped relationship between AI washing and the corporate technological gap, where the gap is measured as the difference between the patent output of frontier firms in the industry and that of the focal firm. In the specification containing the full set of controls, the linear term for AI washing is positive and significant (0.104) and the quadratic term is negative and significant (−0.0196), both at the 1% level. The implied turning point is 2.6531, and a formal test of the slope confirms a range from −0.0718 to 0.1988. The finding holds under alternative measures of the key variables, stricter fixed-effect specifications and controls for potential omitted variables. Figure 1 plots the estimated relationship.

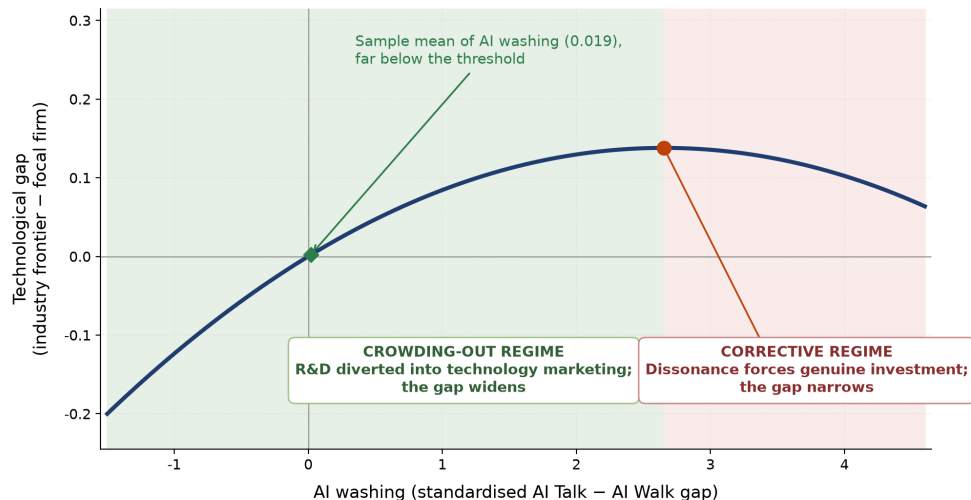


Figure 1. AI washing and the corporate technological gap, reproduce from ref. [3]

The economic interpretation is a sequence rather than a single effect. At low levels, moderate AI washing helps firms obtain external resources by shaping a technological image and catering to market expectations. Because those resources do not correspond to substantive investment in AI capability, the firm's distance from the industry frontier widens. As washing intensifies further, resource misallocation and erosion of the technological advantage continue. Beyond the turning point, however, AI-related disclosure diverges so far from actual capability that cognitive dissonance becomes acute, and the firm is compelled to increase R&D investment and strengthen its AI capability, so the gap begins to narrow. Low levels of washing make a firm more false; high levels force it to become more real.

This result carries an uncomfortable methodological implication. The turning point is estimated in a sample where the mean of AI washing is 0.019 and the standard deviation is 0.769, with values ranging from -2.413 to 4.469 . The critical threshold therefore lies far above the mean, and the observed narrowing of the gap applies only to a small number of extreme firms. The practical relevance of the corrective branch depends entirely on how many firms actually enter that region, and the available evidence cannot answer that question.

7. AI washing and capital markets

The third outcome chain concerns investors, information intermediaries and prices. Here the literature provides its clearest evidence that the short-run benefits and long-run costs of AI washing are two sides of the same mechanism. Attention and investor optimism: The AI label attracts attention, and attention translates into optimism when narratives are unverifiable. Liu et al. [2] test this channel directly by measuring investor sentiment from stock forum posts, classified by machine learning algorithms as positive, neutral or negative, with sentiment defined as positive posts minus negative posts divided by their sum. AI washing significantly increases investor optimism, with a coefficient of 0.003 significant at the 5% level. Exaggerated AI claims raise investor expectations and push prices beyond what fundamentals can support. Evidence from US firms, using BERT-based text classification, shows a comparable pattern: unsubstantiated AI claims initially attract investors but ultimately produce negative market reactions and sustained underperformance [10].

Media amplify rather than constrain this process. Liu et al. [2] find that the effect of AI washing on crash risk is stronger for firms with more positive news coverage, reasoning that positive coverage lends legitimacy to a firm's AI claims and thereby intensifies the valuation bubble. This is a striking asymmetry: the media are

normally understood as information intermediaries that reduce asymmetry, but in the presence of unverifiable technological claims they become a transmission channel for the narrative. The finding reframes media coverage as a contextual amplifier of disclosure manipulation rather than a restraint on it.

The second mechanism is the degradation of the information environment. Liu et al. [2] measure information opacity using analyst forecast dispersion, defined as the standard deviation of analysts' annual earnings per share forecasts divided by the absolute value of the mean forecast, and find that AI washing significantly widens dispersion, with a coefficient of 0.051 significant at the 5% level. AI washing therefore worsens the corporate information environment in a measurable way: the most professional intermediaries available become worse at predicting the earnings of washing firms.

This result is theoretically important because it constitutes the condition for eventual correction. Opacity is not only a precondition for successful washing but also a consequence of it, and accumulated opacity is what allows bad news to be withheld. Both mechanisms are necessary for a crash: optimism creates the bubble, and opacity allows the bubble to accumulate in silence.

The ultimate consequence is an asymmetric increase in the probability of extreme negative returns. Liu et al. [2] show that AI washing raises stock price crash risk, measured by the negative conditional skewness of firm-specific weekly returns and by the down-to-up volatility ratio. The coefficients are positive and significant at the 1% level in both specifications, at 0.023 and 0.011 respectively; a one standard deviation increase in AI washing corresponds to increases in crash risk equivalent to 3.4% and 2.5% of the respective sample standard deviations. The result holds when AI patent applications and AI job postings are used as alternative measures, when province-year and industry-year fixed effects are added, on a propensity-score-matched sample, and in a two-stage least squares specification using the provincial Baidu AI search index as an instrument, with a first-stage F-statistic of 18.914.

The theoretical chain is a story of bad news hoarding. Exaggerated AI capability cannot generate matching cash flows, so negative performance signals are inevitable. Because the inflated valuation can only be sustained by withholding information, managers conceal those negative signals. Bad news accumulates inside the firm until the suppressed volume exceeds a critical threshold and is then released abruptly, triggering a sharp downward revision of investor expectations and a violent price decline [11-13].

An important conceptual distinction emphasised by Liu et al. [2] separates this result from general volatility in the AI sector. Genuine AI firms face uncertainty about future cash flows and may therefore exhibit high return volatility, and evidence linking AI investment to crash risk through reduced transparency and heightened managerial optimism [14] reinforces the need for the distinction; but given timely and transparent disclosure, that volatility is symmetric in the return distribution and does not produce asymmetric concentration of extreme negative returns. Washing firms, by contrast, lack substantive technological capability and rely on strategic concealment to sustain their valuations, producing systematic left-tail asymmetry. Narrative-driven crash risk and the volatility inherent in legitimate AI innovation are therefore empirically distinct. This distinction matters for how the evidence is read: finding that AI-related stocks are volatile is not informative about washing, whereas finding left-tail asymmetric risk conditional on a talk-action gap is directly informative.

8. The detection of AI washing

Whether AI washing pays depends not only on what firms say but also on whether the parties receiving the claims can judge their truth. This section assembles the evidence on detection capacity. Liu and Li [1] provide the organizing framework. In signaling theory, the effectiveness of a signal depends not only on the signal

itself but also on the strength of the constraints in the transmission environment; external governance can optimize the signaling environment and strengthen receivers' ability to identify false reputation [7]. The claim is precise: the effectiveness of a false signal is proportional to the weakness of the transmission environment, not merely to the attractiveness of the signal. This converts the detection question into an empirical one about which governance mechanisms actually bind.

9. Conclusion

AI washing is not simply corporate boasting about AI. It is a disclosure strategy with identifiable economic consequences, operating through a specific mechanism—the unverifiability of technology claims—and affecting corporate financing, innovation and capital market stability simultaneously. The evidence assembled in this review establishes four results. AI washing improves firms' ability to obtain bank loans, because the AI label is consistent with lenders' priors and cheap to send [1]. It has a non-monotonic effect on innovation, crowding out R&D and widening the technological gap at low levels while triggering a corrective response at high levels [3]. It inflates investor optimism, degrades the information environment and raises stock price crash risk [2]. And the credit it attracts may be of lower quality than its quantity suggests, because analogous evidence on certified financing shows that more credit can coexist with shorter maturities and more severe maturity mismatch [4]. Across all three outcome chains, external governance—analyst coverage, audit quality, regulatory intensity and institutional quality—weakens the effect, with regulation the most powerful constraint.

The field is nonetheless at an early stage, and its gaps are structural rather than incidental. The construct is not yet settled, the measurement of substantive AI capability remains crude, and all available estimates come from a single institutional setting. Above all, the literature risks accumulating dependent variables without accumulating explanation: adding one more outcome that correlates with AI washing will not advance understanding of the phenomenon.

A more productive agenda would answer four questions systematically. Why do firms wash—which managerial incentives, financing pressures and peer dynamics sustain the behavior? Who believes them—which receivers are systematically misled, and under what conditions? Who can detect them—which governance mechanisms actually bind, and in what order? And what does washing buy in the short run against what it costs in the long run—not merely how much credit it attracts, but what kind of credit, what kind of innovation, and what kind of risk it leaves behind? Answering these questions would convert a set of findings about AI washing into an explanation of why unverifiable technology narratives have become a durable feature of corporate finance, and of what can be done about them.

References

- [1] Liu, W., & Li, M. (2026). The impact of AI washing on enterprises' access to bank loans: From the perspective of external governance. *Finance Research Letters*, 98, 109884. <https://doi.org/10.1016/j.frl.2026.109884>
- [2] Liu, Z., Ran, M., & Wang, M. (2026). All that glitters? Corporate AI washing and stock price crash risk. *Finance Research Letters*, 99, 109882. <https://doi.org/10.1016/j.frl.2026.109882>
- [3] Sun, Z., Wen, Y., Zhao, L., Almgren, I., & Galgotia, A. (2026). Unveiling AI washing: Bridging corporate technological gaps through a cognitive dissonance lens. *Technological Forecasting and Social Change*, 225, 124511. <https://doi.org/10.1016/j.techfore.2025.124511>
- [4] Liu, M., Shen, L., & Yan, Y. (2025). Government-led green supply chain demonstration and corporate maturity mismatch: Evidence from China. *Finance Research Letters*, 75, 106870.

<https://doi.org/10.1016/j.frl.2025.106870>

- [5] Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California Management Review*, 54(1), 64–87. <https://doi.org/10.1525/cmr.2011.54.1.64>
- [6] Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2), 483–504. <https://doi.org/10.1287/orsc.2015.1039>
- [7] Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, 37(1), 39–67. <https://doi.org/10.1177/0149206310388419>
- [8] Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, 5, 179–194. [https://doi.org/10.1016/0165-4101\(83\)90011-3](https://doi.org/10.1016/0165-4101(83)90011-3)
- [9] Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- [10] Song, X., Hou, W., Ouyang, Z., & Hao, F. (2026). AI washing: Strategic disclosure and backlash. *Finance Research Letters*, 95, 109684. <https://doi.org/10.1016/j.frl.2026.109684>
- [11] Jin, L., & Myers, S. C. (2006). R2 around the world: New theory and new tests. *Journal of Financial Economics*, 79(2), 257–292. <https://doi.org/10.1016/j.jfineco.2004.11.003>
- [12] Hutton, A. P., Marcus, A. J., & Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. *Journal of Financial Economics*, 94(1), 67–86. <https://doi.org/10.1016/j.jfineco.2008.10.003>
- [13] Kim, J.-B., Li, Y., & Zhang, L. (2011). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics*, 100(3), 639–662. <https://doi.org/10.1016/j.jfineco.2010.07.007>
- [14] Bai, X., Zhao, W., & Liu, M. (2026). Artificial intelligence and stock price crash risk: Evidence from China: A pre-registered study. *Pacific-Basin Finance Journal*, 96, 103039. <https://doi.org/10.1016/j.pacfin.2025.103039>