

Can Regulation Curb AI Adoption Narrative—Practice Decoupling? An Agent-Based Model of the A-Share Market

ming sun

sunming@zqu.edu.cn

Zhaoqing University

Research Article

Keywords: AI adoption, narrative–practice decoupling, agent-based modelling, financial markets, information disclosure, regulatory design

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11410728/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Firms increasingly communicate their adoption of artificial intelligence (AI), but the intensity of such narratives may not correspond to the extent of AI implementation in their existing business operations. This paper examines how the gap between firms' public AI narratives and observable AI-related activities interacts with investor behaviour and regulatory design. We build an agent-based model of firms, retail and institutional investors, and a regulator, informed by firm-level data on 3,133 A-share listed firms (2021–2024). AI narrative intensity (Talking) is measured from annual-report text, while observable AI implementation (Doing) is proxied by AI-related hiring and patenting; their standardised difference measures narrative–practice decoupling, not proof of deceptive intent. Under a common AI-hype shock, we compare nine stylised regulatory mechanisms with a no-regulation baseline across four intensity levels (200 ensemble simulations). Four results emerge. The market's own pricing discipline is weak ($\beta_D \approx -0.37$ at $\lambda = 0$). Effects stratify sharply: repeat-offender escalation (-2.07) and targeted enforcement (-1.80) are strongest, while short selling and label-only disclosure coincide exactly with the no-regulation baseline (zero effect under this implementation). Ablation shows that structural design—escalation rather than raw penalty size—drives effectiveness. A Pareto search identifies repeat-offender escalation plus targeted enforcement as a recurring core pair (85.2% of non-dominated solutions) and the best tested combination ($\beta_D = -2.125$), with deterrence non-monotonic in the number of mechanisms. These findings are conditional on the model's behavioural assumptions; policy implications are illustrative and require validation against real-market outcomes.

JEL classification: C63; G14; G18; O33; K22

1 Introduction

1.1 Research Background

When a firm's annual report advertises an AI transformation that its investment budget does not fund, the market confronts a disclosure problem that price discovery alone does not resolve: capital may flow toward the most articulate narrators rather than the most substantive adopters. This talk–do gap—what we call **AI narrative decoupling**—is not a transient friction. It is a persistent tension between the strategic value of narrative and the informational function of prices, and it determines both where scarce capital is allocated and whether disclosure carries credible information. Narrative is cheap to produce and AI claims are hard to verify: a firm can relabel an existing algorithm an “AI platform” at negligible cost, whereas genuine adoption requires hiring, patenting and sustained investment that appears only slowly in financial statements. The gap therefore has a testable direction: if investors cannot separate the two, valuation should respond to what firms say rather than to what they do.

Scope of AI adoption. This study examines firms that adopt, or seek to adopt, artificial intelligence in their existing business operations, rather than firms whose primary business is the development or sale of AI products and services. Adoption may serve productivity, cost reduction, managerial decision-

making or product quality; firms may also communicate AI ambitions to attract investor attention or support valuation even when implementation remains limited. We therefore distinguish Talking (public AI narrative intensity) from Doing (observable AI-related implementation), and treat their difference as narrative–practice decoupling rather than as proof of deception. The analysis asks how this gap interacts with investor behaviour, firm valuation and alternative regulatory rules in an agent-based model of the A-share market.

1.2 Literature Review

Classical theory supplies the benchmark. Akerlof (1970) shows that when quality is unobservable, costless claims cannot sustain a separating equilibrium; Spence (1973) shows that credible separation requires signals whose cost differs across firm types; Becker (1968) and Polinsky and Shavell (2000) formalise enforcement as an expected penalty (magnitude $\times$ detection probability) and establish the case for escalating sanctions against repeat offenders. Under the efficient-market benchmark (Fama 1970), prices should track fundamentals rather than narrative, so a systematic talk–do gap is a matter for disclosure regulation. Recent work sharpens the institutional picture: transparency is necessary but not sufficient (Fung et al. 2007); “comply or explain” has been weakly enforced (MacNeil and Li 2006); non-traditional actors account for much fraud detection (Dyck et al. 2010); short-sale constraints cause systematic overpricing (Miller 1977; Diamond and Verrecchia 1987); and economics has turned to agent-based and narrative-based modelling (Mantegna and Stanley 2000; Bouchaud and Potters 2003; Farmer and Foley 2009; Shiller 2019), treating markets as interacting particle systems whose regularities emerge from micro-level rules. Shiller’s narrative economics adds that stories move prices through contagion irrespective of truth (Shiller 2019); cheap-talk theory questions whether disclosure alone can deter unsupported claims (Crawford and Sobel 1982; Farrell and Rabin 1996); and the “greenwashing” literature offers a direct reference (Lyon and Maxwell 2011; Delmas and Burbano 2011). What distinguishes this paper is that it answers counterfactual questions that difference-in-differences, instrumental variables and regression discontinuity cannot: how pricing shifts between fundamentals and narrative, and how regulatory rules shift that response. Drawing on Mantegna and Stanley (2000), Bouchaud and Potters (2003) and Farmer and Foley (2009), we treat investor agents as “particles”, β_D as a “response coefficient” and λ as a “control parameter”; because the model is stochastic, results are summarised across many independent runs under identical shocks with rules varied across runs.

Recent research sharpens the institutional picture without resolving the comparison problem. Transparency is necessary but not sufficient (Fung et al. 2007); “comply or explain” has in practice been weakly enforced, with market discipline permitting frequent non-compliance (MacNeil and Li 2006); non-traditional actors—employees, the media and industry regulators—account for a large share of fraud detection, whereas standard governance actors detect relatively little (Dyck et al. 2010); and short-sale constraints lead to systematic overpricing (Miller 1977; Diamond and Verrecchia 1987). In parallel, economics has turned to agent-based and narrative-based modelling of market dynamics (Mantegna and Stanley 2000; Bouchaud and Potters 2003; Farmer and Foley 2009; Shiller 2019), reflecting the view that aggregate outcomes can be studied by simulating heterogeneous behaviour rather than by

assuming a representative agent. This tradition treats markets as interacting particle systems whose aggregate regularities—fat tails, clustered volatility, threshold-like responses—emerge from micro-level rules rather than from a single equilibrium condition.

Shiller’s narrative economics adds that stories move prices through contagion irrespective of their truth—a mechanism agent-based models are well suited to capture (Shiller 2019). Cheap-talk theory further questions whether disclosure alone can deter unsupported claims when communication is inexpensive and enforcement is weak (Crawford and Sobel 1982; Farrell and Rabin 1996), and the structurally analogous “greenwashing” literature offers a direct reference for AI narrative decoupling (Lyon and Maxwell 2011; Delmas and Burbano 2011). Mechanism-design theory suggests that a well-designed mechanism can make truthful disclosure the more attractive choice—the goal of regulation is not to “make penalties heavier” but to change the incentives firms face; “targeted transparency” (Fung et al. 2007) and the critique of weak “comply or explain” enforcement (MacNeil and Li 2006) provide an analytical framework for labelling-type mechanisms.

What distinguishes this paper from conventional econometric methods is that it answers counterfactual questions: difference-in-differences, instrumental variables and regression discontinuity can test whether A affects B, but cannot ask how pricing shifts between fundamentals and narrative, or how regulatory rules shift that response. Drawing on Mantegna and Stanley (2000), Bouchaud and Potters (2003) and Farmer and Foley (2009), we treat investor agents as “particles”, β_D as a “response coefficient” and λ as a “control parameter”; because the model is stochastic, results are summarised by distributions across many independent runs, and relationships confounded in real markets by regulatory endogeneity can be examined under identical shocks with rules varied across runs.

1.3 Problem Analysis and Research Objective

Three gaps nevertheless remain. First, a **theoretical gap**: existing comparisons of regulatory instruments rest on static games or on single mechanisms, and none places firms, retail investors, institutional investors and a regulator inside one dynamic market in which enforcement rules are varied under a common information shock. Second, an **empirical gap**: real regulatory regimes change slowly, and enforcement is endogenous—regulators inspect firms that already look suspicious—so observational comparisons are contaminated by selection and cannot answer the counterfactual question of what would happen under a different rule. Third, an **integration gap**: behavioural finance and game theory supply rich mechanism hypotheses—recidivism escalation, targeted enforcement, whistleblowing and short selling—but no common testbed in which they can be compared under identical conditions; retail investor attention has been shown to shape stock price crash risk in China (Wen et al. 2019), yet how investor-side attention moderates regulatory effects is not captured in a unified framework.

To bridge these gaps, this paper proposes an agent-based simulation model informed by statistical-physics intuitions and tests its implications using annual-report text and financial panel data on **3,133 A-share listed firms** over 2021–2024. Under a single, identical AI-hype shock, we conduct a counterfactual comparison of nine mainstream international regulatory mechanisms together with a no-regulation

baseline, and trace how each rule reshapes the simulated relationship between narrative decoupling and firm valuation. Each cell is run at four levels of regulatory intensity $\lambda \in \{0, 0.3, 0.5, 0.7\}$ with five independent ensembles over 504 trading days, so that mechanism differences can be attributed to the rule rather than to the shock.

This paper makes four contributions. First, we **propose** a unified comparison framework: treating the pricing coefficient β_D of narrative decoupling as a response coefficient and regulatory intensity λ as a control parameter turns “whether regulation works” into a comparable question across mechanisms (§ 2–§ 3). Second, we **construct** an agent-based model of four agent types—firms, retail investors, institutional investors and a regulator—anchored in firm-level annual-report and financial data, and use it to rank mechanisms over 200 ensemble simulations (§ 3–§ 4.1). Third, we **reveal**, through component ablation and Pareto optimisation, that at an equal penalty multiplier a mechanism's structural design explains its effectiveness better than the size of the penalty multiple, and that the number of combined mechanisms is non-monotonic in average deterrence (§ 4.2–§ 4.4). Fourth, we **offer** a four-tier graded disclosure framework with a clearing exit and a practical RegTech path, presented as conditional policy implications rather than model-proven optima (§ 5).

2 Theoretical Background and Model Propositions

The four statements below are modelling assumptions and interpretive frameworks, not proved theorems; their observable implications are examined through numerical experiments. The term “phase transition” is used only in an analogical sense (a threshold-like jump in regulatory effect as intensity rises); we do not fit a power law to a critical exponent (see § 4.1 and § 5.4).

P1 (Law of large numbers → macro-measurability of bubbles): As the number of investors N grows, the bubble amplitude A_{bubble} is largely determined by the industry distribution of Talking/Doing and the regulatory intensity, up to a small finite-size correction; convergence across four market sizes can therefore be checked numerically.

P2 (Simulation-based inference): Repeated independent simulations characterize the distribution of model outcomes under the specified initial conditions and behavioural rules. Hence N independent runs can proxy for “observing one real market path for a long time” in supporting counterfactual comparison.

P3 (Threshold-like response to intensity): At low regulatory intensity λ the market shows little punishment of AI narrative decoupling ($\beta_D \approx 0$); as λ rises, the simulated β_D becomes negative in a nonlinear, threshold-like manner rather than as a sharp transition. The core experiment scans λ to locate, for each mechanism, the range over which its effect strengthens; a strict critical value is left to a finer grid.

P4 (Noise versus regulatory cost): The larger the market noise σ , the stronger the λ needed for the same effect. A high-retail, high-noise market therefore needs stronger regulation for the same effect, which helps explain why markets of different maturity may need differentiated regulatory intensity.

3 Agent-Based Model and Experimental Design

3.1 Model Construction

The model is built on Mesa with four agent types: 3,133 firm agents (carrying talking, doing, Dissonance and Tobin's Q, updating narratives on the annual-report cycle), 3,000 retail households (trading on narrative signals and noise with a 30% stop-loss), 100 institutional investors (trading on a talking-plus-doing score and able to short under certain mechanisms), and one regulator (inspecting, labelling and deducting valuation every 20 steps). β_D is estimated by an OLS regression of Tobin's Q on Dissonance with HC3 robust standard errors across all 3,133 firms; it is a model-implied response coefficient—the simulated association between Dissonance and valuation under a given rule—not a causal estimate of investor punishment. It is estimated separately in each simulated cross-section and summarised by its mean and distribution across ensembles; standard errors are HC3-robust. The bubble indicator is $A_{\text{bubble}} = \max(\text{Tobin's Q})/\text{median}(\text{Tobin's Q})$. The nine mechanisms span punitive, labelling and game-theoretic paradigms plus a no-regulation baseline (Table 1). Each runs at $\lambda \in \{0, 0.3, 0.5, 0.7\}$ (0.3 light: disclosure/inquiry; 0.7 heavy: investigation, penalty, market ban); the ablation adds $\lambda = 1.0$ to examine over-regulation.

The nine mechanisms span the major global regulatory paradigms, grouped by mode of action into punitive, labelling and game-theoretic types, alongside a no-regulation baseline (Table 1). Each is run at four levels of the regulatory-intensity parameter $\lambda \in \{0, 0.3, 0.5, 0.7\}$ (higher λ denotes stronger enforcement; 0.3 light: basic disclosure/inquiry letter; 0.7 heavy: investigation, penalty, market ban); the ablation study additionally uses $\lambda = 1.0$ to examine over-regulation.

In operational terms: A deducts 1%, 3% and 7% of valuation at $\lambda = 0.3/0.5/0.7$; F imposes $0.03 \times 2^{s-1}$ (s = cumulative violations), with the strike count reset to zero on compliance; E exposes with probability $\min(0.5, 0.1 \times D)$ and deducts 5%; D complies with 60% probability and explains (2% deduction) otherwise; I audits 5% of firms each round, deducting 8% or 3%; and G lets institutions hold short positions of $\leq 5\%$ of capital with a $\pm 20\%$ close.

Table 1
Institutional archetypes and modes of action of the nine regulatory mechanisms

Type	Mechanism	Institutional archetype / country	Mode of action
Punitive	A Blanket threshold	Command-and-control (China / EU GDPR)	Deducts valuation if $ Dissonance > \text{threshold}$
	B Targeted enforcement	Targeted supervision (US SEC)	Penalizes only $Dissonance > 0$
	C Targeted + transparency	Graded disclosure (SEC Reg S-K)	Warning $\rightarrow$ risk flag $\rightarrow$ rectification
	F Repeat-offender escalation	Recidivism (three-strikes)	$0.03 \times 2^{s-1}$, reset on compliance
Labelling	D Comply-or-explain	UK Corp. Gov. Code	60% comply / 40% explain
	E Whistleblower	US Dodd-Frank (SEC)	$\min(0.5, 0.1D)$ exposure, - 5%
	H Sunshine disclosure	US SEC Reg FD	Label only, no penalty
	I Random audit	US IRS random audit	Audits 5% of firms per round
Game-theoretic	O No regulation	Laissez-faire (control)	No inspection or penalty
	G Short-selling	SEC Reg SHO / China margin	Shorts on decoupled firms ($\leq 5\%$)

Table 2
Agents, objectives, information and decision rules.

Agent	Objective / motive	Information observed	Decision rule	Key parameters
Firm (3,133)	Operating value; listing/valuation signal	Own talking, doing, Dissonance, Tobin's Q	Update narrative on the annual-report cycle; doing evolves with investment	decay $\kappa = 0.17$; firm noise $\sigma = 0.10$
Retail investor (3,000)	Portfolio return	Narrative signal, price, position	Buy on positive narrative signal; 30% stop-loss	$\sigma_{\text{retail}} = 0.15$
Institutional investor (100)	Risk-adjusted return	Talking-plus-doing score, price	Trade on combined score; may short under G	$\sigma_{\text{inst}} = 0.03$; short $\leq 5\%$ capital
Regulator (1)	Enforce chosen rule	Dissonance, violation history	Inspect/label/deduct every 20 steps	$\lambda \in \{0, 0.3, 0.5, 0.7\}$; $h_{\text{penalty}} = 0$ for H

3.2 Experimental Design

The scenario is anchored to the A-share AI hype ignited by ChatGPT in 2023, over about two years (February 2023 to end-2024), corresponding to 504 trading days. The data base is the annual-report text and financial panel of 3,133 A-share listed firms (2021–2024). “Talking” is a text-constructed AI-narrative indicator built by NLP over annual-report text from a purpose-built dictionary (AI-keyword intensity plus disclosed AI-capital signals; method in (Loughran and McDonald 2011)); “doing” is built from AI-related hiring and patents; the two are weakly correlated ($r = 0.36$). The dictionary and coding rules were validated by two independent coders on a $\sim 10\%$ sample (Cohen’s $\kappa = 0.72\text{--}0.89$; two-week re-coding agreement 91.2%), with precision 92.4% and recall 88.7% against human-coded text. Both indices are z-scored over the pooled firm-year sample, so Dissonance = $z(\text{Talking}) - z(\text{Doing})$; valuation is Tobin’s Q. The firms span 81 CSRC industry categories. For the hype structure we make a modelling assumption rather than an empirical estimate: the top 20% by initial talking (626) are assigned genuine AI investment and the remaining 80% (2,507) merely “claim to be positioning”; firms with doing > 0 already account for about 33% in 2023, so 20% is a conservative lower bound. In all 10 settings (nine mechanisms plus baseline) $\times 4 \lambda \times 5$ ensembles = 200 simulations, an identical shock is injected on day 31; all cells share paired seed baselines, so the regulatory rule is the only design factor varied across runs. Each run covers 504 trading days with 3,133 firms and 3,100 investors; the 200 runs total about 630 million agent decisions (decision events, not independent samples).

This experiment is essentially a shock–response comparison. In all 10 settings (nine regulatory mechanisms plus a no-regulation baseline) $\times 4 \lambda$ levels $\times 5$ ensembles = 200 full simulations, an **identical** hype shock is injected on the 31st trading day: 626 (20%) “genuine-AI” firms see a slight rise in talking/doing, while 2,507 (80%) “hype-riding” firms see a surge in talking with doing unchanged. Thus “the shock itself” is not a source of differences in outcomes; the regulatory rule is the only design factor varied across runs. To keep comparisons clean, all mechanism–intensity cells share the same paired seed baselines, so that any residual differences at zero intensity reflect seed-level noise; mechanism effects are therefore measured relative to the common no-regulation baseline rather than to one another. Each simulation runs 504 trading days (about two years, covering two annual-report cycles), each iteration involves 3,133 firms and 3,100 investors, and the 200 simulations total about **630 million agent decisions** (these are decision events, not independent samples). Key parameters: retail signal noise $\sigma_{\text{retail}} = 0.15$, institutional noise $\sigma_{\text{inst}} = 0.03$, and narrative decay rate $\kappa = 0.17$.

Table 3
Data-to-model mapping.

Data / indicator	Use in the model	Detail
Annual-report AI narrative	Talking (initial state)	53-term three-tier dictionary; keyword density plus AI-capital disclosure; z-scored over pooled firm-years
AI-related hiring and patents	Doing (initial state)	Recruitment and patent counts; equal-weighted z-scores; excludes AI product revenue
Tobin's Q	Valuation state; β_D regression	Winsorised Q; OLS of Q on Dissonance with HC3 SE in each simulated cross-section
Industry / firm characteristics	Heterogeneity; robustness splits	81 CSRC categories; size and investor-structure splits
2023 AI hype	Common exogenous shock (day 31)	Scenario assumption: top 20% by initial talking treated as genuine adopters

4 Results and Robustness

4.1 Ranking of the Nine Mechanisms

Table 4 reports the ebb-period pricing coefficient β_D of the nine mechanisms at four levels of regulatory intensity (response curves in Fig. 2). At $\lambda = 0$, all nine mechanisms share the same seed-paired baseline, $\beta_D = -0.369$ (sd 0.030)—the baseline simulation produces a negative association between Dissonance and Tobin's Q, but this "natural baseline" is far from sufficient to curb hype. Regulatory effects thereby show huge dispersion:

Table 4

β_D response curves of the nine regulatory mechanisms (bracketed values in the $\lambda = 0.7$ column are 95% CIs computed as $t(4)$ intervals across the 5 ensembles; all values are simulation outputs)

Rank	Mechanism	$\lambda = 0.0$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 0.7$ [95% CI]	Effect
1	F Repeat-offender escalation	-0.369	-2.065	-2.065	-2.065 [- 2.114, - 2.017]	+++
2	B Targeted enforcement	-0.369	-0.601	-1.202	-1.804 [- 1.853, - 1.754]	++
3	C Targeted + transparency	-0.369	-0.369	-0.642	-1.471 [- 1.518, - 1.423]	+
4	A Absolute blanket threshold	-0.369	-0.535	-0.903	-1.245 [- 1.286, - 1.205]	+
5	E Whistleblower	-0.369	-0.670	-0.670	-0.670 [- 0.693, - 0.647]	$\pm$
6	D Comply-or-explain	-0.369	-0.608	-0.608	-0.608 [- 0.693, - 0.523]	$\pm$
7	I Random audit	-0.369	-0.508	-0.508	-0.508 [- 0.551, - 0.464]	$\pm$
8	H Sunshine disclosure	-0.369	-0.369	-0.369	-0.369 [- 0.406, - 0.332]	-
—	No regulation (baseline)	-0.369	-0.369	-0.369	-0.369 [- 0.406, - 0.332]	—
9	G Short-selling	-0.369	-0.369	-0.369	-0.369 [- 0.406, - 0.332]	-

Note: +++/++/+/±/- grade effect strength by $|\beta_D|$ in the ebb period; "—" marks the baseline. Values are simulation outputs; the $\lambda = 0.7$ 95% CIs use $t(4)$ across the 5 ensembles.

The results reveal three regularities. First, **regulatory effects are sharply stratified**: repeat-offender escalation and targeted enforcement form the first tier, well ahead of “targeted + transparency” and “blanket threshold”, while short selling and the label-only sunshine design coincide exactly with no regulation (zero effect under this implementation). Second, **effective mechanisms respond rapidly**: for repeat-offender escalation, β_D jumps from - 0.37 to - 2.07 at $\lambda = 0.3$ and then saturates; we do not fit a power-law scaling, and a strict critical value is left to a finer λ grid. Third, **regulation is non-monotonic**: comply-or-explain is essentially flat across λ (- 0.608 for $\lambda \geq 0.3$), showing no additional benefit from higher intensity in the tested range. (The framework and architecture are shown in Fig. 1.)

4.2 Mechanism Attribution

To identify the sources of a mechanism’s effectiveness, we perform a component ablation of repeat-offender escalation (a 2×2×4 factorial design; Table 5, Panel A). The escalating multiplier (2^{s-1}) contributes about 56% of the strength: the variant with escalation has a mean β_D of – 2.326 versus – 1.239 without it ($\Delta\beta = 1.087$), and removing it drops repeat-offender escalation from champion to the level of the blanket threshold. This corroborates Polinsky and Shavell (2000): the power comes from intertemporal deterrence—violating today doubles tomorrow’s penalty. The exponential form is also consistent with prospect theory (Kahneman and Tversky 1979), in which losses loom larger than equal gains.

By contrast, the "clearing exit" (resetting the strike count upon compliance) plays a limited role within the time horizon of this simulation (the difference in β_D with versus without clearing is only 0.000); its value lies more at the institutional level (avoiding a "permanent stigma") and needs to be verified over a longer horizon. The deeper reason is organizational inertia: clearing takes effect only after completing the full cycle "violation → repentance (Dissonance ≤ 0) → violation again", yet in the 504-day simulation almost no agent completes this cycle—once in a high-decoupling state, a firm is locked in by its narrative/behavioural strategy. **Organizational inertia therefore overwhelms institutional incentives.**

Table 5
Mechanism attribution and structural contrast (ablation experiments)

Panel A: Repeat-offender escalation (F) component attribution				
Component	With (β)	Without (β)	$\Delta\beta$	Share
Escalating $\times 2^{s-1}$	–2.326	–1.239	1.087	~ 56%
Clearing exit	–2.326	–2.326	0.000	~ 0%
Base penalty	–1.239	–0.369 (baseline)	0.870	~ 44%
Panel B: B targeted vs A blanket – penalty-multiplier ablation				
Variant	Mechanism	Multiplier	$\lambda = 0.7$	$\lambda = 1.0$
B	Targeted	1.0	–2.064	–1.979
A	Blanket	1.0	–1.428	–0.714
B_2×	Targeted	2.0	–2.282	–1.911
A_2×	Blanket	2.0	–1.490	–0.517

Note: Panels A–B are dedicated ablation batches (seeds and scale differ from Table 4), so values are not comparable to Table 4. $\Delta\beta = |\beta_{\text{With}}| - |\beta_{\text{Without}}|$; the "Base penalty" baseline is no regulation ($\beta_D = -0.369$). In Panel B, at an equal penalty multiplier the B–A gap at $\lambda = 0.7$ (0.64) exceeds the doubling gain (0.22).

4.3 Structural Comparison

The comparison between targeted enforcement and the absolute blanket threshold (Table 5, Panel B) shows that, at an equal penalty multiplier, $B > A$ at all λ and the gap widens with λ : at the extended $\lambda = 1.0$, A rebounds because of over-penalisation (β_D rises from -1.428 to -0.714) while B remains stable. The structural reason is “signal contamination”: the blanket threshold penalises |Dissonance|, so genuine implementers (negative Dissonance) and braggers (positive Dissonance) receive the same penalty; implementers are “punished no matter what they do”, and their best response is to reduce doing, blurring the distinction between real and fake AI. Targeted enforcement penalises only Dissonance > 0 , so the dominant strategy for implementers is to keep implementing—consistent with stronger incentives for compliance. Over-regulation is also visible: at $\lambda = 1.0$ firms strategically compress talking rather than raise doing, and the signal is drowned in noise, consistent with a Goodhart-type response.

Temporally, each scenario is divided into pre-injection (steps 10–30), hype (31–100) and ebb (201–504) windows. All mechanisms show very similar β differences during the hype period (-0.11 to -0.19): when hype has just been injected, the market has not yet had time to distinguish real from fake. Differentiation appears in the ebb period, when effective mechanisms keep punishing firms that “bragged but did not act”, pushing β sharply negative, whereas the label-only sunshine design and short selling coincide exactly with the no-regulation baseline (zero effect under this implementation). This “same first, different later” structure shows that regulation takes effect only as information is gradually released.

Moreover, the nine mechanisms mainly work by **altering the simulated Dissonance–valuation relationship** rather than **eliminating bubbles**: the bubble indicator A_{bubble} fluctuates between 5.9 and 6.8 across mechanisms, with little difference between them. Punitive mechanisms can thus price braggers effectively (β more negative) but hardly compress the right tail of the valuation distribution; bubble containment may require more direct macroprudential tools (trading restrictions, leverage constraints), supporting the view that different regulatory tools should each play their part.

4.4 Combination Optimization

Since any single mechanism has loopholes, we conduct a Pareto search with fast non-dominated sorting (Deb et al. 2002) over 300 candidate combinations in a three-objective space (β_D deterrence / bubble control A_{bubble} / a retail-investor cumulative P&L proxy; Figs. 3–4). “Repeat-offender escalation + targeted enforcement” (F + B) is a recurring core pair, appearing in 85.2% of the 27 non-dominated solutions; the best combination is F + B with $\beta_D = -2.125$, a further 2.9% gain over the strongest single mechanism (full-scale F, -2.065) and about 5.8× the no-regulation benchmark (ratio to the Table 4 baseline of -0.369 ; the combination batch’s own no-mechanism baseline is -0.301 , i.e. $\approx 7.1\times$). This is a change in the simulated coefficient, not in real-world violation rates. The number of mechanisms is non-monotonic in average deterrence—two are strongest, four remain effective, six or more fall back markedly—because stacking too many sources creates “regulatory congestion”. Comply-or-explain, though modest alone ($\beta_D = -0.61$), interacts positively with F and B: by mandating disclosure it supplies the information infrastructure needed to enforce the other mechanisms.

Because regulation is not costless, the three-objective space includes a welfare dimension approximated by a retail-investor cumulative P&L proxy: if highly decoupled firms are priced more efficiently, retail losses from chasing highs in the ebb period fall. On the Pareto frontier, combinations with F + B at their core are not simultaneously outperformed across the three objectives, suggesting no systematic conflict among them. This welfare characterisation must be read with caution: it relies on a single proxy and excludes compliance costs, administrative costs and institutional P&L, so it is directional rather than a quantitative estimate of net welfare (see § 5.4).

4.5 Robustness Checks

We verify the core conclusions across 178 variants (parameter sensitivity, investor structure, market size and seed level). Across narrative-decay rates, hype scenarios, four investor structures and three market sizes, F remains first and B second; seed-level tests give F a mean β_D of -2.303 (CV 1.9%) and B -1.367 (3.1%), with all seeds of the same sign; the ranking of the two leading mechanisms was preserved in all 178 variants, and even under continued “de-retailisation” the F + B core stays optimal.

To test statistical significance, we run 60 seeds for each of the ten settings at the M scale (300 firms $\times$ 504 days; 600 simulations) and conduct pairwise Welch t-tests with Holm correction (Table 6). 41 of the 45 pairwise comparisons remain significant after Holm correction (the complete 45-pair matrix, with unadjusted and Holm-adjusted p-values and effect sizes, is provided in the replication package), and the core stratification $F > B > C > A > E \approx D > I$ holds (Cohen’s $d = 2.92$ for F vs B); E and D do not differ ($p = 1.00$), while H and G coincide exactly with “no regulation” ($p = 1.00 / 1.00$; $d = 0.00$), i.e. zero effect under this implementation. The ranking at the M scale is essentially the same as at the full scale, so the stratification conclusion does not depend on simulation scale.

Table 6

Pairwise significance tests across mechanisms (M scale: 300 firms $\times$ 504 trading days, $\lambda = 0.7$, $n = 60$ simulations per mechanism; the tabulated mean β_D is the within-run post-period mean at the M scale and is therefore not directly comparable to the full-scale values in Table 4)

Rank	Mechanism	Mean β_D	95% CI	p(Holm) vs none	Cohen's d
1	F Repeat-offender escalation	-1.096	[- 1.111, - 1.082]	5.5×10^{-107}	+ 15.8
2	B Targeted enforcement	-0.938	[- 0.952, - 0.925]	1.9×10^{-98}	+ 13.6
3	C Targeted + transparency	-0.755	[- 0.768, - 0.742]	6.2×10^{-85}	+ 10.4
4	A Absolute blanket threshold	-0.671	[- 0.684, - 0.658]	2.3×10^{-77}	+ 9.0
5	E Whistleblower	-0.337	[- 0.352, - 0.323]	7.1×10^{-27}	+ 2.7
6	D Comply-or-explain	-0.324	[- 0.354, - 0.295]	1.3×10^{-11}	+ 1.5
7	I Random audit	-0.241	[- 0.257, - 0.225]	1.5×10^{-5}	+ 0.9
8	G Short-selling	-0.186	[- 0.202, - 0.171]	1.00	+ 0.0
9	H Sunshine disclosure	-0.186	[- 0.202, - 0.171]	1.00	+ 0.0
—	No regulation (baseline)	-0.186	[- 0.202, - 0.171]	—	—

Note: more negative β_D = stronger punishment; p is Holm-corrected pairwise Welch (41 of 45 significant) and, as the ensembles are not fully independent, should be read with the 95% CIs.

Excluding AI-native sectors. Because the sample is not restricted to non-AI industries, we re-run the $\lambda = 0.7$ comparison after excluding software and information-technology services (I65), which removes 312 firm entries and leaves 3,015 firms. The ranking is unchanged: repeat-offender escalation (- 2.271) > targeted enforcement (- 1.972) > targeted + transparency (- 1.582) > blanket threshold (- 1.368) > no regulation (- 0.277). The result supports reading the sample as firms that adopt AI in their existing operations, rather than as AI-product firms.

Sensitivity to the hype-structure assumption. The 20/80 split is a scenario assumption, so we vary the share of firms assigned genuine AI investment (0.1, 0.3 and 0.4) and re-run the four key settings at $\lambda = 0.7$. Table 7 reports the results. The ordering F > B > A > no regulation is preserved at every share; the absolute magnitude of F rises from - 1.943 at a 0.1 share to about - 2.11 at 0.3–0.4, and the no-

regulation baseline becomes less negative as more firms are assigned genuine adoption. The ranking conclusion is therefore robust to this assumption, while absolute levels should be read as conditional on the chosen scenario.

Table 7
Sensitivity of the $\lambda = 0.7$ results to the share of firms assigned genuine AI adoption.

Share of genuine adopters	F	B	A	No regulation
0.1	-1.943	-1.758	-1.232	-0.428
0.3	-2.113	-1.809	-1.248	-0.319
0.4	-2.114	-1.781	-1.206	-0.263

What is and is not validated. The empirical panel estimate of the Dissonance–valuation association is small and statistically indistinguishable from zero ($\beta = 0.010$, $p = 0.21$), which is consistent with the paper’s premise that market discipline is weak; the simulated no-regulation baseline ($\beta_D \approx -0.37$) is a scenario benchmark under the common hype shock, not a calibrated replication of that estimate. Accordingly, the model is used to compare mechanisms under identical assumptions, not to forecast real-world magnitudes.

5 Discussion and Conclusion

5.1 Main Conclusions

Taking 3,133 A-share listed firms as the micro foundation, this paper builds an agent-based model and compares nine regulatory mechanisms plus a no-regulation baseline under an identical AI-hype shock, yielding four conclusions. First, the market’s own punitive force is weak and far from sufficient ($\beta_D \approx -0.37$ without regulation). Second, regulatory effects are sharply stratified, with repeat-offender escalation (-2.07) and targeted enforcement (-1.80) strongest and short selling (-0.37) and the label-only sunshine design (-0.37) at the level of no regulation. Third, at an equal penalty multiplier, structural differences between mechanisms explain effectiveness better than penalty magnitude—repeat-offender escalation relies on the intertemporal deterrence of “the longer you drag it out, the more expensive it gets”, targeted enforcement on the incentive alignment of penalising only narrative decoupling. Fourth, the best combinations involve only a few mechanisms, with F + B strongest (-2.125) and 2–4 mechanisms preferred.

5.2 Policy Implications: Tiered Regulation and RegTech

Recommendation 1: move from a “blanket threshold” to targeted, tiered regulation. We propose an illustrative four-tier design for A-share AI disclosure (thresholds are illustrative institutional choices, not model-proven optima): green (Dissonance ≤ 0.5 , no intervention), yellow (> 0.5 , inquiry letter requiring an explanation of the match between narrative and actual AI inputs), orange (> 1.0 or no improvement after

inquiry, public warning, fine and valuation deduction), and red (> 1.5 with three cumulative orange episodes, refinancing restrictions and executive accountability). A fifth element—a clearing exit—lets any firm in the orange/red tier descend to green if Dissonance ≤ 0.5 for two consecutive cycles; it is included for its long-term value in avoiding a “permanent stigma”, and its short-term deterrent effect is limited (§ 4.2).

Recommendation 2: automate auditing through RegTech. Tiered regulation presupposes low-cost, high-frequency measurement of each firm’s decoupling degree. With large language models it is already feasible to scan AI narratives in annual reports, extract claimed investment amounts and cross-validate them against actual AI-related hiring and patenting. This substantially raises the efficiency of manual sampling and is the key to moving the agent-based framework from a “research tool” to “regulatory infrastructure”; in the present model the “doing” side is built from hiring and patents, while AI-capital disclosure contributes to the narrative side.

Recommendation 3: tie incentives to auditable inputs and make decoupling a disclosure indicator. Subsidies and tax breaks should be tied not to a firm’s claim of “positioning in AI” but to auditable AI inputs (e.g. AI-related hiring and patents, or audited AI-revenue share), closing the signalling-incentive loophole. Because the registration-based reform is disclosure-centred, standardising Dissonance as a disclosure indicator would disadvantage highly decoupled firms in IPO review and refinancing—not an added penalty but the natural operation of capital-market infrastructure, and hence a durable design feature. State-owned listed firms should take the lead in adopting a unified AI-disclosure standard and third-party auditing, using peer pressure to lower market-wide compliance costs.

5.3 Cross-Market Applicability of the Conclusions

The direct scope of this paper is A-share listed firms nationwide, but the relative ranking across mechanisms—repeat-offender escalation $>$ targeted enforcement $>$ pure transparency—appears stable across the industries and firm sizes examined. Three points matter for cross-market migration. First, short selling in a mature ecosystem (US hedge funds, securities lending, options) is likely to be considerably better than the negligible effect observed here; in the A-share market, securities lending and short selling remain constrained by limited lendable supply and quotas, so the zero effect is consistent with that institutional setting, and Miller (1977) and Diamond and Verrecchia (1987) apply only where short-selling infrastructure is developed. Second, the whistleblower may be far more effective under the SEC’s monetary-reward scheme (10–30% of fines); our “no monetary incentive, exposure capped at 50%” assumption corresponds to a low-incentive scenario. Third, the finding that a pure label creates little deterrence is likely to generalise, although in the US class actions may turn a label into part of a litigation evidence chain. Institutional transplantation therefore cannot be copied wholesale but must fit local investor composition, short-selling infrastructure and enforcement tools.

5.4 Limitations and Outlook

This paper has several limitations. First, each firm is characterised by only two scalars, talking and doing, whereas real AI narratives and investment are far more multidimensional. Second, short selling is

modelled as a one-off open-and-close position without dynamic adjustments or derivatives. Third, investors are relatively homogeneous, and fourth, firms do not interact strategically. Fifth, the narrative measure is a dictionary-based, text-constructed proxy rather than a trained classifier, and AI-capital disclosure saturates for many firms. Statistically, the significance of β_D comes from an OLS fit of a simulated cross-section and agents are not independently sampled, so p-values should not be read frequentistically; λ takes only four points, $\{0, 0.3, 0.5, 0.7\}$ (the ablation adds 1.0), without a finite-size scaling fit, and a strict identification of λ_c is left to a finer grid; ablation and part of the robustness checks are single-seed runs without reported uncertainty intervals; and the welfare dimension is approximated only by a retail-investor P&L proxy. Future work will test optimal mechanism combinations, anchor the pricing kernel to limit-order-book data, and develop a real-time early-warning dashboard.

Declarations

Author Contribution

M.S. conceptualised and designed the study, developed the agent-based model and the simulation code, performed the formal analysis and the numerical experiments, and wrote and revised the manuscript.

Data Availability

The study uses two categories of material. (i) Annual reports of Chinese A-share listed companies, 2021–2024. The raw texts were obtained under licence from the CSMAR and Wind commercial databases. The licence does not permit redistribution, so the raw texts are not publicly available; they can be obtained directly from these vendors, and the firm identifiers used in this study are listed in the replication package so that the extraction can be repeated by any licensee. (ii) Replication package for "Can Regulation Curb AI Adoption Narrative-Practice Decoupling? An Agent-Based Model of the A-Share Market". This package contains the derived firm-level indicators (Talking, Doing, Dissonance and controls for 3,133 firms), all simulation code, parameter files and result files, and reproduces every table and figure in the paper. It is available from the corresponding author on request during review and will be deposited in Zenodo upon acceptance (DOI to be assigned). No other data were used.

References

1. Akerlof GA (1970) The market for "lemons": quality uncertainty and the market mechanism. *Q J Econ* 84:488–500. <https://doi.org/10.2307/1879431>
2. Becker GS (1968) Crime and punishment: an economic approach. *J Polit Econ* 76:169–217. <https://doi.org/10.1086/259394>
3. Bouchaud JP, Potters M (2003) Theory of financial risk and derivative pricing. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511753893>

4. Crawford VP, Sobel J (1982) Strategic information transmission. *Econometrica* 50:1431–1451. <https://doi.org/10.2307/1913390>
5. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6:182–197. <https://doi.org/10.1109/4235.996017>
6. Delmas MA, Burbano VC (2011) The drivers of greenwashing. *Calif Manage Rev* 54:64–87. <https://doi.org/10.1525/cmr.2011.54.1.64>
7. Diamond DW, Verrecchia RE (1987) Constraints on short-selling and asset price adjustment to private information. *J financ econ* 18:277–311. [https://doi.org/10.1016/0304-405X\(87\)90042-0](https://doi.org/10.1016/0304-405X(87)90042-0)
8. Dyck A, Morse A, Zingales L (2010) Who blows the whistle on corporate fraud? *J Finance* 65:2213–2253. <https://doi.org/10.1111/j.1540-6261.2010.01614.x>
9. Fama EF (1970) Efficient capital markets: a review of theory and empirical work. *J Finance* 25:383–417. <https://doi.org/10.2307/2325486>
10. Farmer JD, Foley D (2009) The economy needs agent-based modelling. *Nature* 460:685–686. <https://doi.org/10.1038/460685a>
11. Farrell J, Rabin M (1996) Cheap talk. *J Econ Perspect* 10:103–118. <https://doi.org/10.1257/jep.10.3.103>
12. Fung A, Graham M, Weil D (2007) Full disclosure: the perils and promise of transparency. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9780521699617>
13. Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291. <https://doi.org/10.2307/1914185>
14. Loughran T, McDonald B (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *J Finance* 66:35–65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>
15. Lyon TP, Maxwell JW (2011) Greenwash: corporate environmental disclosure under threat of audit. *J Econ Manag Strategy* 20:3–41. <https://doi.org/10.1111/j.1530-9134.2010.00282.x>
16. MacNeil I, Li X (2006) “Comply or explain”: market discipline and non-compliance with the Combined Code. *Corp Gov Int Rev* 14:486–496. <https://doi.org/10.1111/j.1467-8683.2006.00524.x>
17. Mantegna RN, Stanley HE (2000) An introduction to econophysics: correlations and complexity in finance. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511755767>
18. Miller EM (1977) Risk, uncertainty, and divergence of opinion. *J Finance* 32:1151–1168. <https://doi.org/10.1111/j.1540-6261.1977.tb03317.x>
19. Polinsky AM, Shavell S (2000) The economic theory of public enforcement of law. *J Econ Lit* 38:45–76. <https://doi.org/10.1257/jel.38.1.45>
20. Shiller RJ (2019) Narrative economics: how stories go viral and drive major economic events. Princeton University Press, Princeton. <https://doi.org/10.1515/9780691189970>
21. Spence M (1973) Job market signaling. *Q J Econ* 87:355–374. <https://doi.org/10.2307/1882010>
22. Wen F, Xu L, Ouyang G, Kou G (2019) Retail investor attention and stock price crash risk: evidence from China. *Int Rev Financ Anal* 65:101376. <https://doi.org/10.1016/j.irfa.2019.101376>

Figures

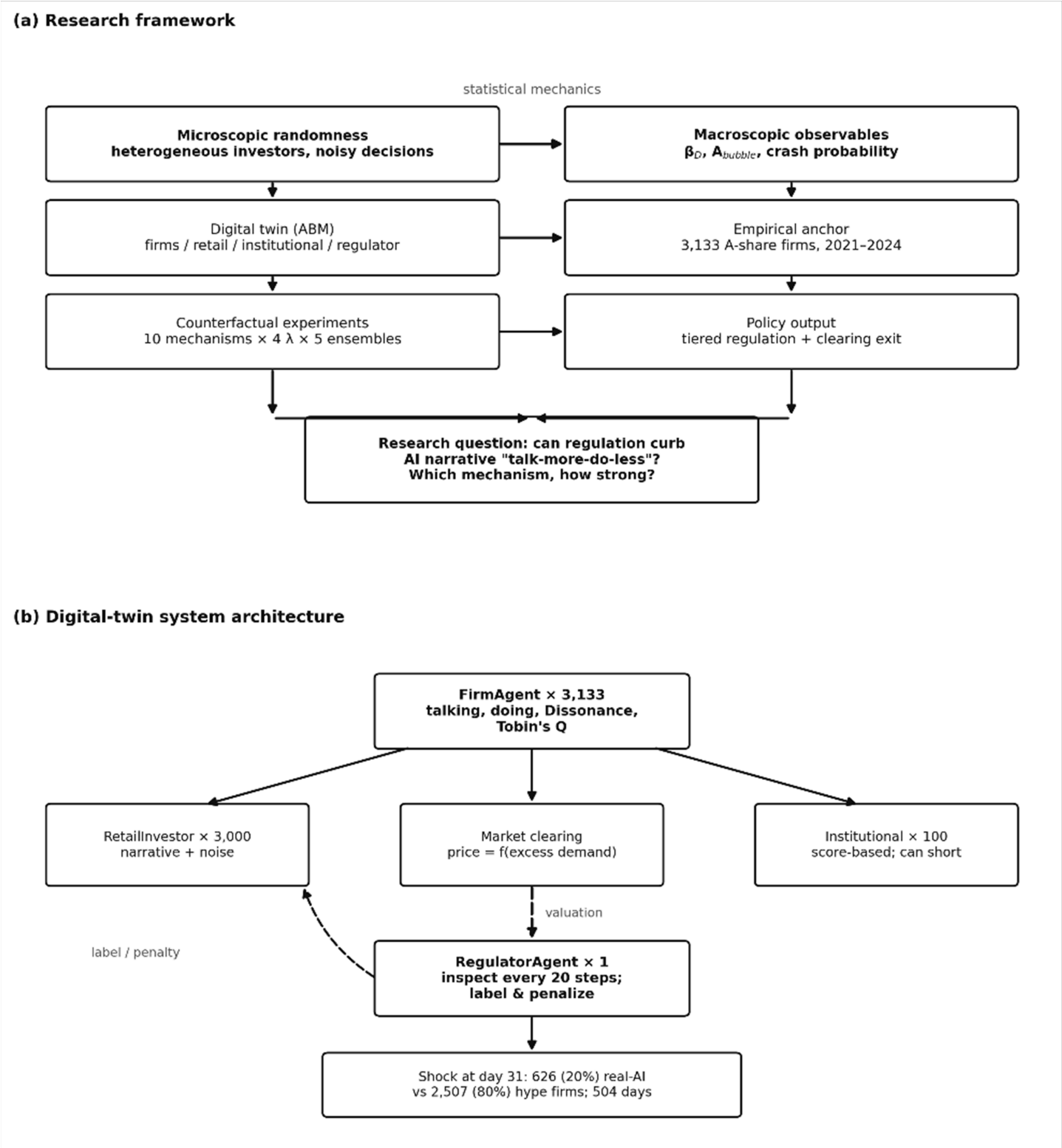


Figure 1

Research framework and agent-based system architecture

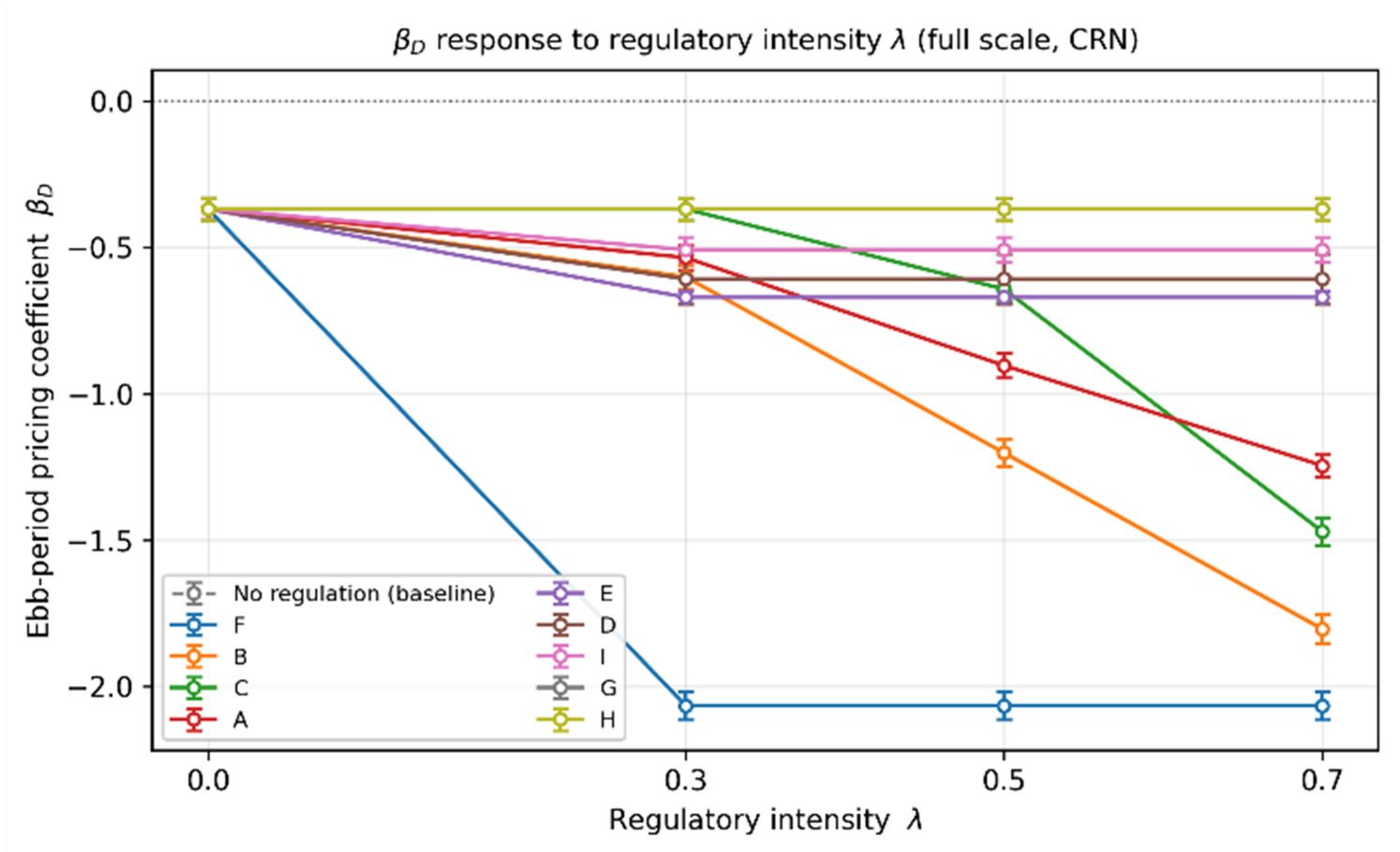


Figure 2

β_D response to regulatory intensity λ across nine mechanisms and the no-regulation baseline.

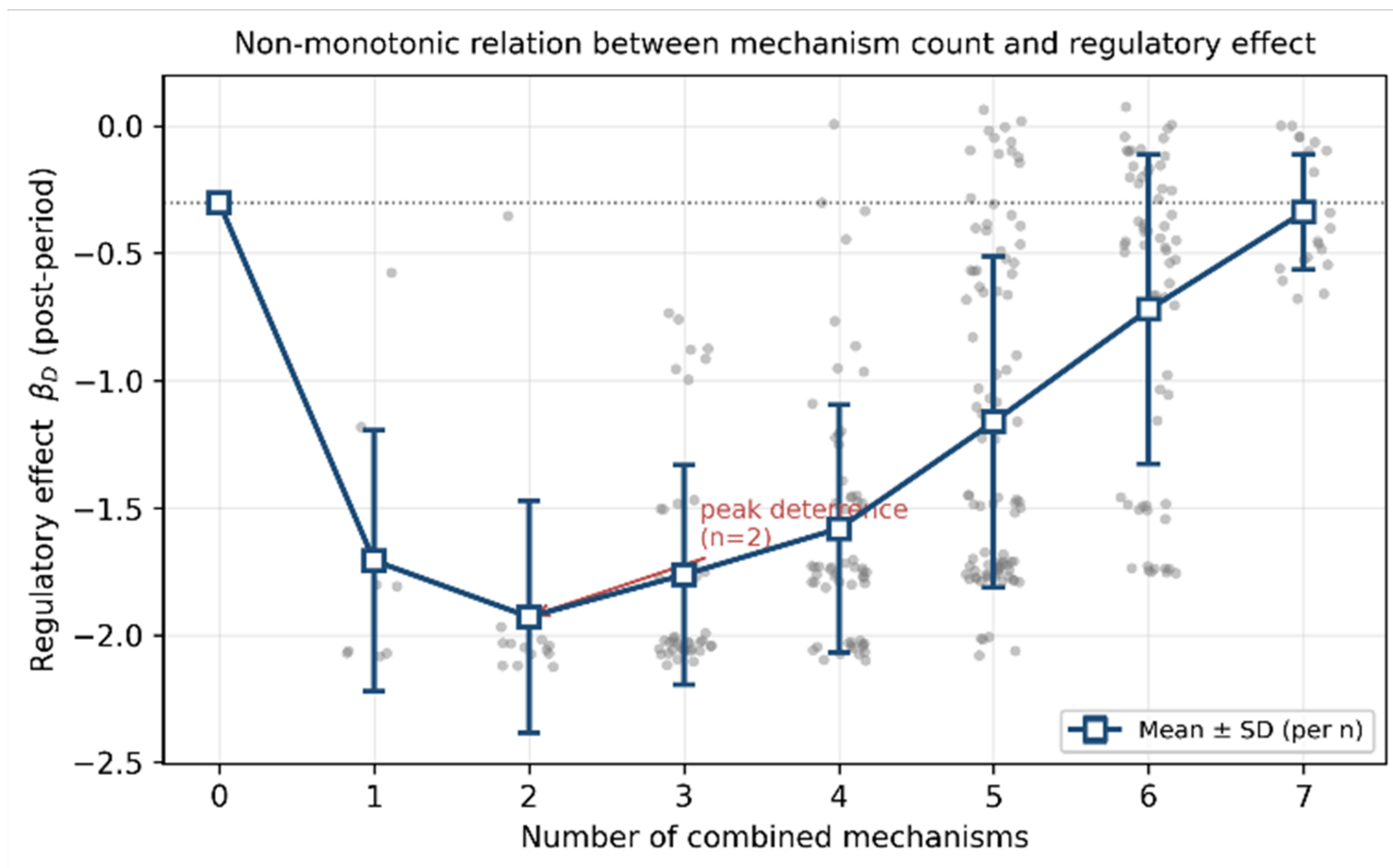


Figure 3

Non-monotonic relation between the number of mechanisms and regulatory effect.

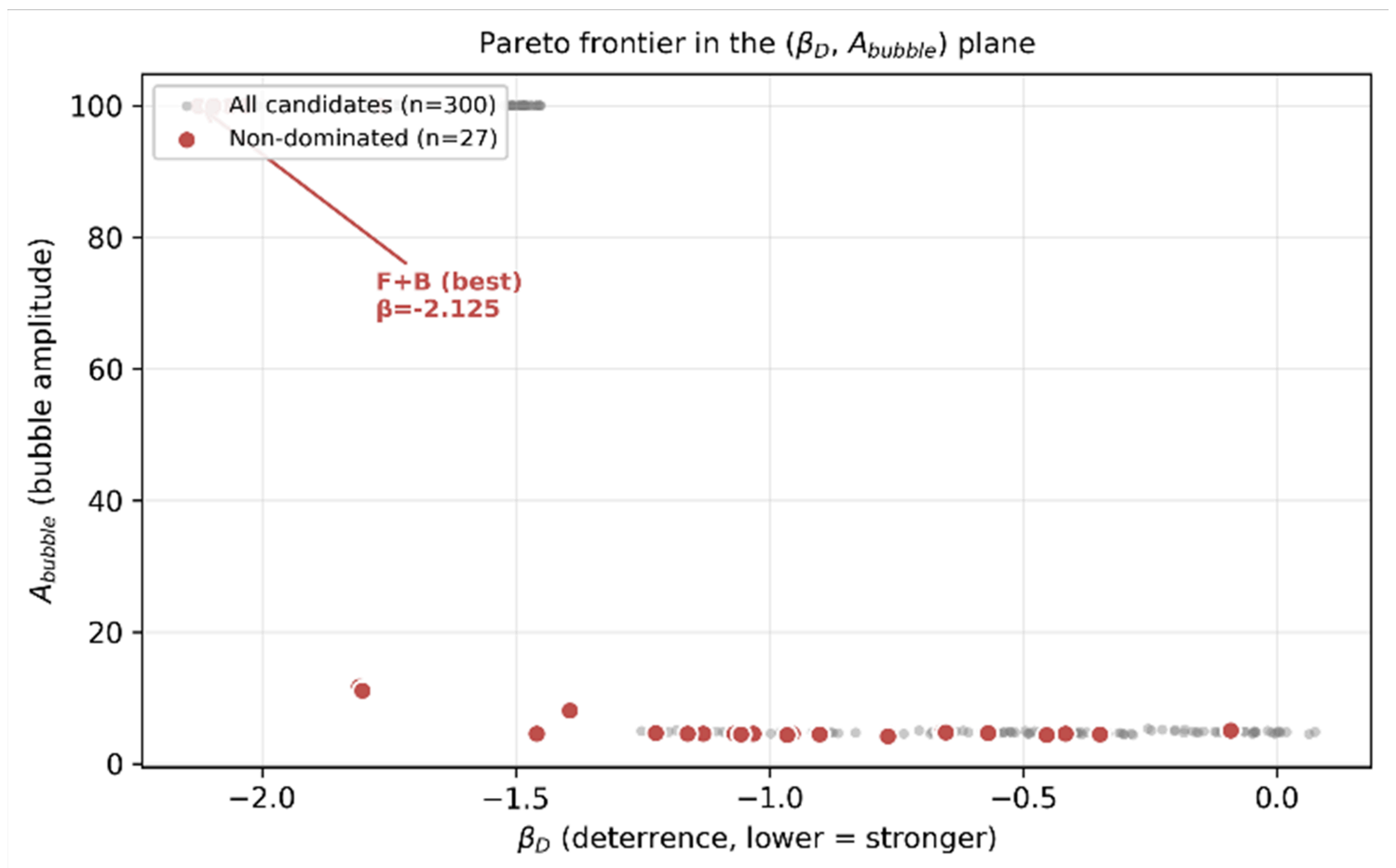


Figure 4

Pareto frontier in the (β_D, A_{bubble}) plane (27 non-dominated solutions).