The Cyber Risk of Operating AI: Evidence from Corporate Data Breaches

Fatih Kansoy*
University of Oxford

Yuhao Huo**
University of Pennsylvania

This version: November 2026

Latest version: fatih.ai/cyber_risk_operating_ai.pdf

Abstract

Most corporations adopt artificial intelligence inside daily workflows tied to personal records rather than as commercial software for outside sale. We ask whether twelve-month statutory breach notifications track internal AI operation differently from commercial AI supply. In a panel of 28,266 US corporate Form 10-K filings (2008–2025), we link human-validated multi-model classifications to public breach registries and archived regulatory letters inside industry-by-year peer groups. Our estimates show that internal AI operators experience substantially higher breach notification rates than AI-discussing peers without internal deployment across post-filing breach onsets, two-phase human calibration, and confirmed own-system compromises in regulatory letters. Commercial AI supply and a third-party vendor placebo sit near zero. Because excess notices concentrate in active Item 105 risk-factor prose, archival filings leave workflow vulnerability only partly separated from threat-informed disclosure and render every estimate a conditional association. Boards, insurers, and supervisors should scrutinise live internal AI data access and permissions rather than commercial AI credentials.

Keywords: artificial intelligence; data breaches; breach notifications; 10-K filings; risk factors; large language models; information security

*Corresponding author. University of Oxford. fatih.kansoy@economics.ox.ac.uk.
**University of Pennsylvania. martinhuo@engineering.upenn.edu.

1. Introduction

In February 2024, the payments company Klarna reported that an artificial intelligence (AI) assistant had handled two-thirds of its customer-service chats during its first month (Klarna 2024). That assistant processed refunds, returns, payment disputes, and repayment plans directly across sensitive customer accounts. Klarna built none of those models itself. Instead, the company connected an external model from OpenAI to its own databases and support operations. Most corporations now adopt AI in the same way by embedding outside-built models inside administrative and customer-facing systems that hold personal records.

Whether running AI inside daily business workflows raises or lowers corporate data-breach risk is far from obvious. On the one hand, internal operation sends customer, patient, and employee records through prompts, retrieval indexes, and logs and grants staff wider permissions across those databases (Tabassi 2023, Berente et al. 2021). A 2024 supervisory survey, conducted by the Bank of England and the Financial Conduct Authority, found that three-quarters of UK financial firms used AI and that one-third of reported tools came from external providers (Bank of England and Financial Conduct Authority 2024). Nearly half of those firms understood the models only partially. On the other hand, enterprise analytics platforms can reduce complexity-driven breaches (Tanriverdi, Kwon, and Im 2025) just as cloud providers patch flaws faster than individual buyers (August, Niculescu, and Shin 2014). Commercial AI developers face direct targeting (Kamiya et al. 2021) and software bugs that exposed user sessions even at frontier builders (OpenAI 2023). Dual-role firms might also use internal engineering to protect workflows (Cohen and Levinthal 1990, Wang et al. 2023). Knowing where breach exposure sits matters directly for boards, federal cybersecurity regulators (Securities and Exchange Commission 2023), and cyber insurers.

Are companies that operate AI inside their own workflows named more often in subsequent statutory breach notifications than peers in the same industry and filing year? Does any gap separate internal operation from commercial supply or shrink among dual-role firms with internal engineers? Public breach notices and filings reflect reporting discretion (Amir, Levi, and Livne 2018). We therefore test whether primary breach letters, strictly post-filing onset timing, and blinded human validation can separate internal workflow risk from third-party vendor pass-through, delayed reporting, selective risk drafting, and automated reading error across annual-report sections.

Measuring internal AI operation separately from commercial supply avoids raw keyword noise (Loughran and McDonald 2011) and reaches beyond hiring or single-country surveys. From 100,274 annual SEC Form 10-K filings (1994–2026), we screen 55,705 cue-bearing reports across Item 1 business descriptions, Item 1A risk factors, and post-2023 Item 1C cybersecurity governance. A primary language-model family coded all 55,705 reports while a second independently coded all 55,445 filings with business or risk-section cues, after which a third family blindly adjudicated contested main-sample cases alongside agreed controls. Every positive call requires a verbatim supporting excerpt. Our 2008–2025 main panel covers 28,266 read filings across 4,451 public US

firms with complete financial controls, dated three-digit industry codes, and closed twelve-month outcome windows matched to Privacy Rights Clearinghouse statutory breach notices (Romanosky, Telang, and Acquisti 2011, Sen and Borle 2015). Comparison peers are read AI-discussing filings rather than unread silent reports. To verify data quality, we audit all 162 notified windows from 2024 and 2025 against archived letters, obtain blinded human coding on 349 masked filings, and inspect 24 listed commercial banks for hosting visibility.

Our baseline linear probability model compares twelve-month breach notifications across AI-discussing firms inside the same three-digit industry and filing year with controls for firm size, business and risk-section lengths, digital and customer-data wording, cybersecurity text, and prior-year breaches. To separate internal exposure from general technology scale, we enter commercial AI supply in the same equation and interact both roles to test dual-role capability. We also split Item 1 business descriptions from risk sections and use masked-passage reading to distinguish active workflows from cautious disclaimers. Audited regulatory letters separate own-system compromises from vendor notices. Additional tests isolate strictly post-filing breach onsets, compare reporting lags, exclude defensive security-AI adopters, and contrast electronic intrusions against physical losses. Human-validated estimators then adjust for automated reading error alongside within-firm fixed-effects models using firms that change operating status (Mundlak 1978).

Across the full panel, we find that companies reporting internal AI operation appear substantially more often in subsequent twelve-month breach notifications than same-industry peers. That conditional gap equals roughly two-thirds of the sample baseline notification rate and stands out most strongly during the 2024–2025 adoption wave, even though the cross-era increase stays imprecise. Commercial AI development and supply show no comparable elevation. In direct role comparisons, the estimated association for commercial AI suppliers stays close to zero across both periods and separates statistically from internal operation during 2024–2025 after showing a borderline contrast across the full eighteen-year panel. Dual-role engineering moderation remains statistically unresolved across every period.

Primary regulatory letters show that breaches inside the firm's own network explain most of the verified 2024–2025 operating difference. In our audit of all 162 windows from 2024 and 2025, archived regulatory letters verify issuer identity in nearly every case and attribute roughly four-fifths of the verified operating gap to own-system compromises. Pure third-party vendor incidents do not explain the gap. The operating association sits near zero across the 26 pure vendor windows, even though that small count cannot rule out modest upstream vendor risks. Positive associations also persist when we restrict outcomes to breaches beginning strictly after the filing date, shift the observation window past ninety or one hundred and eighty days, or exclude companies that use AI for cyber defence and fraud screening. Among dated post-filing episodes, operators show no faster reporting lags. Breach notices concentrate in electronic hacking, unintended disclosure,

and insider intrusions rather than physical losses despite noisier separation from untyped state notices.

Splitting disclosures by report section shows that elevated breach incidence concentrates among firms reporting internal AI only in risk sections rather than business descriptions. Even with flexible firm-size decile controls or without post-2023 Item 1C-only disclosures, pure Item 1A risk factors drive that placement gap. Sceptics might suspect that risk-section AI text reflects mere boilerplate about future tools (Barrios et al. 2025). In blinded masked-passage reading, nearly nine in ten of these risk-section-only passages state active current workflows rather than empty disclaimers. Confirmed active statements retain a positive breach association. Under Regulation S-K Item 105 (Securities and Exchange Commission 2020), materiality rules direct companies to disclose exposed AI workflows inside Item 1A risk factors just as managers who already sense weak security controls draft sharper corporate risk prose (Wang, Kannan, and Ulmer 2013, Campbell et al. 2014, Florackis et al. 2023). Annual reports alone cannot separate workflow exposure from threat-informed drafting.

Automated reading errors do not manufacture our main association. Blinded human validation matters when model disagreements cluster in the long risk prose of large firms with higher breach rates since outcome-correlated error can bias raw regressions in either direction (Yang et al. 2018). Using resolved labels from two blinded human coders who achieve over ninety-seven per cent weighted agreement across 349 masked validation filings, two-phase regression calibration (Chen, Hong, and Tamer 2005) keeps strict operational estimates positive across both the full panel and 2024–2025. Counting generic AI risk warnings unexpectedly raises calibrated estimates because warning phrases appear widely inside automated comparison strata and coincide with elevated breach rates in contested cases. One-step orthogonal score correction (Robins, Rotnitzky, and Zhao 1994, Egami et al. 2023), which drops the surrogacy assumption of calibration through inverse inclusion weights, matches the full-panel estimate across wide confidence intervals yet loses precision sharply in 2024–2025 under heavy sampling weights. Recent error-adjusted inference therefore relies on calibration.

Separating internal AI operation from commercial supply changes how scholars and practitioners evaluate corporate cyber exposure. Prior work tracks AI talent through hiring (Babina et al. 2024), divides buyers from developers in surveys (Calvino and Fontanelli 2026), or links IT and AI choices to risk within hospitals, banks, and IT firms or through job postings and patenting (Angst et al. 2017, Kim and Kwon 2019, McLemore and Mihov 2026, Xiao et al. 2026, Srinivas, Zhang, and Hong 2026, Rishabh, Mihet, and Jang-Jaccard 2026). Across US industries, pairing overlapping roles with verified own-system breaches redirects attention towards internal workflows rather than outbound products. Our placement and validation findings also clarify Item 1A partial identification and show why applied text studies need downstream calibration when warning definitions shift estimates (Ludwig, Mullainathan, and Rambachan 2026). Boards should focus oversight on internal AI connectors, permissions, and logs rather than assuming commercial engineering protects

live records. Cyber underwriters, who face unobservable security effort (Cezar, Cavusoglu, and Raghunathan 2014), should use operational disclosures for targeted diligence rather than rigid surcharges because 2025 forecasts improve risk ranking modestly at a fixed date while probabilities drift during rapid adoption. Public filings also omit hosting visibility for supervisors. None of our 24 audited commercial banks disclosed whether customer-facing AI ran in-house or on external servers.

As a limitation of this analysis, our observational estimates capture conditional associations rather than causal parameters. Item 105 disclosure discretion, persistent customer-record holdings left unresolved by underpowered within-firm switching models, and the concentration of operating filings in 2024–2025 (Carter, Schnepel, and Steigerwald 2017) prevent causal identification without exogenous deployment shocks or internal telemetry. Section 2 develops competing hypotheses from prior research. In Sections 3 and 4, we describe the Form 10-K reading pipeline, human validation, breach records, and empirical specifications. Sections 5 and 6 present the empirical findings, validation and out-of-sample checks, and implications for governance, cyber insurance, and future research.

2. Related literature and hypotheses

Empirical work on corporate cybersecurity shows that digital expansion raises breach risk unless technical expertise and strong controls keep pace. Larger firms with wider digital footprints face higher exposure (Sen and Borle 2015, Kamiya et al. 2021). Yet new enterprise software pushes security in opposite directions across organisational boundaries since electronic health records roughly tripled accidental hospital breaches (Kim and Kwon 2019) while enterprise analytics platforms cut complex failures across multi-hospital health systems (Tanriverdi, Kwon, and Im 2025). Across public firms, higher IT innovativeness likewise raises breach incidence unless senior executives hold technical skill (Wang et al. 2023). Breach rates instead fall when organisations invest in proactive rather than symbolic security controls (Kwon and Johnson 2014, Angst et al. 2017) alongside security-aware IT budgets, domain-expert chief information officers, and financial-audit spillovers (Li, Leung, and Yue 2023, Haislip, Lim, and Pinsker 2021, Liu 2025). Because existing studies track firm-wide IT intensity or isolated security functions, they leave open whether engineering skill gained from commercial AI products protects internal corporate deployments.

Observed breach records and annual reports pass through managerial filters before researchers see them. State breach-notice mandates cut identity theft by roughly 6 per cent (Romanosky, Telang, and Acquisti 2011) and sharply lowered recorded privacy breaches at firms with sensitive consumer records (Kim, Wang, and Wu 2024). Even under mandatory laws, managers frequently underreport cyberattacks (Amir, Levi, and Livne 2018), shift breach notices onto busy news days to mute equity losses (Foerderer and Schuetz 2022), and exercise timing discretion over statutory disclosures (Lin and She 2025). Annual Form 10-K risk factors, required under federal securities rules, add a second

disclosure filter because cybersecurity risk text conveys forward-looking hazards that predict future breach announcements (Wang, Kannan, and Ulmer 2013, Campbell et al. 2014, Florackis et al. 2023). In corporate AI disclosures, companies with heavy AI talk and sparse technical talent likewise gain neither efficiency nor patents unless they outsource (Barrios et al. 2025). Disclosed AI-breach links therefore blend exposure with managerial reporting and drafting choices.

Most empirical studies of corporate AI infer technology use from workforce skills or national surveys rather than workflow roles inside annual reports. Across labour studies, measures of AI-skilled employees and online job vacancies link technical talent to product innovation and firm growth (Acemoglu et al. 2022, Babina et al. 2024). Labour proxies do not cleanly divide internal operators from commercial vendors. By contrast, enterprise surveys put early US production AI below 6 per cent (McElheran et al. 2024), count roughly three French AI buyers per developer (Calvino and Fontanelli 2026), and tie German product updates to AI users and world-first inventions to developers (Rammer, Fernández, and Czarnitzki 2022). Those enterprise surveys examine productivity and innovation rather than corporate data breaches.

Recent studies link firm-level AI to cyber risk with conflicting conclusions across proxies. Elevated operational losses and breach rates accompany larger AI worker shares at major US banks (McLemore and Mihov 2026), substantive AI deployment rather than pre-adoption announcements at IT firms (Xiao et al. 2026), and automation-focused rather than augmentation-focused job advertisements (Srinivas, Zhang, and Hong 2026). Broader AI talent instead aids recovery after natural disasters (Han et al. 2025). Measures of post-ChatGPT workforce task exposure likewise associate with lower corporate breach likelihood (Wang et al. 2026). Under heightened cyber risk, external AI adopters slow patent output while frontier builders stay undeterred (Rishabh, Mihet, and Jang-Jaccard 2026). No cross-industry panel tests whether subsequent twelve-month breach notices track internal AI operation rather than commercial AI development and supply.

Live AI inside daily workflows places customer records, access rights, and staff under a different control regime from that of commercial software development. Governance standards distinguish organisations that operate live models on protected records from commercial model distributors (Tabassi 2023) because autonomous, continuously updated systems inside live workflows create distinct managerial control hazards (Berente et al. 2021). At the same time, breach gaps can arise without new AI vulnerability if operators are larger technology targets, revise risk text after threat alerts, detect incidents more often, report prior attacks late, or hold larger customer databases. Table 1 maps seven testable accounts across these competing forces.

Under an operational-exposure account, internal AI elevates corporate breach risk through three custodian-specific channels absent from outbound software sales. Daily use feeds personal records into training corpora, retrieval indexes, prompts, and logs. When firms connect AI models to internal databases, attackers gain broader API pathways, wider staff permissions, and new entry points vulnerable to credential theft and indirect prompt injection (Greshake et al. 2023). External model APIs and cloud hosts introduce interdependent security hazards across corporate boundaries

(Kunreuther and Heal 2003, Tabassi 2023) through correlated software flaws (Chen, Kataria, and Krishnan 2011), moral hazard in vendor protection effort (Cezar, Cavusoglu, and Raghunathan 2014), and upstream compromises that trigger client breach notices (Crosignani, Macchiavelli, and Silva 2023). Even when external vendors patch software promptly (August, Niculescu, and Shin 2014), legal duties to notify affected consumers remain on the enterprise custodian.

Hypothesis 1 (Operational AI Exposure). *Conditional on three-digit industry-by-year shocks, firm size, filing text, and prior breach history, reported internal AI operation associates positively with subsequent twelve-month data-breach notifications, even when restricted to episodes that begin after the filing date.*

Commercial AI development and supply offer a direct within-report benchmark because outbound software builders, unlike internal operators, share technical standing and AI vocabulary without routine internal use of their own consumer records. Pure outbound sales predict zero rise in breach notices. In practice, commercial vendors that host multi-tenant client records on shared servers can attract buyer breach notices even when vendor patches cut routine exposures (August, Niculescu, and Shin 2014). That hosting ambiguity leaves standalone commercial AI supply theoretically unsigned, even though custodian exposure still predicts a stronger breach association for internal AI operation than for commercial AI supply. By contrast, an AI-firm scale account (Kamiya et al. 2021) treats all AI-discussing firms as prime technology targets and predicts positive associations of similar size across both roles.

Hypothesis 2a (Role Asymmetry). *Under custodian-exposure logic, internal AI operation carries a stronger positive association with subsequent breach notifications than commercial AI development or supply, even when vendor client hosting makes standalone commercial supply non-zero.*

In-house AI engineering could instead insulate dual-role companies that both develop AI products commercially and operate AI internally. Model builders gain absorptive capacity to spot technical failure modes (Cohen and Levinthal 1990). Because software knowledge travels poorly across firm boundaries (von Hippel 1994), internal engineers can strengthen vendor oversight (Mayer and Salomon 2006), soften the breach risk of innovative IT (Wang et al. 2023), and spread defensive controls across corporate networks (Liu 2025). That protective offset works only if corporate leaders reward security alongside rapid product release (Gordon and Loeb 2002, Anderson and Moore 2006) and place internal AI engineers inside daily enterprise workflow teams rather than isolated commercial product units (Bresnahan, Brynjolfsson, and Hitt 2002).

Hypothesis 2b (Dual-Role Capability Moderation). *Under cross-unit capability spillover, commercial AI development or supply weakens the breach association of internal AI operation among dual-role firms. Organisational separation between product units and internal workflows or larger enterprise AI footprints produces zero offset.*

Breach onset dates, defensive security software, and notice delays separate operational exposure from pre-filing timing and detection intensity. Public registries record breaches only after discovery and delay (Romanosky, Telang, and Acquisti 2011, Amir, Levi, and Livne 2018, Foerderer and Schuetz

2022, Lin and She 2025). If operators merely report pre-filing attacks late or adopt defensive cyber-AI tools after an intrusion, any positive cross-firm association disappears once pre-filing breach onsets and defensive security-AI adopters are removed from the estimation sample. A separate reporting-intensity rival, grounded in partial-identification bounds (Manski 1990, Horowitz and Manski 1995), holds actual breach rates fixed and assumes that AI operators detect or disclose identical attacks more often. Under higher detection or reporting intensity, faster internal discovery shortens onset-to-notification lags while observed notices rise indiscriminately across breach types rather than selectively in cyber intrusions over untyped or physical losses.

Federal Form 10-K disclosure rules create a sharp institutional test of where material operational AI hazards appear in corporate reports. Under Regulation S-K Item 105 and post-2023 federal cybersecurity governance rules (Securities and Exchange Commission 2020, 2023), statutory materiality standards direct consequential operational AI hazards out of general Item 1 business descriptions and into mandatory Item 1A corporate risk factors. That materiality sorting predicts stronger risk-section elevation, active present-tense operational prose, and selectivity towards cyber intrusions over untyped or physical losses. Threat-informed risk prose produces that exact same placement and cyber-type pattern when managers warn investors about weak internal controls (Wang, Kannan, and Ulmer 2013, Campbell et al. 2014, Dyer, Lang, and Stice-Lawrence 2017, Florackis et al. 2023). By contrast, hypothetical AI buzzwords cluster in generic disclaimers (Barrios et al. 2025).

Hypothesis 3 (Materiality Sorting and Breach-Mechanism Specificity). *Where internal AI operation reflects material workflow exposure, the association concentrates in risk-section-only disclosures above business descriptions, holds across active operational statements, and favours cyber, disclosure, and insider intrusions over untyped notices without shorter reporting lags.*

Persistent holdings of personal customer records, which change slowly over time, create a final confounder because large consumer databases raise both the commercial return to internal AI use (Eisfeldt, Schubert, and Zhang 2023) and a corporation's value as a cyber target (Kamiya et al. 2021). Because within-firm comparisons across reported AI status changes absorb fixed personal-record holdings and stable corporate disclosure habits (Mundlak 1978), estimates should fall near zero if stable cross-sectional differences explain every breach gap. Yet within-firm regressions also remove persistent cross-firm variation, magnify classical classification attenuation (Aigner 1973, Griliches and Hausman 1986), and leave outcome-related error unsigned (Yang et al. 2018) as short late-panel horizons widen statistical noise. Sections 3 and 4 translate Table 1 into our empirical tests.

3. Data and measurement

Annual SEC Form 10-K filings supply longitudinal role records to test whether internal AI operation elevates corporate breach risk beyond commercial supply across US industries. Because

TABLE 1. Discriminating Empirical Predictions Across Maintained and Rival Accounts

	AI-role accounts		Selection, disclosure, and reporting rivals				
Observable empirical comparison	Operational exposure	Capability spillover	AI-firm scale	Disclosure selection	Reporting intensity	Pre-filing cyber AI	Data stock
Panel A: Core AI-role comparisons and capability moderation							
Internal AI operation (Hypothesis 1)	+	+	+	+/0[a]	+	+	+
Commercial AI supply	0/?[b]	?/≤0	+	0	0/?	0	0
Operation minus supply (Hypothesis 2a)	+	+	0	+/0[a]	+/?	0	+
Dual-role moderation (Hypothesis 2b)	0/≥0	−	0/+	0	0	0	0
Panel B: Breach timing, pathways, and reporting diagnostics							
Post-filing onsets, delayed windows	+	+	+	+[a]	+[c]	0	+
Without security or fraud AI	+	+	+	+[a]	0/+[c]	0	+
Electronic vs. untyped notices (Hypothesis 3)	+	+	0	+	0	0	+
Operator onset-to-notification lag	0	0	0	0	−	+	0
Panel C: Disclosure placement and within-firm variation							
Risk-section-only above Item 1 (Hypothesis 3)	+	+	0	+	0	0	0/+
Within-firm status changes	+[d]	+	0	0/+	+/0	0	0

Notes: Predicted signs (+ positive, − negative, 0 zero, ? unsigned) hold conditional on three-digit SIC-by-year fixed effects, firm size, business and risk-section lengths, digital, customer-data, and cybersecurity text (Wang, Kannan, and Ulmer 2013, Jamilov, Rey, and Tahoun 2026), and prior-year breach notices. AI-firm scale reflects targeting of technology-intensive filers regardless of role. Pre-filing cyber AI predicts higher notices only before exclusion of pre-filing onsets and defensive security AI. Risk-section elevation above Item 1 with active operational prose fits both Regulation S-K Item 105 materiality (Securities and Exchange Commission 2020) and threat-informed risk drafting.

[a] Under disclosure selection, threat-informed risk prose predicts positive signs unless cybersecurity controls absorb threat text (Wang, Kannan, and Ulmer 2013). Purely hypothetical AI disclaimers predict zero (Barrios et al. 2025).

[b] Zero under pure outbound software supply. Unsigned (?) when multi-tenant hosting attracts client breach attributions while vendor patches lower routine exposures (August, Niculescu, and Shin 2014).

[c] Under reporting intensity, faster discovery attenuates delayed-window estimates by shifting notices earlier. Excluding security-AI adopters yields zero if higher detection rests solely on defensive AI tools (Manski 1990, Horowitz and Manski 1995).

[d] Within-firm switching estimates attenuate towards zero under classical classification error (Aigner 1973, Griliches and Hausman 1986) and carry unsigned bias under outcome-related error (Yang et al. 2018).

federal securities rules govern Item 1 business descriptions, Item 1A material risk factors, and post-2023 Item 1C cybersecurity governance (Securities and Exchange Commission 2023), statutory materiality standards lead firms to report active artificial intelligence (AI) deployments rather than unannounced pilots. We assemble 100,274 Form 10-K reports from 1994 through 2026 (84,037 through 2023 and 16,237 in 2024–2026) in line with corporate-text practice (Loughran and McDonald 2016, Hoberg and Phillips 2016, Gentzkow, Kelly, and Taddy 2019, Hassan et al. 2019). Simple keyword counts misclassify boilerplate mentions outside active business use (Loughran and McDonald 2011, Frankel, Jennings, and Lee 2022). Our first stage therefore uses cue matching to route 55,705 cue-matched filings into passage reading (55,445 in Item 1 or Item 1A plus 260 in Item 1C alone) and leave 44,569 unread. An expanded second screen of the initial 44,656 no-cue reports sent 87 candidates to our primary reader and found zero AI roles (Appendix B.1). Comparison peers are read AI-discussing filings rather than silent reports.

Under the criteria in Table 2 and Appendix B.1, structured text reading (Ash and Hansen 2023, Dell 2025) assigns two non-exclusive binary indicators to each screened report. Internal AI operation (A_{if}) equals one when Item 1, Item 1A, or Item 1C states that firm i currently uses AI or named machine learning in internal workflows or live customer services. Generic automation, future plans, pilots, and third-party AI never qualify. Commercial AI development or supply (S_{if}) equals one when a report establishes that the firm builds or sells AI products or services. External tool buyers and generic cloud tenants do not qualify. Every positive call requires a verbatim excerpt of at most 40 words. While genuinely ambiguous passages stay coded unclear outside main regressions (Appendix B.2), Panel B of Table 2 shows how verbatim proof divides active deployment statements from conditional third-party AI warnings. We record the exact disclosure section because Item 1A risk prose reflects perceived threats and persists across years (Campbell et al. 2014, Dyer, Lang, and Stice-Lawrence 2017, Cohen, Malloy, and Nguyen 2020). Separate Item 1C tags preserve comparability across the 2023 rule change (Securities and Exchange Commission 2023).

To guard against artefacts from a single language model, three separate model families evaluate consistency across the archive. Our primary model (`OpenAI GPT-6`) coded all 55,705 cue-matched filings. A second family (`Microsoft-Decision-1`) independently read all 55,445 Item 1 and Item 1A filings with 96.5 per cent agreement on operation ($\kappa = 0.814$) and 96.8 per cent on supply ($\kappa = 0.823$). Raw agreement moves from 98.4 per cent before 2024 to 90.2 per cent in 2024–2026 as corporate AI prose grows denser while chance-corrected κ holds steady at 0.783 and 0.780 (Artstein and Poesio 2008, Gilardi, Alizadeh, and Kubli 2023, Ziems et al. 2024, Appendix B.2). Outside 114 Item 1C-only extractions, the two readers agree on 27,384 of the 28,152 dual-read main-sample filings and contest 768. Blind adjudication by `Anthropic Claude` resolves 480 contested cases (381 aligned with our primary reader), confirms 77 of 80 agreed controls, and repeats 57 of 63 same-day calls.

Automated disagreements do not behave like random noise (Aigner 1973). Dual-read agreement reaches 98.7 per cent across comparison filings and 87.6 per cent across business-section operators.

TABLE 2. Classification Rules and Illustrative Excerpts from Form 10-K Filings

Category or section	Classification rule or verbatim filing excerpt	Assigned class
Panel A: Operational and supply classification criteria applied to retrieved filing passages		
Internal AI operation	Passages state that the issuer currently uses AI in its own workflows or in a live service it runs for customers, even when an external vendor built the model. Named AI methods, including machine learning and neural networks, qualify. Generic analytics, algorithms, or automation do not.	Operating ($A_{if} = 1$)
Excluded from operation	Building, training, or testing an AI product. Selling or licensing software that customers run themselves. Plans, pilots, or hypothetical risks. AI use by customers, suppliers, competitors, or attackers. Industry background or discontinued historical use.	Comparison ($A_{if} = 0$)
AI development or supply	Passages establish ongoing or completed company development of AI models or commercial supply of AI products or services. Buying or licensing external AI tools, or leasing generic cloud compute, does not qualify.	Supply ($S_{if} = 1$)[a]
Evidence requirement	Every positive classification requires a verbatim supporting excerpt from Item 1, Item 1A, or Item 1C. Statements of active current use qualify wherever disclosed. Genuinely ambiguous passages are coded unclear and excluded.	Verbatim quote
Panel B: Verbatim excerpts from 2024–2025 filings classified as operating by the primary reader		
Business section (Item 1)	"we have introduced functionality based on generative AI to improve the user experience on our website"	States operation (S)
Risk factors only (Item 1A)	"We use, and may continue to expand our use of, machine learning and artificial intelligence ("AI") technologies to deliver our services and operate our business."	States operation (S)
Risk factors only (Item 1A)	"We or our third-party vendors, clients or counterparties may develop or incorporate AI technology in certain business processes, services or products."	Mentions AI only (M)[b]

Notes: Panel A summarises the decision rules for retrieved passages from the business description (Item 1), risk factors (Item 1A), and post-2023 cybersecurity governance disclosures (Item 1C). All readers rely only on text dated on or before the filing date. The primary language-model family coded all 55,705 cue-matched filings. A second family independently coded all 55,445 filings with Item 1 or Item 1A cues. A third family then blindly adjudicated all 768 contested main-sample filings alongside 80 agreed controls. Panel B displays three 2024–2025 filings classified as operating by the primary reader with AI passages confined to a single filing item. The third column reports the blinded masked-passage classification after issuer names are replaced with "the Company" and dates and identifiers are removed.

[a] Internal AI operation (A_{if}) and commercial AI development or supply (S_{if}) are coded separately and can both equal one when a firm runs its own AI-enabled service or both builds and uses AI internally.

[b] Illustrates one of the 47 risk-section-only operating filings whose excerpt the masked passage audit classifies as mentioning AI only rather than stating active current deployment (Section 5.1).

Across risk-section-only operating filings, agreement falls to 69.0 per cent (Appendix B). Primary-only contested filings also belong to larger, more frequently notified issuers (7.60 mean log assets and 7.67 per cent breach incidence) than second-only filings (6.68 and 3.44 per cent). Because outcome-correlated error can bias raw regressions in either direction, we conduct a stratified 349-filing blinded two-coder human audit alongside multi-model checks (Yang et al. 2018, Fong and Tyler 2021, Ludwig, Mullainathan, and Rambachan 2026) and calibrated validation estimators (Robins, Rotnitzky, and Zhao 1994, Chen, Hong, and Tamer 2005, Egami et al. 2023) evaluated in Sections 4 and 5.3. Design-weighted human agreement across those 349 filings reaches 97.66 per cent on five categories ($\kappa = 0.9016$) and 99.13 per cent on strict binary operation ($\kappa = 0.9319$). Against resolved human labels, our primary strict classifier achieves 77.48 per cent positive predictive value, 99.89 per cent negative predictive value, 98.51 per cent sensitivity, and 97.91 per cent specificity.

Two bounded diagnostics test whether pretrained language models draw on entity memory (Sarkar and Vafa 2024, Glasserman and Lin 2024, Lopez-Lira, Tang, and Zhu 2025) and whether public filings reveal where AI models run. Across 40 masked filings with firm names replaced by "the Company" and dates and identifiers stripped, blind third-reader agreement stays at 82.5 per cent across three classes and 90.0 per cent on binary operation. Those masked runs (33 and 36 of 40) show zero reversals between yes and not established and sit near the 57 of 63 unmasked repeat rate (Appendix B.3). When research assistants audited annual and supplemental filings for 24 listed commercial banks (Appendix G), 118 of 120 fields matched source text even though none of the 24 banks disclosed whether customer-facing AI ran in-house or on outside servers. Public filings hide hosting location. That 24-bank audit maps a visibility limit rather than sample-wide accuracy.

External statutory breach notices discipline corporate self-reports with dated regulatory filings (Romanosky, Telang, and Acquisti 2011, Sen and Borle 2015, Edwards, Hofmeyr, and Forrest 2016, Romanosky 2016, Amir, Levi, and Livne 2018, Kamiya et al. 2021). From the 5 August 2026 Privacy Rights Clearinghouse snapshot for notices through 31 July 2026, we aggregate 103,366 raw rows across 20 state agencies and the US Department of Health and Human Services into 39,134 source groups. Entity resolution and 26 source-verified additions map those groups into 2,241 corporate episodes across 1,056 firms under attribution rules that separate notifying entities, compromised third-party providers, affected clients, and SEC registrants (Appendix A.2). Binary outcome Y_{if} equals one when an attributed episode's report interval falls wholly within twelve calendar months after filing date d_f. Episodes across that boundary and windows past 6 July 2026 stay unresolved rather than zero (Appendix A.2). That rule yields 867 notified windows across 465 firms in the main sample (3.07 per cent) and 869 in the full sample (report dates alone add two boundary cases). Across 1,426 dated archive rows, the median breach-to-report lag stands at 77 days (78 days across 537 sample-linked 2015–2021 episodes).

Across all 162 notified 2024–2025 windows, manual inspection of archived regulatory PDFs and state registry records verifies issuer identity in 157 windows (96.9 per cent, Appendix A.2). Own-system compromises (82) outnumber pure vendor notices (26) by over three to one. Because the

remaining 49 verified windows come from tabular state rows without narrative detail, this primary-letter audit allows Section 5.2 to test whether recent breach gaps stem from corporate-network compromises rather than upstream supplier pass-throughs.

Public agency files record an incident only when a firm detects it, notifies under statutory personal-data rules, posts publicly, and enters the archive before 5 August 2026. Non-personal outages and undisclosed attacks stay unseen. Outside our audited 2024–2025 cohort, historical notifier links cannot separate every direct breach from client pass-through notices. Regulatory letters never identify AI as the entry point. Subsidiaries that operate under distinct trade names can also escape company name matching. Proportional underreporting shrinks percentage-point gaps yet preserves rate ratios. Because higher detection or faster notice by AI operators would instead push estimates upwards, our onset, lag, and sensitivity checks test that reporting channel directly.

TABLE 3. Sample Coverage, Breach Rates, and Filing Characteristics by Period and AI Role

	2008–2023		2024–2025		All years	
Variable	Comparison	Operating	Comparison	Operating	Comparison	Operating
Panel A: Sample coverage and 12-month breach notifications						
Observations (filings)	21,869	890	3,956	1,551	25,825	2,441
Unique firms[a]	3,653	424	2,283	1,005	4,155	1,113
Filings followed by a breach notice	650	55	75	87	725	142
Breach notification rate (%)	2.97	6.18	1.90	5.61	2.81	5.82
Panel B: AI supply overlap, prior breaches, and firm size						
AI development or supply (%)	2.27	77.87	6.93	59.38	2.98	66.12
Prior 12-month breach notice (%)	2.71	4.49	2.68	6.00	2.70	5.45
Log total assets (mean)	6.90	7.16	6.78	7.57	6.88	7.42
Log total assets (s.d.)	2.21	2.28	2.38	2.32	2.24	2.31
Panel C: Operating filings by where AI operation is disclosed						
Business section (Item 1)[b]	–	782	–	889	–	1,671
Risk sections only (Items 1A/1C)[c]	–	108	–	662	–	770

Notes: Summary statistics for the main estimation sample of 28,266 Form 10-K filings across 4,451 public firms from 2008 to 2025 with complete covariates, a closed twelve-month post-filing breach window, and a dated three-digit SIC code. Filings from 2025 (submitted through 2 July 2025) use 2025–2026 notifications recorded through 6 July 2026 to complete their twelve-month windows. Filings from 2026 lack a closed window and are excluded. Operating filings establish active internal AI use ($A_{if} = 1$). Comparison filings contain a retrieved AI cue that was read without establishing internal operation ($A_{if} = 0$).

[a] Firm counts are distinct issuers within each cell and do not sum across periods or role columns because firms can file in both eras and change operating status over time.

[b] Includes filings establishing operation in Item 1 alone or jointly in Item 1 and the risk sections.

[c] Comprises 699 filings with operational evidence in Item 1A alone, 58 in Item 1C alone, and 13 in both Item 1A and Item 1C.

Sequential filters narrow the 55,705 cue-matched filings to 29,692 candidates from 2008–2025 with closed twelve-month windows through 6 July 2026, definitive operation labels, Compustat links, and full covariates (Table 3 and Appendix A.1). That cutoff omits 4,192 cue-matched 2026

filings whose windows reach 2027. Iterative removal of firm and year singletons forms the 28,536-filing full sample across 4,484 firms, after which a dated historical three-digit SIC rule yields the 28,266-filing main sample across 4,451 firms. In every panel, log total assets and historical industry codes draw from the latest Compustat record dated within 730 days before d_f (Appendix A.1). Six disclosure and history controls join firm size through log Item 1 and Item 1A words, standardised digital, customer-data, and Item 1A cybersecurity wording outside Item 1C, and a prior-year breach indicator. Balance-sheet assets and text wording proxy customer-record volume only imperfectly.

Across the main sample, 5.82 per cent of operating filings are followed within twelve months by a breach notice against 2.81 per cent of comparison filings (Table 3 and Figure 1). That gap widens in 2024–2025 as comparison incidence falls to 1.90 per cent. Across all years, roughly two-thirds of operating filings (66.1 per cent) also report commercial AI supply. Disclosures of internal AI operation jump by roughly three-quarters after 2023 from 890 filings across 2008–2023 to 1,551 in 2024–2025 alone. Risk-section-only filings account for most of that recent surge because 662 of all 770 such disclosures appear in 2024–2025. The comparison breach rate drops from 3.95 per cent in 2023 to 1.45 per cent in 2025 as wider AI entry reshapes the comparison pool. Because snapshot recording delays (1.7 to 9.2 per cent) and 2023 industry or asset-fifth reweighting explain little of that decline (Appendix D.3), these compositional shifts motivate the industry-by-year, era, placement, and validation designs in Section 4.

4. Empirical design

Our baseline linear probability model compares twelve-month breach notifications across reported AI roles inside three-digit industry-by-year cells

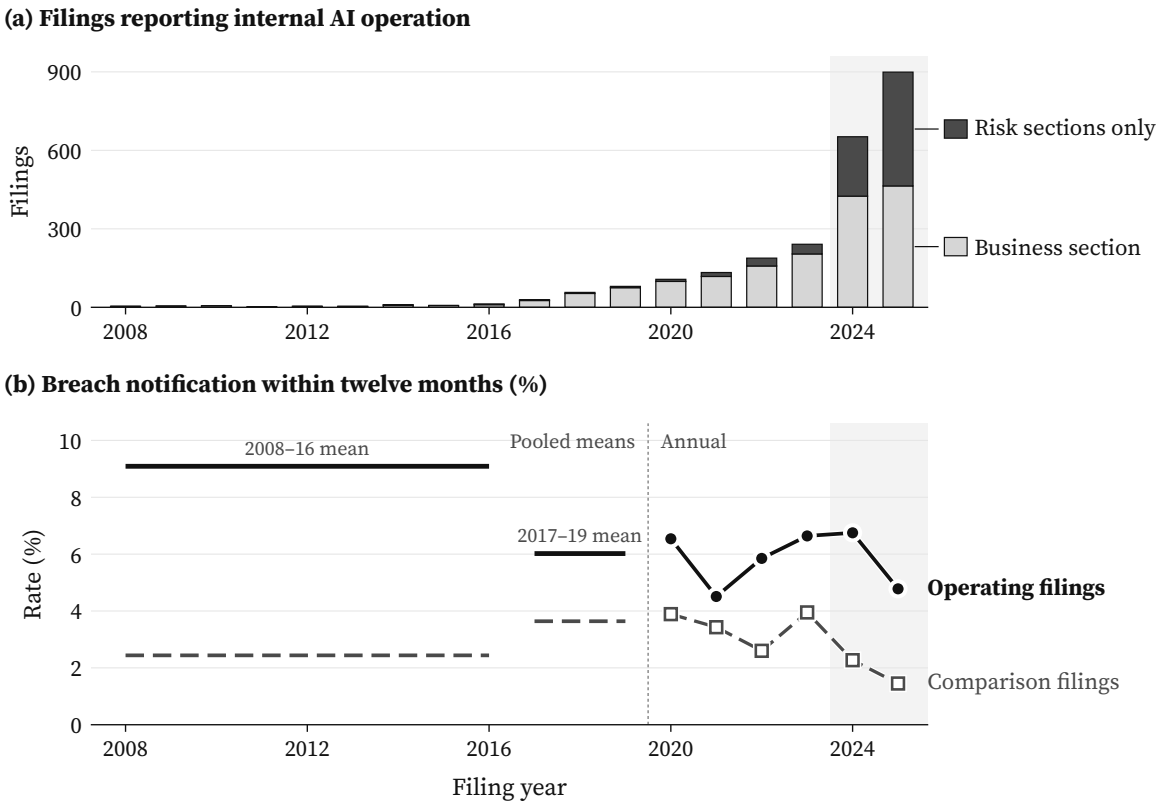
$$100\,Y_{if} = \beta A_{if} + \theta S_{if} + X_{if}'\gamma + \delta_{j(i,f),\tau(f)} + \varepsilon_{if}. \tag{1}$$

Here $Y_{if} \in \{0,1\}$ is the twelve-month notification indicator of Section 3 scaled into percentage points. Indicators A_{if} and S_{if} record internal AI operation and commercial AI supply. Vector X_{if} holds the seven scale, disclosure-text, and prior-breach controls of Section 3. Each absorbed cell $\delta_{j(i,f),\tau(f)}$ interacts filing year $\tau(f)$ with the registrant's dated three-digit SIC code $j(i,f)$. Baseline slope β therefore identifies across same-industry, same-year peers. Because absorbed singleton cells add zero information to within-cell slopes and fixed-effects logit estimators suffer incidental-parameter bias and rare-event separation (Angrist and Pischke 2009, Greene 2004, Beck 2020), our linear probability model retains all 28,266 main-panel filings.

Two companion models isolate within-firm status switches and post-2023 calendar shifts. Across the full sample of 28,536 filings, firm fixed effects α_i and filing-year effects $\lambda_{\tau(f)}$ replace industry-by-year cells in

$$100\,Y_{if} = \beta_W A_{if} + \theta_W S_{if} + X_{if}'\gamma_W + \alpha_i + \lambda_{\tau(f)} + u_{if}. \tag{2}$$

FIGURE 1. Calendar support for the operating comparison



Notes: 28,266 filings. Panel (a) plots filings whose evidence of current AI operation includes the business section (lower light-grey bars) or lies only in the risk sections (upper dark-grey bars), by filing year. Panel (b) plots the share of filings followed within twelve months by a public notification naming the firm for operating filings (solid black) and comparison filings (dashed dark grey). Years with fewer than 100 operating filings are pooled into multi-year means (2008–2016 and 2017–2019, horizontal plateaus) before annual series begin in 2020. Sources: SEC Form 10-K filings and Privacy Rights Clearinghouse.

Within-firm slope β_W draws identification from 820 firms that switch operating status across filings. All 28,536 observations still inform $\lambda_{\tau(f)}$ and γ_W. To estimate era slopes jointly, we interact $Z_{if} = (A_{if}, S_{if}, X'_{if})'$ with period indicators E_f for 2008–2023 and T_f for 2024–2025 in one equation

$$100\, Y_{if} = E_f Z'_{if} b_E + T_f Z'_{if} b_T + \delta_{j(i,f),\tau(f)} + v_{if}. \tag{3}$$

We estimate b_E and b_T together across the 890 pre-2024 and 1,551 post-2023 operating filings so that $\beta_T - \beta_E$ captures cross-period covariance inside shared firms. Separate subsample regressions omit that within-firm correlation.

To test disclosure placement across report items, a fourth equation splits A_{if} into 1,671 business-section (B_{if}) and 770 risk-section-only (R_{if}) operating filings in

$$100\, Y_{if} = \beta_B B_{if} + \beta_R R_{if} + \theta S_{if} + X'_{if}\gamma + \delta_{j(i,f),\tau(f)} + e_{if}. \tag{4}$$

An extended specification from our blinded passage audit divides R_{if} into 664 active statements R_{if}^{S} (β_S) and 106 mentions or unclear passages R_{if}^{MU} (β_{MU}). A separate specification adds nine asset-decile indicators to linear log assets. Equations (1)–(4) map directly onto the empirical rows of Table 1 under Hypothesis 1 ($\beta > 0$) and Hypothesis 2a ($\beta - \theta > 0$). We leave θ unsigned because unobservable vendor hosting can attract client breach notices and compress $\beta - \theta$ even when internal workflow risk stays real. An interaction term $A_{if}S_{if}$ tests Hypothesis 2b ($\eta < 0$) on whether commercial engineering capability lowers operational risk net of larger AI footprints. Inside (4), Hypothesis 3 draws on Regulation S-K Item 105 (Securities and Exchange Commission 2020) to predict higher risk-section slopes than business-section slopes ($\beta_R - \beta_B > 0$) and positive elevation on active statements ($\beta_S > 0$). Active prose cannot separate workflow exposure from threat-informed disclosure (Wang, Kannan, and Ulmer 2013, Florackis et al. 2023).

Across the rival accounts in Table 1, contrasts between β_W and β evaluate persistent personal-record holdings (Mundlak 1978). Sparse switches and misdated transitions still amplify attenuation under classical error and leave outcome-related error unsigned (Aigner 1973, Griliches and Hausman 1986, Yang et al. 2018). Additional onset and exclusion checks drop security-AI adopters (Appendix D.1) and isolate strictly post-filing onsets and shifted windows (Section 5.2). Pathway audits confirm that primary notices reflect electronic cyber mechanisms (Section 5.2). Manski partial-identification bounds separately quantify how much higher reporting sensitivity would need to be to explain the gap (Manski 1990, Horowitz and Manski 1995). Across Section 5.3 and Appendix E, frontier checks add event-time leads, weak-IV tests (Staiger and Stock 1997, Lee et al. 2022), concentrated exponential QMLE (Wooldridge 1999, Santos Silva and Tenreyro 2006, Correia, Guimarães, and Zylkin 2020), entropy balancing (Hainmueller 2012), proportional selection (Oster 2019), and 2025 out-of-sample forecast evaluations.

Two validation estimators address outcome-correlated error across our thirteen-stratum human sample (Section 5.3). Under two-phase regression calibration (Chen, Hong, and Tamer 2005), design-weighted human stratum means replace automated role indicators across the main frame without conditioning on breach outcomes. Consistency requires conditional outcome surrogacy given true roles and controls. Because simple stratum means also approximate stratum-level conditional expectations, a covariate-conditional check relaxes that stratum-mean assumption directly. One-step orthogonal score correction (Robins, Rotnitzky, and Zhao 1994, Egami et al. 2023) starts from residualised calibrated roles, adjusts the score with an inverse-probability Jacobian, and adds weighted human-minus-calibrated residuals. On the positive-support frame of 28,158 filings, that score adjustment drops surrogacy at the cost of higher variance when inverse inclusion weights are large (Appendix B.4).

Baseline inference clusters standard errors across 4,451 main-sample firms and 4,484 full-sample firms (Abadie et al. 2023). We add finite-sample degree-of-freedom adjustments and connected-firm checks for common ownership or shared incidents (Appendix C.2). Because operating filings concentrate in 2024–2025, standard asymptotic year clustering relies on few effective

calendar clusters (Cameron and Miller 2015, Carter, Schnepel, and Steigerwald 2017, MacKinnon, Nielsen, and Webb 2023). Section 5.3 and Appendix E therefore report two-way clustering (Cameron, Gelbach, and Miller 2011), six-point Webb wild cluster bootstraps by filing year (Cameron, Gelbach, and Miller 2008, Roodman et al. 2019), and within-cell firm-persistent permutation tests. Human-validated intervals take the 2.5th and 97.5th percentiles across 1,000 joint Phase-1 firm-cluster and Phase-2 stratified resamples. Appendix C.2 adjusts our main hypothesis family for step-down multiplicity (Holm 1979).

5. Results

5.1 Main results and disclosure placement

Internal AI operators appear far more often in subsequent statutory breach notices than industry peers in the same year. In our baseline three-digit industry-by-year model, internal AI operation associates with 1.99 percentage points higher twelve-month notification incidence, equal to roughly two-thirds of the 3.07 per cent sample mean (Table 4, column 3). Holm adjustment confirms that gap ($p = 0.009$, Appendix C.2). Mixed-cell raw incidence more than doubles peer rates. Across columns (1) and (2), prior-breach and cybersecurity controls trim the full-sample slope by roughly one-fifth. Because comparison filings still mention AI keywords without active internal use, every estimate captures a conditional gap among AI-discussing firms rather than a contrast against silent nonusers.

Commercial AI supply shows no elevation. Inside the same baseline regression, its conditional slope sits at -0.23 percentage points and stays statistically indistinguishable from zero (Table 4, column 3). Direct evaluation of Hypothesis 2a yields a pooled operation-minus-supply contrast of 2.23 percentage points ($p = 0.053$, Appendix C.2) that remains borderline at the five per cent threshold under strict human calibration (Table 6). Pooled estimates align with custodian exposure without settling population-wide separation across all eighteen years. As Section 3 notes, annual reports rarely disclose whether commercial AI vendors host buyer records on their own servers. That visibility limit can compress the role contrast yet leave internal workflow risk intact.

Calendar splits show that precision rests on recent filings. Estimates in Figure 2 and Appendix Table A14 move from an imprecise 1.84 percentage points across 890 pre-2024 operating filings to 2.55 points in 2024–2025, when operation exceeds commercial supply by 3.05 points. Wide pre-2024 intervals leave the shift inconclusive.

Dual-role engineering moderation stays unresolved. Raw averages in Appendix Table A19 show that dual-role firms incur lower breach incidence than pure operators. Once controls and fixed effects enter, conditional interaction estimates under Hypothesis 2b span zero both across the full panel (-1.55 percentage points) and in 2024–2025 (-0.12 percentage points). Sparse pure-operator counts and unobserved internal security budgets prevent firm conclusions on whether in-house commercial engineering offsets operational risk.

TABLE 4. Internal AI Operation and Subsequent Corporate Breach Notifications

	Dependent variable: Breach notification in $(0, 12]$ months ($\times 100$, pp)						
	Full sample: Broad FE		Main sample: Industry × year FE				Within firm
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Internal AI operation	2.61***	2.02***	1.99***				0.74
	(0.75)	(0.63)	(0.70)				(1.00)
Internal AI operation by disclosure location							
Business section (Item 1)				0.97	0.93	0.92	
				(0.84)	(0.84)	(0.79)	
Risk sections only (Items 1A/1C)				3.39***			
				(1.01)			
States operation (S)					2.92***	2.63**	
					(1.08)	(1.07)	
Mentions / unclear (M/U)					6.20**	5.33*	
					(3.04)	(3.02)	
AI development or supply	−0.31	−0.07	−0.23	0.19	0.23	0.17	0.51
	(0.63)	(0.55)	(0.60)	(0.64)	(0.64)	(0.63)	(0.89)
Prior 12-month breach notice		24.55***	21.97***	21.93***	21.93***	20.41***	−2.99
		(2.85)	(2.76)	(2.76)	(2.75)	(2.46)	(1.87)
Cybersecurity discussion (s.d.)		0.23	0.14	0.13	0.13	0.05	−0.07
		(0.18)	(0.19)	(0.19)	(0.19)	(0.20)	(0.26)
Firm size (log total assets)	1.41***	1.04***	1.12***	1.12***	1.12***	3.80***	0.50**
	(0.15)	(0.09)	(0.11)	(0.11)	(0.11)	(0.88)	(0.24)
Text-length & digital/data controls[a]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Asset-size decile indicators[b]	No	No	No	No	No	Yes	No
Broad industry & year fixed effects	Yes	Yes	No	No	No	No	No
SIC3 × year fixed effects	No	No	Yes	Yes	Yes	Yes	No
Firm & year fixed effects	No	No	No	No	No	No	Yes
Observations (filings)	28,536	28,536	28,266	28,266	28,266	28,266	28,536
Firms (clusters)	4,484	4,484	4,451	4,451	4,451	4,451	4,484
Mean of dependent variable (%)	3.05	3.05	3.07	3.07	3.07	3.07	3.05
Adjusted R^2	0.042	0.097	0.131	0.131	0.131	0.141	0.195

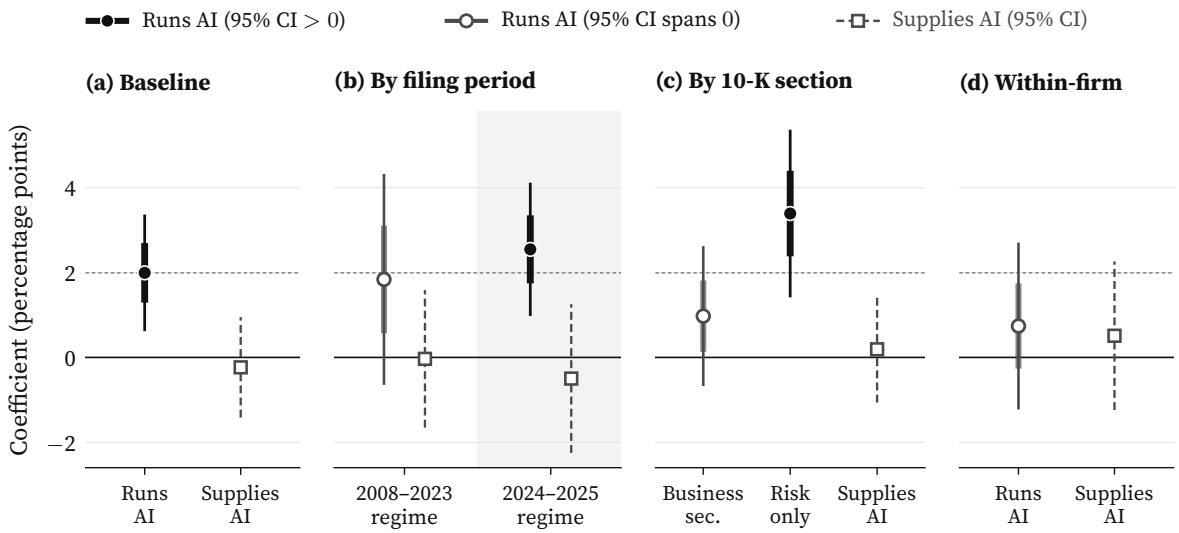
Notes: Linear probability models estimating whether a regulatory data-breach notification names the firm in the twelve months following the Form 10-K filing date ($100 \times Y_{if}$). Firm-clustered (CR1) standard errors are reported in parentheses (t-distribution with $G - 1$ degrees of freedom). The full sample comprises all 28,536 filings with complete controls after dropping singletons in column (7); the main sample comprises the 28,266 filings with a historical three-digit SIC code (3,357 three-digit SIC × filing-year cells). Column (4) partitions operating filings into Item 1 business-section disclosures (1,671 filings) and risk-section-only disclosures (770 filings). Column (5) splits risk-section-only filings using the blinded passage audit into confirmed statements of current operation (664 filings) and AI mention-only or unclear excerpts (106 filings). Across the 531 mixed three-digit SIC × year cells containing both operating and comparison filings, raw operating notification incidence is 5.59% versus 2.38% for cell-reweighted comparison peers. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (two-sided).

[a] Every specification controls for log Item 1 business words, log Item 1A risk-factor words, standardised digital-technology wording, and standardised customer-data wording (full covariate slopes in Appendix Table A13).

[b] Adds nine indicators for sample deciles of log total assets alongside linear log total assets.

Inside Form 10-K (Table 4, column 4, and Figure 2), business-section operation in Item 1 yields an imprecise slope of 0.97 percentage points against 3.39 percentage points for risk-section-only disclosures. That 2.42 percentage-point placement gap excludes zero under firm clustering (Appendix D.2). One-way filing-year bootstrap intervals remain wider. Across risk subsections, pure Item 1A risk factors drive the elevation at 3.59 percentage points against an inconclusive Item 1C estimate. That split rules out a pure artefact of the 2023 federal cybersecurity rule. When we omit filings whose AI cues appear only in Item 1C, the baseline estimate stays virtually unchanged (Appendix Table A18).

FIGURE 2. Conditional breach associations for operating and supplying AI across specifications



Notes: Point estimates with ±1 standard-error (thick inner) and 95 per cent (outer) firm-clustered intervals, in percentage points of the twelve-month notification rate, from Table 4 and Table A14. Filled black circles denote operating coefficients whose 95 per cent intervals exclude zero, open circles denote operating coefficients whose intervals span zero, and open squares with dashed stems denote commercial AI development or supply coefficients estimated in the same regression. Panels (a) to (c) absorb three-digit industry-by-year effects and all seven controls across 28,266 filings (column 3 of Table 4, the period model of Table A14, and column 4 of Table 4); panel (d) absorbs firm and year fixed effects across 28,536 filings (column 7 of Table 4). The horizontal dashed line marks the baseline operating estimate of 1.99 percentage points.

One natural objection holds that risk-factor AI text reflects empty disclaimers about hypothetical future tools rather than active systems (Barrios et al. 2025). In Appendix B.3, blinded masked-passage checks show that 86.2 per cent of risk-section-only operating passages describe active current AI workflows (S) rather than pure boilerplate. A same-family check confirms those calls. In column (5) of Table 4, confirmed present-tense risk statements carry a positive slope of 2.92 percentage points pooled and 3.60 points in 2024–2025 (Appendix Table A20). Mention-only or unclear excerpts (M/U) yield a noisy estimate that does not separate reliably from confirmed statements. Automated present-tense prose still cannot establish human ground truth or isolate technical architecture from risk-management wording.

Larger corporations attract more attackers (Kamiya et al. 2021). Nearly a quarter (24.5 per cent) of risk-section-only operators sit in the top size decile. In column (6) of Table 4, nine asset-decile indicators join linear log assets to test whether nonlinear size effects drive that elevation. Confirmed active statements (S) stay positive at 2.63 percentage points pooled and 3.50 points in 2024–2025. Across size quintiles, raw notification rates preserve that rank order inside the largest fifth of firms (Appendix Table A12). Size controls still leave one boundary. Risk-section concentration in active prose fits Regulation S-K Item 105 materiality screening (Securities and Exchange Commission 2020) just as well as threat-informed drafting (Wang, Kannan, and Ulmer 2013, Florackis et al. 2023) by managers who sharpen risk disclosures after they sense weak controls.

5.2 Incident pathways, breach mechanisms, and timing

Public chronology rows cannot prove direct compromise (Romanosky 2016, Crosignani, Macchiavelli, and Silva 2023). A matched issuer may have suffered an internal intrusion or appeared only as an affected client of a breached outside supplier. To trace exact incident pathways, we audited all 162 notified 2024–2025 windows against archived regulatory letters and state registry records (Section 3 and Appendix Table A5). Primary sources confirm that issuer or subsidiary identity matches in 157 of the 162 windows (96.9 per cent) once false homonyms and spurious vendor links are removed. Across those verified windows, the operating slope stands at 2.33 percentage points (Table 5, Panel A). Confirmed compromises inside the issuer's own systems span 82 windows and produce 1.85 percentage points, equal to 79.5 per cent of that verified slope. Forty-nine narrative-free state-table rows add 0.59 percentage points. The small negative vendor placebo reconciles those two slices exactly to the verified total (Table 5 notes). Because state tables omit incident prose, that 79.5 per cent share tracks a linear slope decomposition rather than a strict bound on all own-system breaches.

Vendor propagation does not explain these notices. Although the 26 pure third-party vendor incidents in row 5 of Table 5 yield a slope near zero (−0.11 percentage points, negligible role contrast), that small sample cannot rule out modest upstream vendor risks. Estimates stay virtually unchanged once upstream vendor links are excluded or multi-window episodes are deduplicated across both eras (Table 5, rows 3 and 7). Under the joint filter of confirmed own-system compromise and strictly post-filing breach onset in row 6, the recent slope stays positive at 1.04 percentage points across 51 verified windows. The enterprise custodian still governs permissions, credentials, and logs inside its own perimeter even when an outside vendor supplies part of an AI workflow (August, Niculescu, and Shin 2014).

Because breach categories differ sharply in baseline incidence, we normalise each category slope by its own mean to test mechanism specificity under Hypothesis 3 (Romanosky 2016). Combined electronic hacking, unintended disclosure, and insider intrusions (Table 5, row 9) produce a pooled slope of 1.82 percentage points, equal to 1.10 times the category mean. External hacking stays positive in both eras. Against pure physical or card loss (−0.26 times its mean), that normalised

TABLE 5. Primary-Source Attribution, Incident Pathways, and Breach-Mechanism Regressions

	Pooled 2008–2025 ($N = 28{,}266$)			Recent 2024–2025 ($N_{\text{recent}} = 5{,}507$)		
Outcome slice	Windows	Operation (β)	Op. − Supply	Windows	Operation (β)	Op. − Supply
Panel A: Source-audited 2024–2025 entity and pathway slices and full-panel attribution checks						
1. Baseline all linked episodes	867 142/725	1.99*** (0.70)	2.23* (1.15)	162 87/75	2.55*** (0.80)	3.05** (1.47)
2. Entity-verified (drop 11 false links)	–	–	–	157 83/74	2.33*** (0.78)	2.69* (1.44)
3. Entity-verified, no vendor links[a]	855 140/715	2.01*** (0.69)	2.32** (1.12)	131 73/58	2.44*** (0.73)	2.68** (1.35)
4. Confirmed own-system compromises	–	–	–	82 48/34	1.85*** (0.58)	2.09** (1.05)
5. Pure third-party vendor placebo	–	–	–	26 10/16	−0.11 (0.30)	0.01 (0.54)
6. Confirmed own-system, onset $> d_f$[b]	–	–	–	51 31/20	1.04** (0.48)	1.39 (0.86)
7. Deduplicate multi-window episodes	861 141/720	2.02*** (0.70)	2.31** (1.15)	157 86/71	2.66*** (0.79)	3.30** (1.46)
Panel B: Privacy Rights Clearinghouse breach mechanisms						
8. External hacking or malware (HACK)	351 79/272	1.33*** (0.50)	1.32 (0.80)	97 50/47	1.61** (0.63)	1.79 (1.10)
9. Cyber, disclosure, or insider intrusion	467 100/367	1.82*** (0.62)	2.01** (1.02)	114 63/51	2.09*** (0.67)	2.19* (1.22)
10. Exclude physical or card loss	755 137/618	2.10*** (0.69)	2.38** (1.16)	161 86/75	2.58*** (0.80)	3.08** (1.47)
11. Untyped state notices (UNKN)	288 37/251	0.28 (0.37)	0.37 (0.65)	47 23/24	0.49 (0.48)	0.90 (0.93)

Notes: Linear probability models ($100 \times Y_{if}$, percentage points) with three-digit SIC-by-year fixed effects and baseline controls across 28,266 filings (4,451 firms) from equations (1) and (3). Firm-clustered (CR1) standard errors are in parentheses; smaller figures beneath Windows report operating/comparison notified window counts (op./comp.). Tabular-only state-verified rows (49 windows; 25/24) yield 0.59 (0.48), so $1.85 + 0.59 - 0.11 = 2.33$ across own-system, tabular-only, and pure vendor rows. Stacked $\hat{\beta}/\bar{Y}$ equals 1.10 on cyber intrusions (row 9) versus 0.27 on untyped notices (row 11; $\Delta = 0.83$, two-sided $p = 0.131$, one-sided $p = 0.065$) and -0.26 on pure physical loss ($\Delta = 1.36$, $p = 0.002$). Conditional reporting-sensitivity ratios equal $2.22\times$ pooled and $2.41\times$ recent on cyber intrusions ($1.69\times$ and $2.15\times$ on all notifications). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[a] Pooled column drops upstream-flagged vendor links (855 windows; raw upstream-excluded 156 recent windows give 2.54^{***} (0.79)); recent column drops all 26 pure and three mixed source-audited vendor windows (131 windows).

[b] Reflects two corrected state discovery-date entries (51 windows); uncorrected archive dates (53 windows; 32/21) yield 1.05^{**} (0.48) and contrast 1.32 (0.87). All 87 entity-verified strict post-onset windows yield 1.16^{*} (0.60).

cyber contrast is sharp ($p = 0.002$). Untyped state notices (UNKN) often conceal unlabelled electronic breaches and yield a 0.27-times ratio ($p = 0.131$) that gives Hypothesis 3 partial mechanism support. Superior internal detection or disclosure remains a rival account because firms frequently withhold undetected attacks (Amir, Levi, and Livne 2018). Under equal true incidence in the Table 5 notes, Manski partial-identification bounds show that operator reporting sensitivity would need to exceed comparison peers by factors of 2.22 pooled and 2.41 in 2024–2025 (Manski 1990, Horowitz and Manski 1995).

Prior-onset breaches cannot explain the gap. Once we drop every undated or prior-onset window, the strict post-filing filter leaves a positive slope of 1.04 percentage points across 431 pooled windows and 1.26 points in 2024–2025 (Appendix Table A23, Panel A). That filtered slope preserves the same proportional scale against its lower base rate as the full baseline. In both eras, delayed observation windows that discard notices filed within ninety or one hundred and eighty days of the annual report retain undated state notices and remain positive. Annual-report dates lag initial deployment. Onset filters therefore exclude reverse causation from earlier intrusions only among episodes with recorded start dates.

Among timing rivals, managerial discretion between breach discovery and public notice (Amir, Levi, and Livne 2018, Foerderer and Schuetz 2022) could bunch faster disclosures inside twelve months even without higher breach incidence. Median onset-to-report lags across dated episodes after filing in Panel B of Appendix Table A23 run longer rather than shorter for AI operators (69 versus 50 days pooled and 67 versus 58 days in 2024–2025). Those lags weigh against faster public notice. Undated rows and true detection rates lie outside the archive. Attacked firms might also adopt defensive AI. Under both core and broader cyber-defence and fraud-screening exclusion or recoding rules, Appendix Table A18 confirms that the baseline slope holds between 1.65 and 1.94 percentage points. Undisclosed internal security tools remain unobserved.

5.3 Human validation, within-firm evidence, and scope

Model disagreements clustered inside the dense risk-factor prose of large firms that generate the most breach notices. Because outcome-correlated misclassification creates unsigned bias rather than simple classical attenuation (Aigner 1973, Yang et al. 2018), we audited 349 masked filings with two blinded coders who reached 99.13 per cent design-weighted strict agreement ($\kappa = 0.9319$, Table 6, Panel A). The automated strict classifier posts high sensitivity and specificity with only 77.48 per cent positive predictive value. Over one-fifth of machine-positive calls lack human confirmation. That risk-section ambiguity explains why pooled slopes vary across the nine multi-model rules in Appendix Table A9. In 2024–2025, every firm-clustered interval across those nine rules stays above zero. Strict two-phase regression calibration (Chen, Hong, and Tamer 2005) in Panel B of Table 6 keeps operating slopes positive at 2.04 percentage points pooled and 2.76 points in 2024–2025 (with a 3.80-point recent role contrast). Covariate-conditional calibration, Coder 1 alone, and the positive-support frame return nearly identical estimates. False positives do not create the gap.

TABLE 6. Stratified Human Validation ($n_{\text{eval}} = 349$) and Misclassification-Corrected Estimates

Panel A: Inter-coder reliability and classifier performance ($N = 28{,}266$ frame; $n_{eval} = 349$ across 13 strata)				
Validation diagnostic	Five-way	Strict (A/C)	Inclusive (A/C/E)	Supply (B/C)
Classified prevalence in evaluation frame (%)	–	8.64	8.64	8.47
Human design-weighted prevalence (%)	–	6.79	10.26	5.17
Unweighted agreement (%) / Cohen's κ	86.82 / 0.8267	91.69 / 0.8323	93.41 / 0.8636	96.56 / 0.9149
Design-weighted agreement (%) / Cohen's κ	97.66 / 0.9016	99.13 / 0.9319	98.63 / 0.9230	99.25 / 0.9269
Classifier sensitivity / specificity (%)	–	98.51 / 97.91	71.00 / 98.49	98.76 / 96.45
Positive / negative predictive value (%)	–	77.48 / 99.89	84.35 / 96.74	60.26 / 99.93

Panel B: Misclassification-corrected linear probability estimates ($100 \times Y_{if}$; standard errors in parentheses)				
Sample, estimator, and coding rule	Frame N	Operation (β)	Supply (θ)	Op. − Supply
1. Pooled 2008–2025 post-stratified frame				
Uncorrected three-reader baseline[a]	28,266	1.99*** (0.70)	−0.23 (0.60)	2.23* (1.15)
Two-phase calibration, strict (A/C)[b]	28,266	2.04** (0.86)	−0.26 (0.85)	2.30 (1.46)
Two-phase calibration, inclusive (A/C/E)	28,266	2.59*** (0.89)	−0.31 (0.80)	2.90** (1.37)
One-step score corr. (DSL), strict (A/C)[c]	28,266	2.09 (2.52)	−0.58 (3.00)	2.67 (4.59)
One-step score corr. (DSL), inclusive (A/C/E)[c]	28,266	1.64 (1.96)	−0.37 (2.89)	2.01 (3.89)
2. Recent 2024–2025 post-stratified cohort				
Uncorrected three-reader baseline[a]	5,507	2.55*** (0.79)	−0.50 (0.88)	3.05** (1.44)
Two-phase calibration, strict (A/C)[b]	5,507	2.76*** (0.95)	−1.04 (1.16)	3.80** (1.77)
Two-phase calibration, inclusive (A/C/E)	5,507	3.49*** (0.99)	−0.46 (1.16)	3.95** (1.58)
One-step score corr. (DSL), strict (A/C)[c]	5,507	0.47 (2.77)	2.27 (4.38)	−1.80 (5.02)
One-step score corr. (DSL), inclusive (A/C/E)[c]	5,507	0.62 (2.16)	2.40 (4.46)	−1.77 (4.55)
3. Exact positive-inclusion support frame ($\pi_{if} > 0$; 99.62% of main sample)				
Uncorrected three-reader baseline	28,158	2.00*** (0.71)	−0.24 (0.60)	2.23* (1.17)
Two-phase calibration, strict (A/C)	28,158	2.05** (0.84)	−0.25 (0.87)	2.30 (1.48)
One-step score corr. (DSL), strict (A/C)[c]	28,158	2.09 (2.55)	−0.58 (2.94)	2.67 (4.78)

Notes: Blinded two-coder human validation across 13 strata ($n_{\text{eval}} = 349$ excluding one worked example; 124 A, 64 B, 31 C, 76 D, 54 E; design-weighted shares 4.50, 2.87, 2.30, 86.87, and 3.47%; full-frame supply prevalence 8.43%). Panel B standard errors (in parentheses) come from 1,000 joint Phase-1 firm-clustered and Phase-2 stratified bootstrap replications. Paired 400-bootstrap contrasts on strict calibration: Coder 1 gives 2.03^{**} $[0.64, 3.68]$ ($p = 0.018$) pooled and 2.64^{***} $[0.73, 4.69]$ ($p = 0.010$; contrast 3.52^{**} $[0.54, 7.16]$) recent (Coder 2 minus Coder 1: $+0.01$, $p = 0.948$; $+0.12$, $p = 0.561$); Inclusive minus Strict gives $+0.54^{**}$ $[0.09, 1.04]$ ($p = 0.029$) pooled and $+0.73^{**}$ $[0.11, 1.50]$ ($p = 0.047$) recent under calibration vs. -0.45 ($p = 0.812$) and $+0.15$ ($p = 0.955$) under DSL; DSL minus calibration gives $+0.05$ ($p = 0.985$) pooled and -2.29 ($p = 0.316$; contrast -5.60, $p = 0.254$) recent. Stratum S12 ($N = 21{,}312$, 18 D, weight $1{,}184.0$) carries a 15.33% one-sided binomial upper bound vs. 10.27% unadjusted tipping share ($2{,}188.5$ filings). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[a] Analytic firm-clustered CR1 standard errors (95% t_{G-1} CIs: $[0.62, 3.37]$ pooled and $[1.01, 4.09]$ on standalone $N = 5{,}507$).

[b] Percentile 95% bootstrap CIs on strict calibration: $[0.50, 3.85]$ (β) and $[-0.24, 5.56]$ ($\beta - \theta$) pooled; $[1.11, 4.76]$ (β) and $[0.71, 7.40]$ ($\beta - \theta$) recent; $[0.63, 3.73]$ on exact positive-support frame. Covariate-conditional calibration gives 2.03 pooled and 2.71 recent (contrast 3.75).

[c] Percentile 95% bootstrap CIs on strict one-step DSL: $[-2.09, 8.34]$ pooled and $[-5.35, 5.95]$ recent; solving the full DSL linear system directly yields 2.09 pooled and 0.03 recent ($\theta = 3.52$).

One validation result contradicted our pre-specified expectation. We expected inclusive human coding of nominalised AI risk warnings (E) to dilute estimated slopes. Instead, inclusive calibration rises reliably above strict estimates to 2.59 percentage points pooled and 3.49 points in 2024–2025 (Table 6). Roughly 82.9 per cent of weighted E mass sits inside automated comparison strata (Appendix B.4), where widespread generic risk warnings and high breach rates among contested human calls amplify rather than attenuate calibrated exposure (Yang et al. 2018, Egami et al. 2023).

Estimator assumptions govern 2024–2025 precision. When one-step orthogonal score correction (DSL) drops calibration's surrogacy assumption via inverse inclusion weights (Robins, Rotnitzky, and Zhao 1994, Egami et al. 2023), its pooled strict slope of 2.09 percentage points tracks calibration closely across wide confidence intervals (Table 6). Heavy inverse inclusion weights in 2024–2025 inflate bootstrap variance 8.56-fold and pull the strict point estimate down to 0.47 percentage points. Small heavily weighted strata, above all the 18 validation draws inside early comparison stratum S12, make recent one-step score estimates statistically uninformative rather than evidence of zero risk (Appendix B.4).

Within-firm fixed effects trade bias against power. Across the 820 firms that switch operating status across filings, firm and year fixed effects yield an imprecise slope of 0.74 percentage points (Table 4, column 7, and Figure 2, panel d). Firm plus three-digit industry-by-year fixed effects gives 1.43 percentage points against an 80 per cent minimum detectable effect of 2.86 points (Appendix Table A23). Neither pre-adoption event-time leads nor Mundlak and Hausman tests indicate that pre-trends or stable firm traits bias cross-firm estimates (Mundlak 1978). The pre-2023 industry-readiness instrument lacks first-stage relevance (Staiger and Stock 1997, Lee et al. 2022). Limited panel power and attenuation from misdated transitions (Griliches and Hausman 1986) leave within-firm regressions statistically uninformative rather than able to disprove operational exposure.

Calendar concentration imposes a separate boundary. Because over three-fifths of all operating filings (1,551 of 2,441) arrive in 2024 and 2025 after federal cybersecurity disclosure rules (Securities and Exchange Commission 2023), skewed calendar shares leave only 4.74 effective calendar clusters across eighteen years (Carter, Schnepel, and Steigerwald 2017). Two-way firm-by-year clustering and a six-point Webb wild cluster bootstrap by filing year ($p = 0.082$) push inference near the ten per cent threshold (Appendix Tables A22 and A23). A sample without 2024 likewise lowers the operating slope to 1.29 percentage points and widens its interval across zero (Appendix D.1). Even so, within-cell firm-persistent permutation rejects random reassignment cleanly. Sensitivity checks show that positive slopes still survive once we exclude the shared Cencora episode or the three most influential firms.

Functional form does not drive the gap. In Appendix Table A22, concentrated industry-by-year exponential QMLE regressions show that internal AI operation carries a relative risk of 1.585 and a 3.10 percentage-point marginal gap (Wooldridge 1999). Entropy balancing yields 2.64 percentage points pooled and 3.10 points in 2024–2025 (Hainmueller 2012). Under proportional-selection

bounds (Oster 2019), unobservables must select 3.63 or 1.11 times as strongly as observables to eliminate the baseline estimate.

Retrospective tests evaluate 2025 out-of-sample prediction. Across 2,693 filings in 2025, separate internal-operation and commercial-supply indicators outperform a combined AI indicator on fixed-origin probability loss by 0.042 points and lift precision-recall area by 1.29 points over cybersecurity text (Appendix Table A24). Monthly updates struggle under rapid diffusion. During that adoption surge, cross-firm risk rankings remain steady yet probability calibration drifts over time.

Our evidence narrows Table 1 sharply. Primary-letter attribution, the near-zero vendor placebo, and cyber-versus-physical specificity weigh directly against pure AI-firm scale and upstream vendor propagation. Strictly post-filing onsets, longer reporting lags, security-AI exclusions, and masked active-use checks likewise weigh against pre-filing breach onset, faster disclosure, and hypothetical buzzwords. Four limits still bound these findings. On disclosure selection, annual reports cannot separate workflow vulnerability under Item 105 materiality from threat-informed risk drafting by managers who sense weak internal controls. Unmeasured personal-record holdings remain an open cross-firm confounder behind an underpowered 2.86-point within-firm threshold. In the calendar dimension, concentration across 4.74 effective clusters keeps year-level tests near the ten per cent boundary. Wide one-step score intervals in 2024–2025 also leave classification-corrected recent estimates dependent on regression calibration's conditional surrogacy assumption.

6. Conclusion

This paper asked whether corporate data-breach incidence varies with how US public companies deploy artificial intelligence across annual Form 10-K reports. Internal AI operation connects predictive and generative models, user permissions, and corporate pipelines directly to personal records. Inside shared three-digit industry-by-year cells, firms that operate AI internally experience substantially higher twelve-month statutory breach rates than same-sector peers that discuss AI without active internal deployment. Commercial AI supply shows no rise. Across 2024–2025 regulatory archives, primary breach letters confirm that verified intrusions inside the issuer's own perimeter drive the bulk of the audited operating gap beside a near-zero third-party vendor placebo. Positive operational estimates survive post-filing breach start dates, delayed observation windows, and active present-tense risk-factor disclosures. Dual-role offsets remain unresolved. All reported slopes remain conditional.

Our evidence builds on three empirical literatures in information systems and finance. Earlier firm-level AI studies infer adoption from workforce skills and national enterprise surveys (Babina et al. 2024, Calvino and Fontanelli 2026) or link AI labour, deployment, and cyber exposure to bank operational losses, technology-sector breaches, and corporate patent responses (McLemore and Mihov 2026, Xiao et al. 2026, Srinivas, Zhang, and Hong 2026, Rishabh, Mihet, and Jang-Jaccard 2026). This study separates internal AI deployment from commercial software sales across

every US public industry and links both roles to verified issuer breach letters. For mandatory risk-disclosure research (Wang, Kannan, and Ulmer 2013, Campbell et al. 2014, Florackis et al. 2023), our placement audit shows that Item 105 risk-factor location and active present-tense prose fit both created operational vulnerability and threat-informed managerial disclosure. Unresolved dual-role slopes leave absorptive-capacity arguments (Cohen and Levinthal 1990) and symbolic-governance claims (Angst et al. 2017) open across unmeasured vendor hosting contracts. Because automated disagreements cluster inside complex filings of large breached issuers (Yang et al. 2018), downstream human validation inside regression estimators (Chen, Hong, and Tamer 2005, Egami et al. 2023) proves essential in applied text studies where broader risk warnings can shift calibrated slopes unexpectedly. Simple classical attenuation fails under outcome-correlated error.

Corporate governance, cyber insurance, and financial supervision each draw concrete priorities from this role divide. For corporate boards and senior security leaders, oversight should audit internal AI data connectors, staff access permissions, query logs, and continuous model updates (Berente et al. 2021, Tabassi 2023, Anderson and Moore 2006). Commercial engineering alone cannot insulate corporate databases. In cyber insurance, where unobservable security effort constrains policy design (Cezar, Cavusoglu, and Raghunathan 2014), underwriters can use operational risk-factor statements to guide diligence on cloud hosting and data-access controls rather than automatic surcharges (Wang, Kannan, and Ulmer 2013). Cyber underwriting models still require regular recalibration because 2025 probability forecasts drift across fast adoption cohorts even when risk rankings beat cybersecurity-text benchmarks. For securities regulators and financial supervisors (Securities and Exchange Commission 2023, Bank of England and Financial Conduct Authority 2024), annual Form 10-K reports rarely reveal whether corporate AI runs on internal infrastructure or external cloud servers. Silence in annual filings is not evidence of non-use.

Four limits bound archival evidence. Persistent consumer-record holdings can raise both AI returns and target value while sparse within-firm switches leave panel fixed-effects models underpowered against that cross-firm confounder (Mundlak 1978). Public registries miss undetected breaches (Amir, Levi, and Livne 2018). Archival Item 105 prose cannot separate operational exposure from threat-informed risk drafting. In the calendar dimension, 2024–2025 concentration leaves few effective year clusters (Carter, Schnepel, and Steigerwald 2017) and ties recent error-corrected estimates to calibration surrogacy.

Richer data can narrow these limits. As enterprise AI rollouts spread across post-2025 cohorts, longer panels with exogenous cloud shocks, internal deployment logs, cyber-insurance claims, and security telemetry can absorb fixed record stocks, widen calendar cluster support, and separate operational exposure from disclosure choices and unobserved detection. Whether firms connect live AI tools to personal records safely will define corporate cyber governance.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?" *Quarterly Journal of Economics* 138 (1): 1–35.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial Intelligence and Jobs: Evidence from Online Vacancies." *Journal of Labor Economics* 40 (S1): S293–S340.

Aigner, Dennis J. 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation." *Journal of Econometrics* 1 (1): 49–59.

Amir, Eli, Shai Levi, and Tsafrir Livne. 2018. "Do Firms Underreport Information on Cyber-Attacks? Evidence from Capital Markets." *Review of Accounting Studies* 23 (3): 1177–1206.

Anderson, Ross, and Tyler Moore. 2006. "The Economics of Information Security." *Science* 314 (5799): 610–613.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Angst, C. M., E. S. Block, J. D'Arcy, and K. Kelley. 2017. "When Do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches." *MIS Quarterly* 41 (3): 893–916.

Artstein, Ron, and Massimo Poesio. 2008. "Inter-Coder Agreement for Computational Linguistics." *Computational Linguistics* 34 (4): 555–596.

Ash, Elliott, and Stephen Hansen. 2023. "Text Algorithms in Economics." *Annual Review of Economics* 15 (1): 659–688.

August, Terrence, Marius Florin Niculescu, and Hyoduk Shin. 2014. "Cloud Implications on Software Network Structure and Security Risks." *Information Systems Research* 25 (3): 489–510.

Babina, T., A. Fedyk, A. He, and J. Hodson. 2024. "Artificial Intelligence, Firm Growth, and Product Innovation." *Journal of Financial Economics* 151: 103745.

Bank of England, and Financial Conduct Authority. 2024. "Artificial Intelligence in UK Financial Services – 2024." Joint Survey Report, Bank of England and Financial Conduct Authority. 21 November.

Barrios, John Manuel, John L. Campbell, Ryan G. Johnson, and Christine Liu. 2025. "Signals or Smoke? The Determinants and Informativeness of Corporate Artificial Intelligence (AI) Disclosures." Working Paper 5133107, SSRN.

Beck, Nathaniel. 2020. "Estimating Grouped Data Models with a Binary-Dependent Variable and Fixed Effects via a Logit versus a Linear Probability Model: The Impact of Dropped Units." *Political Analysis* 28 (1): 139–145.

Berente, Nicholas, Bin Gu, Jan Recker, and Radhika Santhanam. 2021. "Managing Artificial Intelligence." *MIS Quarterly* 45 (3): 1433–1450.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics* 117 (1): 339–376.

Calvino, Flavio, and Luca Fontanelli. 2026. "AI Users Are Not All Alike: The Characteristics of French Firms Buying and Developing AI." *Research Policy* 55 (5): 105473.

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics* 29 (2): 238–249.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90 (3): 414–427.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–372.

Campbell, John L., Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu, and Logan B. Steele. 2014. "The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings." *Review of Accounting Studies* 19 (1): 396–455.

Carter, Andrew V., Kevin T. Schnepel, and Douglas G. Steigerwald. 2017. "Asymptotic Behavior of a t-Test Robust to Cluster Heterogeneity." *Review of Economics and Statistics* 99 (4): 698–709.

Cezar, Asunur, Huseyin Cavusoglu, and Srinivasan Raghunathan. 2014. "Outsourcing Information Security: Contracting Issues and Security Implications." *Management Science* 60 (3): 638–657.

Chen, Pei-yu, Gaurav Kataria, and Ramayya Krishnan. 2011. "Correlated Failures, Diversification, and Information Security Risk Management." *MIS Quarterly* 35 (2): 397–426.

Chen, Xiaohong, Han Hong, and Elie Tamer. 2005. "Measurement Error Models with Auxiliary Data." *Review of Economic Studies* 72 (2): 343–366.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen. 2020. "Lazy Prices." *The Journal of Finance* 75 (3): 1371–1415.

Cohen, Wesley M., and Daniel A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35 (1): 128–152.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson Estimation with High-Dimensional Fixed Effects." *Stata Journal* 20 (1): 95–115.

Crosignani, M., M. Macchiavelli, and A. F. Silva. 2023. "Pirates without Borders: The Propagation of Cyberattacks through Firms' Supply Chains." *Journal of Financial Economics* 147 (2): 432–448.

Dell, Melissa. 2025. "Deep Learning for Economists." *Journal of Economic Literature* 63 (1): 5–58.

Dyer, Travis, Mark Lang, and Lorien Stice-Lawrence. 2017. "The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation." *Journal of Accounting and Economics* 64 (2–3): 221–245.

Edwards, Benjamin, Steven Hofmeyr, and Stephanie Forrest. 2016. "Hype and Heavy Tails: A Closer Look at Data Breaches." *Journal of Cybersecurity* 2 (1): 3–14.

Egami, Naoki, Musashi Hinck, Brandon Stewart, and Hanying Wei. 2023. "Using Imperfect Surrogates for Downstream Inference: Design-Based Supervised Learning for Social Science Applications of Large Language Models." In *Advances in Neural Information Processing Systems*, vol. 36: 68589–68601.

Eisfeldt, Andrea, Gregor Schubert, and Miao Ben Zhang. 2023. "Generative AI and Firm Values." Working Paper 31222, National Bureau of Economic Research.

Florackis, C., C. Louca, R. Michaely, and M. Weber. 2023. "Cybersecurity Risk." *Review of Financial Studies* 36 (1): 351–407.

Foerderer, J., and S. W. Schuetz. 2022. "Data Breach Announcements and Stock Market Reactions: A Matter of Timing?" *Management Science* 68 (10): 7298–7322.

Fong, Christian, and Matthew Tyler. 2021. "Machine Learning Predictions as Regression Covariates." *Political Analysis* 29 (4): 467–484.

Frankel, Richard, Jared Jennings, and Joshua Lee. 2022. "Disclosure Sentiment: Machine Learning vs. Dictionary Methods." *Management Science* 68 (7): 5514–5532.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–574.

Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120 (30): e2305016120.

Glasserman, Paul, and Caden Lin. 2024. "Assessing Look-Ahead Bias in Stock Return Predictions Generated by GPT Sentiment Analysis." *The Journal of Financial Data Science* 6 (1): 25–42.

Gordon, Lawrence A., and Martin P. Loeb. 2002. "The Economics of Information Security Investment." *ACM Transactions on Information and System Security* 5 (4): 438–457.

Greene, William. 2004. "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects." *Econometrics Journal* 7 (1): 98–119.

Greshake, Kai, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz. 2023. "Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection." In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISec '23)*, pp. 79–90.

Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (1): 93–118.

Hainmueller, J. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20 (1): 25–46.

Haislip, Jacob, Jee-Hae Lim, and Robert Pinsker. 2021. "The Impact of Executives' IT Expertise on Reported Data Security Breaches." *Information Systems Research* 32 (2): 318–334.

Han, M., H. Shen, J. Wu, and X. M. Zhang. 2025. "Artificial Intelligence and Firm Resilience: Empirical Evidence from Natural Disaster Shocks." *Information Systems Research* 36 (4): 2116–2133.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. 2019. "Firm-Level Political Risk: Measurement and Effects." *Quarterly Journal of Economics* 134 (4): 2135–2202.

von Hippel, Eric. 1994. ""Sticky Information" and the Locus of Problem Solving: Implications for Innovation." *Management Science* 40 (4): 429–439.

Hoberg, Gerard, and Gordon Phillips. 2016. "Text-Based Network Industries and Endogenous Product Differentiation." *Journal of Political Economy* 124 (5): 1423–1465.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70.

Horowitz, Joel L., and Charles F. Manski. 1995. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63 (2): 281–302.

Jamilov, Rustam, Hélène Rey, and Ahmed Tahoun. 2026. "The Anatomy of Cyber Risk." Working Paper 28906, National Bureau of Economic Research. Revised July 2026.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz. 2021. "Risk Management, Firm Reputation, and the Impact of Successful Cyberattacks on Target Firms." *Journal of Financial Economics* 139 (3): 719–749.

Kim, Jeong-Bon, Chong Wang, and Feng (Harry) Wu. 2024. "Privacy Breaches and the Effect of Customer Notification." *MIS Quarterly* 48 (4): 1483–1502.

Kim, Seung Hyun, and Juhee Kwon. 2019. "How Do EHRs and a Meaningful Use Initiative Affect Breaches of Patient Information?" *Information Systems Research* 30 (4): 1184–1202.

Klarna. 2024. "Klarna AI Assistant Handles Two-Thirds of Customer Service Chats in Its First Month." Press release, 27 February. https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/.

Kunreuther, Howard, and Geoffrey Heal. 2003. "Interdependent Security." *Journal of Risk and Uncertainty* 26 (2–3): 231–249.

Kwon, J., and M. E. Johnson. 2014. "Proactive Versus Reactive Security Investments in the Healthcare Sector." *MIS Quarterly* 38 (2): 451–472.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter. 2022. "Valid t-Ratio Inference for IV." *American Economic Review* 112 (10): 3260–3290.

Li, Wilson Weixun, Alvin Chung Man Leung, and Wei Thoo Yue. 2023. "Where is IT in Information Security? The Interrelationship among IT Investment, Security Awareness, and Data Breaches." *MIS Quarterly* 47 (1): 317–342.

Lin, X., and G. She. 2025. "Timely Cybersecurity Disclosure and Information Manipulation." *Management Science* 71 (11): 9308–9327.

Liu, Lisa Yao. 2025. "Financial Statement Audits and Data Breaches." *Management Science* 71 (8): 6340–6366.

Lopez-Lira, Alejandro, Yuehua Tang, and Mingyin Zhu. 2025. "The Memorization Problem: Can We Trust LLMs' Economic Forecasts?" Working Paper 5217505, SSRN.

Loughran, Tim, and Bill McDonald. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35–65.

Loughran, Tim, and Bill McDonald. 2016. "Textual Analysis in Accounting and Finance: A Survey." *Journal of Accounting Research* 54 (4): 1187–1230.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan. 2026. "Large Language Models: An Applied Econometric Framework." *Annual Review of Economics* 18 (1): 283–316.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb. 2023. "Cluster-Robust Inference: A Guide to Empirical Practice." *Journal of Econometrics* 232 (2): 272–299.

Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review* 80 (2): 319–323.

Mayer, Kyle J., and Robert M. Salomon. 2006. "Capabilities, Contractual Hazards, and Governance: Integrating Resource-Based and Transaction Cost Perspectives." *Academy of Management Journal* 49 (5): 942–959.

McElheran, Kristina, J. Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas. 2024. "AI Adoption in America: Who, What, and Where." *Journal of Economics & Management Strategy* 33 (2): 375–415.

McLemore, P., and A. Mihov. 2026. "AI and Operational Losses: Evidence from U.S. Bank Holding Companies." *Review of Corporate Finance Studies*: cfag003.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46 (1): 69–85.

OpenAI. 2023. "March 20 ChatGPT Outage: Here's What Happened." Company incident report, 24 March. https://openai.com/index/march-20-chatgpt-outage/.

Oster, E. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

Rammer, Christian, Gastón P. Fernández, and Dirk Czarnitzki. 2022. "Artificial Intelligence and Industrial Innovation: Evidence from German Firm-Level Data." *Research Policy* 51 (7): 104555.

Rishabh, K., R. Mihet, and J. Jang-Jaccard. 2026. "Cyberrisk and AI Firms." *Review of Corporate Finance Studies*: cfag018.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89 (427): 846–866.

Romanosky, Sasha. 2016. "Examining the Costs and Causes of Cyber Incidents." *Journal of Cybersecurity* 2 (2): 121–139.

Romanosky, Sasha, Rahul Telang, and Alessandro Acquisti. 2011. "Do Data Breach Disclosure Laws Reduce Identity Theft?" *Journal of Policy Analysis and Management* 30 (2): 256–286.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using boottest." *Stata Journal* 19 (1): 4–60.

Santos Silva, J. M. C., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics* 88 (4): 641–658.

Sarkar, Suproteem, and Keyon Vafa. 2024. "Lookahead Bias in Pretrained Language Models." Working Paper 4754678, SSRN.

Securities and Exchange Commission. 2020. "Modernization of Regulation S-K Items 101, 103, and 105." Release Nos. 33-10825; 34-89670; 17 C.F.R. § 229.105, 26 August.

Securities and Exchange Commission. 2023. "SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies." Press release 2023-139, 26 July. https://www.sec.gov/newsroom/press-releases/2023-139.

Sen, R., and S. Borle. 2015. "Estimating the Contextual Risk of Data Breach: An Empirical Approach." *Journal of Management Information Systems* 32 (2): 314–341.

Srinivas, S., L. Zhang, and K. Hong. 2026. "The Cybersecurity Consequences of AI Strategy: How Automation-Orientation and Augmentation-Orientation Shape Data Breach Risk." Working Paper 7014258, SSRN.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.

Tabassi, E. 2023. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST AI100-1, National Institute of Standards and Technology. 26 January.

Tanriverdi, Hüseyin, Juhee Kwon, and Ghiyoung Im. 2025. "Taming Complexity in the Cybersecurity of Multihospital Systems: The Role of Enterprise-wide Data Analytics Platforms." *MIS Quarterly* 49 (1): 243–274.

Wang, Le, Xukai Huang, Xin (Robert) Luo, Paul Benjamin Lowry, and Chaofan Yang. 2026. "Breakthrough or Breakdown? Firm-Level Generative AI Exposure and Data Breach Risk." In *Proceedings of the International Conference on Information Systems (ICIS 2026)*, Short paper 6.

Wang, Q., E. W. T. Ngai, D. Pienta, and J. B. Thatcher. 2023. "Information Technology Innovativeness and Data-Breach Risk: A Longitudinal Study." *Journal of Management Information Systems* 40 (4): 1139–1170.

Wang, T., K. N. Kannan, and J. R. Ulmer. 2013. "The Association Between the Disclosure and the Realization of Information Security Risk Factors." *Information Systems Research* 24 (2): 201–218.

Wooldridge, J. M. 1999. "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics* 90 (1): 77–97.

Xiao, Wei, Shan Xu, Jin Li, and Jian Mou. 2026. "Adopting AI, Inviting Risk? An Empirical Analysis of Corporate Cybersecurity Consequences in IT Firms." *Electronic Markets* 36: 83.

Yang, Mochen, Gediminas Adomavicius, Gordon Burtch, and Yuqing Ren. 2018. "Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining." *Information Systems Research* 29 (1): 4–24.

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. "Can Large Language Models Transform Computational Social Science?" *Computational Linguistics* 50 (1): 237–291.

Online Appendix

A0. Summary of empirical checks

Our summary directory (Table A1) collects the measurement, sample, and specification checks across the main text and Online Appendix. For each question, it records what we did, what we found, and where the evidence appears. Section numbers denote main-text sections and letters denote this appendix. Estimates are expressed in percentage points of the twelve-month notification rate (sample mean 3.07 per cent), with 95 per cent firm-clustered intervals in brackets around the 1.99 [0.62, 3.37] baseline estimate.

TABLE A1. Summary of Measurement, Identification, and Sensitivity Checks

Question	Empirical check	Principal result	Location
How is running AI read from a filing?	A language model (OpenAI GPT-6) applies written rules to the business-section and risk-section passages that carry an AI cue. Every yes cites the passages behind it.	2,441 operating and 25,825 comparison filings in the sample; 1,614 of the operating filings also report development or supply.	§3; Table 2; App. B.1
Would another model read the passages the same way?	A second reading by Microsoft-Decision-1 of all 55,445 read filings with passages in the business section or Item 1A.	The same answer on operation for 96.5 per cent of filings; kappa 0.814 [0.802, 0.824].	§3; Table A7
Is agreement lower where operating filings are common?	Agreement by filing year, by kind of cue, and by archive partition.	98.4 per cent in filings to 2023 and 90.2 per cent in 2024 to 2026, the years that hold most operating filings.	§3; Table A7
Which reading is right where the two differ?	A third reading by Anthropic Claude of the 768 contested filings and 80 agreement filings, with no label and no outcome shown. Every yes needs a quotation from the filing.	Of the 480 contested filings it resolves, 381 side with the primary reading and 99 with the second; 77 of 80 agreement filings confirmed; 274 of 274 quotations found word for word.	§3; Table A8
Does the estimate depend on the reading?	The main model under each of the nine readings.	Pooled estimates from 0.83 to 2.15. In 2024 to 2025 they lie between 1.65 and 2.98, every interval above zero.	§5.3; Table A9; Fig. A1
Did the models use what they know about named firms?	Instructions exclude outside knowledge and later events. A masked reread of 40 filings with names, identifiers, and dates removed.	The same answer for 33 of 40 filings, and 36 of 40 on yes against not yes. No answer moved between yes and not established.	§3; Table A10

Table A1 (continued)

Question	Empirical check	Principal result	Location
What about filings that were never read?	All 44,656 archive filings without an initial AI cue went through a second screen with further AI terms, leaving 44,569 unread.	The screen found passages in 33 filings (30 of 40,192 in the main partition and 3 of 4,464 in the tenth). None of the 87 read reports operation or supply.	§3; App. B.1
Whom are operating filings compared with?	Comparison filings are read filings that do not establish operation. Development or supply enters the model as its own indicator.	Of the 25,825 comparison filings, 770 report development or supply only and 25,055 neither role.	§3; Table 3
Do firms that run AI simply hold more data?	Digital and data wording among the controls; the within-firm model with firm and year effects.	Within firms 0.74 [−1.23, 2.71], against 2.02 on the same filings with broad industry and year effects; 820 firms change operating state.	§2; §5.3
Do firms that monitor more simply report more?	Cyber discussion and prior notification added to the model.	Operation falls from 2.61 [1.15, 4.08] to 2.02 [0.77, 3.26] on the full sample with broad industry and year effects.	§5.1; Table 4
Does the SEC's 2023 cybersecurity rule drive the 2024 to 2025 result?	Cyber discussion leaves out Item 1C; the 114 sample filings whose evidence lies only in Item 1C are dropped; the period model.	2.01 [0.62, 3.39] without those filings. 2.55 [0.98, 4.12] in 2024 to 2025 against 1.84 [−0.65, 4.33] before, a difference of 0.71 [−2.14, 3.55].	§3; §5.1; Table A14; Table A18
Are recent notifications fully recorded?	Recording delay measured from two copies of the source; comparison rates reweighted to the 2023 industry and size mix; each filing year omitted in turn.	About 7 and 9 per cent of in-window reports for 2024 and 2025 filings not yet recorded. Reweighted comparison rates of 2.39 and 1.51 per cent, against 3.95 in 2023.	§3; Fig. 1; App. A.2; App. D.3
Did breaches lead firms to adopt security AI?	Core 97 filings (43 firms) and broader 224 filings (112 firms) whose AI serves cyber defence or fraud detection dropped, or recoded as not operating.	Core: 1.84 [0.47, 3.21] dropped, 1.82 [0.47, 3.17] recoded; census: 1.81 [0.41, 3.22] dropped, 1.65 [0.25, 3.05] recoded.	§5.2; App. D.1
Is each notification attributed to the right firm?	Notifier, compromised provider, affected client, and firm kept apart; all 162 notified 2024–2025 windows (210 links) audited using archived records (106 archived regulatory PDF notices and 104 incident-detail fields).	157 of 162 entity-verified (96.9%): 82 own-system (1.85 [0.71, 2.99]), 49 state tabular (0.59), 26 pure vendor client (−0.11 [−0.70, 0.47]), 5 dropped collisions.	§5.2; Table 5; Table A5; App. A.2
Do firms that build their AI fare better?	Operation with and without development of AI.	2.62 [0.59, 4.66] without development and 1.07 [−0.48, 2.63] with it; difference −1.55 [−4.08, 0.98], and −0.12 [−2.98, 2.73] in 2024 to 2025.	§5.1; App. D.2

Table A1 (continued)

Question	Empirical check	Principal result	Location
Is a linear model suitable for a rare outcome?	A logit on the full sample with broad industry and year effects, plus concentrated three-digit industry-by-year exponential QMLE.	Logit average difference 1.63 [0.55, 2.71]; exponential QMLE relative risk 1.585 [1.208, 2.079]; no linear fitted value above one.	§4; App. C.2; App. E
Are errors correlated across firms that share a breach?	Clusters formed from firms joined by a shared episode or a common corporate group.	1.99 [0.61, 3.37] with 4,439 clusters, against [0.62, 3.37] with 4,451 firm clusters.	§4; App. C.2
How does inference adjust for multiple focal tests?	Step-down Holm adjustment across the two focal main-model contrasts and the four full-sample operating specifications.	Holm-adjusted p of 0.009 for operation and 0.053 for operation minus supply in the main sample; 0.002 and 0.004 for pooled full-sample models.	§4; App. C.2
Do a few firms or one breach make the result?	The three firms with the largest influence on the estimate omitted together; the nine firms named in the one episode shared across firms (Cencora) omitted.	1.42 [0.23, 2.61] without the three firms; 1.90 [0.53, 3.26] without the nine.	§5.3; App. D.1
Does the result rest on one year?	Each filing year omitted in turn.	Estimates from 1.29 (2024 omitted) to 2.66 (2025 omitted).	§5.3; App. D.1; Fig. A3
Does it matter where the firm describes its AI?	The location model; a masked reading, blind to outcomes, of the 770 risk-section-only passages.	Risk-section-only operation 3.39 [1.42, 5.37] against business-section operation 0.97 [−0.67, 2.62]. Of the 770 passages, 86.2 per cent state operation.	§5.1; Table 4; App. B.3
Is the location result a firm-size effect?	Indicators for nine of the ten size deciles added to the location model while keeping log total assets; raw rates by size fifth.	Risk-section statements of operation 2.63 [0.53, 4.72]. Risk-section-only operating filings are notified more often than comparison filings in every size fifth.	§5.1; Fig. A2; Table A12
Do pre-filing onsets or faster reporting explain the gap?	Drop known pre-filing onsets, require strict post-filing onset ($d_{\text{breach}} > d_f$), shift windows past 90/180 days, and test reporting lags and weak readiness IV.	Strict post-filing onset 1.04 [0.04, 2.04] pooled and 1.26 [0.05, 2.46] recent; delayed 90-day 1.71 and 2.31; median lags 69 vs. 50 d; data-plus-digital readiness index ($F = 0.09$).	§5.2; §5.3; App. F; Table A23
Does automated coding bias drive or erase the slope?	Two blinded human coders audit 350 stratified filings (349 evaluated) for joint two-phase regression calibration and one-step score correction.	Design-weighted $\kappa = 0.9016$ (0.8267 unweighted) five-way, 0.9319 (0.8323) strict; strict calibration 2.04 [0.50, 3.85] pooled and 2.76 [1.11, 4.76] recent; one-step score correction 2.09 / 0.47.	§5.3; App. B.4; Table 6

Table A1 (continued)

Question	Empirical check	Principal result	Location
Does role separation improve 2025 prediction out of sample?	Retrospective temporal-split fixed (1 Jan) and monthly expanding origins across 2,693 filings in 2025 with 2,000 paired bootstraps.	Role separation lowers fixed-origin Log-Loss $\times 100$ vs. undifferentiated AI (-0.042 ridge L2, $p_{\text{boot}} = 0.040$; -0.052 MLE, $p_{\text{boot}} = 0.018$), while uncalibrated cross-year Log-Loss trails cyber text under base-rate shift.	§5.3; App. F; Table A24

Notes: Sample of 28,266 filings across 4,451 firms from 2008 to 2025, unless a row specifies the 28,536-filing full sample (4,484 firms). Reported point estimates come from the baseline linear probability model with three-digit SIC-by-year fixed effects and seven filing/firm controls, with 95 per cent firm-clustered confidence intervals in brackets, unless noted otherwise. Section numbers denote main-text sections; Tables 1–6 and Figures 1–2 appear in the main manuscript. Figure A1, Figure A2, Figure A3, and all Appendix tables appear in the Online Appendix.

A. Data, sample, and timing

A.1 Sources, sample construction, and variable timing

We read Form 10-K filings from an archive of 100,274 annual reports submitted to the SEC between 1994 and 2026. Of these, 84,037 were filed up to 2023 and 16,237 in 2024 to 2026. Each filing supplies the business section (Item 1) and risk factors (Item 1A), and filings of 2024 to 2026 also include Item 1C, the cybersecurity item mandated by the SEC since 2023 (Securities and Exchange Commission 2023). Every record carries the firm's SEC identifier, corporate name, and filing date, while Compustat annual files supply total assets and historical industry codes under the timing rules detailed below.

Breach notifications come from the Privacy Rights Clearinghouse Data Breach Chronology snapshot of 5 August 2026. That database holds 103,366 notification rows with report dates from 21 June 2006 to 31 July 2026 across 20 state authorities and the US Department of Health and Human Services, from which 8,823 rows link to sample issuers. Massachusetts (28,417 rows), Indiana (12,760), and New Hampshire (8,783) supply the largest counts. The Clearinghouse groups rows into 39,134 source clusters because a single breach often yields multiple filings. To build a filing-level outcome, we join related rows into firm-level episodes and attribute each episode to the affected issuer.

From the raw archive, sample filtering narrows the data into our two estimation panels (Table A2). Primary reading covers 55,705 cue-bearing filings (Appendix B.1). Across the initial 44,656 no-cue filings (40,192 in the ninety per cent partition and 4,464 in the ten per cent subsample), an expanded second screen detected candidate text in 33 filings and routed 87 into primary reading with zero coded as operating or supplying. The remaining 44,569 unread filings stay outside estimation. Among read filings, 55,445 have cues in Item 1 or Item 1A, and 260 filed in 2024–2026 carry cues in Item 1C alone. The complete-control sample of 29,692 filings retains 2008–2025 reports with

a clear yes or not-established operation label, a Compustat match, dated firm size, full covariates, prior-notification history, and a twelve-month window closed before the snapshot date.

Iterative singleton pruning then creates the estimation panels. Once filings that appear alone for their firm or calendar year exit the within-firm model, the full sample spans 28,536 filings across 4,484 firms. A dated historical three-digit SIC code selects the 28,266-filing main sample across 4,451 firms (99.05 per cent of full-sample filings). When absorbing industry-by-year cells, we retain singleton and single-role cells rather than trim post-estimation support.

TABLE A2. Sample Construction From the SEC Archive to the Estimation Panels

Stage of sample construction	Filings	Firms	Notified	Operating
Form 10-K filings in SEC EDGAR archive, 1994–2026	100,274	–	–	–
Unread filings without retrieved AI cue (after 87 audited, 0 AI roles)	44,569	–	–	–
Filings read by the primary model family	55,705	–	–	–
Retrieved passages in Item 1 or Item 1A	55,445	–	–	–
Retrieved passages in Item 1C only (filed 2024–2026)	260	–	–	–
Filed 2008–2025 with complete controls and closed 12-month window	29,692	–	–	–
Full estimation sample (after firm and year singleton removal)	28,536	4,484	869	2,455
Main estimation sample (with historical three-digit SIC code)	28,266	4,451	867	2,441
Main-sample filings with AI passages in Item 1C only	114	–	–	–

Notes: Notified counts filings followed within twelve calendar months by a qualifying regulatory breach notification naming the firm. Operating counts filings establishing active internal AI operation ($A_{if} = 1$). All 114 main-sample filings with AI passages confined to Item 1C were filed in 2024–2025. Dashes indicate archive-level screening stages prior to Compustat issuer linkage and outcome window construction.

Year-by-year counts in Table A3 highlight the late-panel expansion in disclosures. Annual counts of operating filings run between 2 and 13 per year through 2016 before rising to 241 in 2023, 652 in 2024, and 899 in 2025, when they make up one-third of all main-sample filings. Filings whose evidence sits solely inside the risk sections stay rare before 2021 (at most 8 annually) and reach 227 in 2024 and 435 in 2025. Comparison notification rates vary between 1.27 and 4.17 per cent across 2008–2023 before falling to 2.27 per cent in 2024 and 1.45 per cent in 2025. By contrast, operating filings post higher rates in eight of nine years from 2017 to 2025 (with 2019 the lone exception at 1.25 versus 3.72 per cent).

Variable definitions and measurement timing appear in Table A4. A word in a section length begins with a letter and continues with letters, digits, or hyphens. Blank or absent sections render text controls missing. Available sections with zero dictionary matches receive a zero rate. Each wording rate is standardised on the mean and standard deviation across the 29,692 complete-control filings, equal to 0.37 and 0.99 per 1,000 words for digital wording, 0.10 and 0.26 for data wording, and 16.76 and 17.44 qualifying sentences per 1,000 sentences for cyber discussion across every subsample.

The digital vocabulary tracks e-commerce, cloud computing and cloud services, digital and online platforms, channels, stores and portals, mobile applications, data warehouses, enterprise software, software as a service, digital transformation, and application programming interfaces.

TABLE A3. Sample Filings and 12-Month Breach Notification Rates by Filing Year

		Operating filings ($A_{if} = 1$)				Comparison filings ($A_{if} = 0$)		
Filing year	All filings	Filings	Risk-only	Notified	Rate (%)	Filings	Notified	Rate (%)
2008	1,092	4	0	0	0.00	1,088	21	1.93
2009	1,262	5	0	0	0.00	1,257	16	1.27
2010	1,251	6	0	1	16.67	1,245	17	1.37
2011	1,216	2	0	0	0.00	1,214	22	1.81
2012	1,231	4	0	0	0.00	1,227	27	2.20
2013	1,287	4	1	2	50.00	1,283	38	2.96
2014	1,322	10	2	1	10.00	1,312	37	2.82
2015	1,355	7	0	0	0.00	1,348	45	3.34
2016	1,411	13	3	1	7.69	1,398	55	3.93
2017	1,458	29	3	1	3.45	1,429	43	3.01
2018	1,519	57	4	8	14.04	1,462	61	4.17
2019	1,533	80	5	1	1.25	1,453	54	3.72
2020	1,597	107	8	7	6.54	1,490	58	3.89
2021	1,678	133	15	6	4.51	1,545	53	3.43
2022	1,686	188	30	11	5.85	1,498	39	2.60
2023	1,861	241	37	16	6.64	1,620	64	3.95
2024	2,814	652	227	44	6.75	2,162	49	2.27
2025	2,693	899	435	43	4.78	1,794	26	1.45
All years	28,266	2,441	770	142	5.82	25,825	725	2.81

Notes: Main estimation sample ($N = 28{,}266$ filings across 4,451 firms). Filings from 2025 (submitted through 2 July 2025) use 2025–2026 breach notifications recorded through 6 July 2026 to complete their twelve-month post-filing windows; 2026 filings lack a closed window and are excluded. Risk-only counts operating filings whose sole evidence of current AI operation appears in Item 1A or Item 1C. Rate (%) equals notified filings divided by total filings in that cell ($\times 100$).

The eleven data-vocabulary phrases cover customer data, personal information, personally identifiable information, PII, user data, consumer data, client data, patient data, payment card data, consumer information, and customer information. Cyber discussion counts Item 1A sentences passing a security-context screen that match one of 39 expressions across privacy, continuity, recovery, threats, and controls. Each sentence is counted once, omits Item 1C, and differs from the cybersecurity score of Florackis et al. (2023).

Firm size and industry assignment obey strict reporting lags. Firm size uses log total assets from the latest Compustat annual record whose accounting date plus 365 days falls strictly before the filing date and is at most 730 days old so that annual accounts have one full year to become public. Historical SIC codes come from the latest annual record whose accounting date is on or before both the vendor final date and the filing date (also at most 730 days old). Four-digit codes between 0100 and 9799 are truncated to three digits without backfilling from headers or older records. Both measures use Compustat's archival database rather than first-released figures.

A.2 Twelve-month breach windows and firm attribution

For filing date d, the outcome window spans the twelve calendar months $(d, d + 12 \text{ months}]$. Each episode receives a report interval defined by the notification dates naming the firm. A filing is coded

TABLE A4. Variable Definitions, Units, and Measurement Timing

Variable	Operational definition and scaling	Measurement window
Panel A: Outcome and AI-role indicators		
Breach notification (Y_{if})	Indicator ($\times 100$, pp) equal to one when at least one qualifying breach notification naming the firm has its report interval wholly inside $(d_f, d_f + 12$ months] (Appendix A.2)	Twelve months after filing date d_f
Internal AI operation (A_{if})	Indicator equal to one when primary classification establishes current company AI operation in internal workflows or company-operated customer services (Appendix B)	Filing date d_f
AI development or supply (S_{if})	Indicator equal to one when primary classification establishes ongoing or completed AI model development or commercial supply of AI products or services (can overlap A_{if})	Filing date d_f
Panel B: Firm scale, filing-text controls, prior breaches, and industry		
Firm size	Natural logarithm of Compustat total assets	Latest annual record with accounting date $+365$ days $< d_f$
Business-section length	$\log(1 +$ word count of Item 1$)$	Filing date d_f
Risk-section length	$\log(1 +$ word count of Item 1A$)$	Filing date d_f
Digital wording (s.d.)	Matches of 10 digital-technology terms per 1,000 Item 1 words, standardised	Filing date d_f
Data wording (s.d.)	Matches of 11 customer-data terms per 1,000 Item 1 words, standardised	Filing date d_f
Cybersecurity discussion (s.d.)	Item 1A sentences passing a security-context screen and matching 39 security terms per 1,000 Item 1A sentences, standardised	Filing date d_f
Prior breach notice	Indicator equal to one when a qualifying breach notice naming the firm falls wholly inside $[d_f - 12$ months, $d_f)$ (from 1 January 2007)	Twelve months before filing date d_f
Industry (SIC3)	Historical Compustat three-digit Standard Industrial Classification code	Latest annual record $\leq d_f$

Notes: Continuous wording controls are standardised using the mean and standard deviation across all 29,692 complete-control filings. Compustat accounting records are required to be at most 730 days old at the filing date d_f.

positive when at least one attributed episode's report interval falls wholly inside $(d, d + 12$ months]. Windows crossed by an interval that extends past the end of the horizon stay unresolved and exit estimation rather than count as zeros. Prior notification applies the identical rule to the twelve months before d.

Report intervals come directly from Clearinghouse notice dates. Across the 2,245 links between firms and episodes, 1,504 use the earliest report where the notifier matches the exact firm name, 27 use a source-verified subsidiary or former name, and 714 use the earliest report inside an indivisible source group. Notice that these dating rules govern timing rather than legal responsibility. In 2,238 of 2,245 links the interval is a single day, and the widest spans 102 days. Windows must close by 6 July 2026 (thirty days before the archive snapshot). Consequently, 2025 filings enter with outcome windows into 2026, and 2026 filings do not enter estimation. All qualifying notices cover own-system intrusions, external provider exposures, and physical or process failures regardless of AI involvement.

Two separate lags stand between an intrusion and archive entry. Among 1,785 sample-linked episodes with report dates, 1,152 also record breach start dates. The 537 episodes that began between 2015 and 2021 show a median discovery-to-report lag of 78 days across at least four and a half years of follow-up. Ingestion lag from agency publication to Clearinghouse entry is calibrated from our 13 March 2026 and 5 August 2026 snapshots. Of notices posted within two months before March, 47 per cent had not yet appeared in March against 9 per cent for notices at least three years old. Those weekly excess hazards predict that 1.7, 6.7, and 9.2 per cent of in-window reports are unrecorded for 2023, 2024, and 2025 filings (Appendix D.3).

A single breach notice can involve four distinct entities (the notifier, compromised provider, affected client, and SEC registrant). We attribute notifications to the company whose personal records were exposed across both direct own-system notices and client-impact notices from third-party provider breaches. Link provenance is tallied in Table A5. Of the 2,245 firm-episode links, 2,219 come from name matches across 2,221 source groups and 8,823 notices. Another 26 come from verified primary-source additions, and incorrect entity matches were dropped. Across the 2,157 unreviewed notifier-name links, 1,438 match exact notifier strings, 682 come from automated identifier matches, and 37 left the entity tie uncertain under review.

Upstream role tags exist for 88 pre-audit links (32 own-system intrusions, 41 provider breaches, and 15 process or employee disclosures). Those links combine into 2,241 episodes across 1,056 firms (four rest on two links each). Before primary-source inspection, the 87 notified operating windows and 75 comparison windows in 2024–2025 comprised 5 and 5 tagged provider breaches, 4 and 0 own-system hacks, 2 and 1 other tags, and 76 and 69 unreviewed notifier links. Our primary-source audit inspected all 162 windows (201 episodes, 209 filing-episode pairs, and 210 links) against 106 archived regulatory letters and 104 incident details (71 tabular state entries without narrative prose). That audit confirms 93 own-system links (58 operating, 35 comparison), 71 verified state tabular links (39, 32), 35 vendor links (16, 19), and 11 false homonyms (7, 4). Once false matches exit, 157 entity-verified windows remain (82 own-system, 49 tabular-only, and 26 pure vendor windows), which yield 51 own-system post-onset windows (1.04 pp, $[0.10, 1.99]$) and 87 entity-verified post-onset windows (1.16 pp, $[-0.02, 2.34]$). The 2024 Cencora (Lash Group) episode is the only breach that reaches multiple sample firms (4 operating and 5 comparison windows).

B. The measure

B.1 Coding rules and cue retrieval

Under the coding rules illustrated in Table A6, a filing counts as operating AI when its passages show that the issuer itself uses AI today, either in internal workflows or inside a live service run for customers. Completed active deployments qualify. Pilots, integrations still under way, future plans, and hypothetical possibilities do not. Machine learning, neural networks, expert systems, and similar named AI methods qualify without the words artificial intelligence. Algorithms, analytics,

TABLE A5. Links Between Sample Firms and Privacy Rights Clearinghouse Breach Episodes

Linkage source or archival attribution category	Firm–group links	Distinct episodes
Panel A: Linkage construction		
Baseline company-name matched Clearinghouse links	2,219	–
Added after manual primary-source review	26	–
Total firm–episode links and distinct episodes	2,245	2,241
Distinct linked issuers (firms)	–	1,056
Panel B: Upstream archival attribution basis (prior to complete 2024–2025 source audit)		
Notifier name matches issuer name exactly	1,438	–
Company-identifier match without manual role review	682	–
Primary sources reviewed, exact relationship unsettled	37	–
Subtotal: unreviewed notifier-name links	2,157	2,157
Own-system intrusion established in upstream review	32	30
External vendor/provider breach established in upstream review	41	39
Other role (e.g. employee or process disclosure) established	15	15

Notes: A link joins one sample issuer to one Privacy Rights Clearinghouse source group; an episode represents one breach at one issuer and can combine multiple source-group links (four episodes rest on two links each). All 26 manually added links and 62 of the 2,219 baseline links carried upstream role flags prior to the complete 2024–2025 archived-record audit reported in Table 5.

automation, or digitalisation count only when the text ties them explicitly to an AI method. Use by customers, competitors, or attackers, general industry commentary, and discontinued past use are excluded. Active AI use inside research tasks such as target selection or data analysis counts as operation. Development, training, or testing of an AI product for sale does not. Statements of active current use qualify wherever disclosed, including the risk sections.

Commercial AI development or supply covers cases where the firm builds AI models or sells or licenses AI technology, products, or services (Section 3). Adoption of an external vendor's AI product does not count, and leases of computing or data-centre capacity that buyers happen to use for AI are excluded as well. Outbound sales of software that customers operate themselves establish supply without internal operation. A company-operated AI service on customer inputs establishes both roles. Those overlapping rules generate the four role cells in Table 3 (operation only, both roles, commercial supply only, and neither role).

Each question takes three possible answers. Yes means that the retrieved passages establish the role. Not established means that the read text does not show the role without proving non-use. Unclear is recorded when actor, timing, or the boundary between operating and supplying cannot be resolved. Comparison filings consist of read filings marked not established on operation, and filings coded unclear on operation exit the sample. Eight third-reading examples in Table A6 show how those distinctions apply.

A broad keyword screen across business and risk sections determines which filings enter primary reading. One core list matches 69 named AI methods (including artificial intelligence, machine learning, neural network, natural language processing, computer vision, speech recognition, and random forest). A second core list matches seven word-stem pairs within eight words

TABLE A6. Application of the Classification Rules to Eight Contested or Agreement Passages

Verbatim quotation or passage summary	Third-model class	Decision rule illustrated
"in 2023 we implemented a skills intelligence system. This system, enabled by artificial intelligence, can detect known and hidden capabilities"	Operation	Internal workflow deployment currently operating
"This is all accomplished with a blend of human curation and an advanced AI engine that ultimately serves as the foundation of an easy-to-use workflow we branded as our Media Suite."	Operation; also supply	Live service operated for customers on the firm's own AI engine
"the continuing evolution of technology we use, including cloud-based computing, data hosting and artificial intelligence"	Operation	Affirmative statement of current use inside a risk factor
The firm designs and sells instruments with embedded machine learning that its customers operate on their own premises.	Not established	Outbound product supply without company operation
AI appears solely as a technique used by external attackers and as a subject of emerging regulation.	Not established	External threat actor and industry discussion excluded
The firm uses a predictive modelling capability for case management without naming any AI or machine-learning method.	Not established	Generic statistical analytics without named AI method
A risk factor warns that implementing new technologies such as AI, automation and algorithms may carry unintended consequences.	Unclear	Current operational deployment versus prospective risk unresolved
The firm's fraud-management software uses machine learning, without stating whether the firm operates it or licenses it to clients.	Unclear	Internal/managed operation versus outbound licensing unresolved

Notes: Entries in quotation marks reproduce verbatim excerpts cited by the third-family adjudicator; unquoted entries summarise filings for which no positive quotation was returned. Drawn from the 848 filings (768 contested plus 80 agreement) evaluated in the blinded third reading.

(artificial/intelligence, machine/learning, deep/learning, neural/network, natural/language, decision/tree, and pattern/recognition). Current terms in the third core list include acronyms AI, ML, NLP, and LLM, generative AI, expert systems, chatbots, and ChatGPT, Copilot, Watson, Claude, Gemini, and Llama. A fourth list of 29 weaker pairs such as big data, data analytics, and internet of things routes additional filings to reading yet stays tracked separately.

On the initial search, core AI cues appeared in 26,890 filings, weaker cues alone in 28,728, and no cue in 44,656. Weaker cues rarely yield positive operation and account for only 18 primary-reading positives among 28,529 dual-read weak-cue filings. A second screen across all 44,656 initial no-cue filings searched for literal acronyms, recommender systems, and conversational assistants. That screen detected candidate text in 30 of the 40,192 main-partition filings and 3 of the 4,464 tenth-partition filings (33 in total). None of the 87 initial-no-cue filings routed into the 55,705 primary read was classified as operating or supplying AI. The remaining 44,569 unread no-cue filings therefore stay outside estimation. Because unread filings lack text covariates and coded roles, unmeasured classification error across screening pools could shift pooled slopes.

B.2 Primary, second, and third model readings

The primary reading applies OpenAI GPT-6 across all 55,705 cue-bearing archive filings. Ninety per cent of the archive (90,246 filings) ran under one high-reasoning pass, while a stratified ten per cent sample (10,028 filings) combined two GPT-6 reasoning configurations so that supported positive answers in either pass were retained. Among the 2,441 sample operating filings, 2,161 rest on the primary pass alone, 227 on the supplementary pass, and 53 on both. Inter-model agreement against the independent second reader is virtually identical across partitions at 96.3 per cent ($\kappa = 0.809$) in the ten per cent sample (5,544 read filings) and 96.5 per cent ($\kappa = 0.814$) across the remaining 49,901 filings (Table A7).

Each filing entered the reader as numbered literal passage packets wrapped around retrieved cues with gap markers, company identifiers, and filing dates. A packet held at most 12,000 characters and each request contained up to eight packets. For every packet, the model coded current operation, application function, development or supply, plans, and context requests while citing supporting passage IDs under instructions barring outside knowledge or future events. Packets requesting additional text received one follow-up pass with surrounding paragraphs. The full instruction supplied before every packet reads as follows.

> Classify each supplied SEC filing evidence packet independently, at its filing date. These are literal selected blocks from that company's filing; gaps between blocks are omitted text, not continuous prose. Consider ALL supplied blocks together. A packet may be one part of a long filing: no_evidence means only not established in the supplied evidence, never non-adoption.
>
> AI includes explicit artificial intelligence, machine learning, neural/expert systems and equivalent descriptions. Generic algorithms, analytics, automation or digitalisation without an AI-specific link do not establish AI. Do not require the exact phrase artificial intelligence when an explicit equivalent is present.
>
> Return one result per packet id with these overlapping fields:
>
> current_operational_use = yes / no_evidence / unclear. yes requires actual current operational use or a completed deployment described as operating by the focal company itself, internally or in a functioning AI-enabled service that it operates for customers. Merely building/researching/testing an AI product, selling or licensing an AI tool that customers operate, customer/competitor/attacker use, future plans, permissions and generic risk language alone do not establish company operation. Using AI to perform R&D work, such as target selection or data analysis, can qualify as internal operational use; merely constructing/training/testing an AI product does not establish a separate operational use. Use unclear where the passage genuinely leaves company use versus development/supply unresolved. Past development or explicitly discontinued/historical-only use does not establish current deployment. Use unclear when temporal continuity is materially unresolved. A risk discussion can contain an affirmative current-use statement; retain it.
>
> use_function = internal / customer_service / both / unspecified / not_established. If current use is yes, internal means the company's own workflow; indirect customer benefit does not change this. customer_service means AI directly supplies the functioning service the company operates for customers. both requires evidence for both. unspecified means use established but function unavailable. Otherwise use not_established.

development_or_supply = yes / no_evidence / unclear. yes requires actual ongoing or completed company development/research into AI or supply of AI technology/products/services. A pure future intention establishes plans only, not actual development/supply. Using an AI tool does not itself imply developing it. Leasing GPUs, datacentres, compute or network capacity that customers might use for AI does not itself establish AI development/supply or operational use.

planned_or_hypothetical = true only for focal-company prospective, proposed, permitted or conditional AI activity. Competitors', customers', students' or attackers' plans do not qualify. Possible failure of AI already used, or hoped-for benefits of an existing AI activity, does not itself create a plan. Existing use and an additional actual company plan may coexist.

needs_context = true when actor, tense, operator-versus-supplier, antecedents or contradictory evidence prevents a confident material decision. Do not request context merely because these are selected excerpts or because no current use is supported. Retain established positive fields even if another field remains uncertain.

Evidence arrays contain block IDs from that SAME packet. use_evidence supports current use and function (up to 3); supply_evidence supports AI development/supply (up to 2); plan_evidence supports the issuer's prospective activity (up to 2); context_evidence locates the ambiguity (up to 2); basis locates the main relevant source (1 to 3). Positive fields require corresponding evidence; other evidence arrays may be empty. For unclear current use or development/supply, provide corresponding evidence AND set needs_context. context_reason = none / actor / timing / operator_or_supplier / antecedent / conflicting_evidence / insufficient_evidence. Use none iff needs_context=false. No quotations, prose, outside knowledge or prior-model answers.

PACKETS:

Microsoft-Decision-1, an independent model family, conducted the second reading across the same passages for all 55,445 Item 1 and Item 1A cue filings under instructions that barred outside reputation, industry stereotypes, later events, and notification outcomes. Answers take yes, not established, or unclear, and any packet that establishes a role marks the filing yes. Item 1C-only passages (260 archive filings, including 114 sample filings) were excluded from the second reading.

Year-by-year operation agreement in Table A7 exceeds 99 per cent annually through 2017 and stands at 90.1 per cent in 2025 and 88.6 per cent in 2026. Cohen's κ ranges between 0.77 and 0.83 across every year from 2017 onward, averaging 98.4 per cent ($\kappa = 0.783$) up to 2023 and 90.2 per cent ($\kappa = 0.780$) in 2024–2026. Agreement among core-AI-cue filings stands at 92.9 per cent. By contrast, weaker cues elicit near-universal agreement on non-operation.

Within the sample, the second reading answered 28,152 of the 28,266 filings and gives the same operating answer as the primary reading on 27,384 of them. The remaining 768 are the contested filings. In 443 cases the primary reading says operating and the second does not (166 in 2008–2023 and 277 in 2024–2025), 320 are of the reverse kind (185 and 135), and in 5 the second reading answered unclear. The filings that only the primary reading calls operating come from the largest firms, with mean log assets of 7.60 against 7.38 where both readings say operating, 6.68 where only the second does, and 6.89 where neither does. Those primary-only filings are also the most often notified (7.67 per cent against 5.42, 3.44, and 2.80 per cent).

TABLE A7. Inter-Model Agreement Between the Primary and Second Readers on AI Operation

Archive slice	Filings read	Primary yes	Second yes	Agreement (%)	Cohen's κ
Panel A: By filing year (2008–2026)					
2008	1,424	12	13	99.4	0.64
2009	1,433	12	15	99.2	0.59
2010	1,387	10	13	99.4	0.61
2011	1,402	7	12	99.4	0.60
2012	1,373	6	9	99.3	0.37
2013	1,441	10	15	99.1	0.50
2014	1,580	18	23	99.1	0.64
2015	1,645	20	28	98.9	0.63
2016	1,632	27	30	99.0	0.70
2017	1,660	43	42	98.9	0.78
2018	1,753	76	75	98.4	0.81
2019	1,793	118	122	97.2	0.78
2020	1,903	163	161	96.8	0.80
2021	2,144	241	240	95.9	0.80
2022	2,294	352	352	94.7	0.80
2023	2,155	341	337	95.3	0.83
2024	4,209	985	911	91.9	0.78
2025	4,382	1,384	1,247	90.1	0.77
2026	4,192	1,846	1,684	88.6	0.77
Panel B: By filing era (including 1994–2007 archive filings)					
Filings up to 2023	42,662	1,609	1,665	98.4	0.78
Filings from 2024 to 2026	12,783	4,215	3,842	90.2	0.78
All read filings (1994–2026)	55,445	5,824	5,507	96.5	0.81
Panel C: By keyword screening tier					
Core AI cue	26,829	5,806	5,448	92.9	0.79
Weak adjacent cue only	28,529	18	59	99.8	0.31
Expanded second-screen no-cue audit	87	0	0	100.0	–
Panel D: By primary-reader archive partition					
Stratified 10% dual-pass partition	5,544	616	551	96.3	0.81
Remaining 90% single-pass partition	49,901	5,208	4,956	96.5	0.81

Notes: Evaluates all 55,445 archive filings with retrieved passages in Item 1 or Item 1A read independently by OpenAI GPT-6 (primary) and Microsoft-Decision-1 (second). Filings from 1994–2007 enter Panels B–D. Agreement (%) is three-class exact match across yes, not established, and unclear; Cohen's κ is computed over the same three categories and omitted where neither model returns yes.

Anthropic Claude conducts the blind third reading across all 768 contested filings mixed with 80 random agreement filings (20 from each agreement cell across both eras) in separate batches without prior labels or outcomes. A positive answer requires a verbatim supporting quotation of at most 40 words, and every positive matches word for word (274 on operation and 375 on supply). On a 63-filing same-day repeat check, 57 received the exact same answer both times.

Third-reader adjudications in Table A8 resolve 480 contested filings (381 matching the primary reader and 99 matching the second reader) while leaving 288 unclear. For primary-only contested cases, the third reader sides with primary in 106 of 166 filings before 2024 compared with 97 of 277 in 2024–2025 (marking 122 unclear). It confirms 77 of 80 agreement controls and flags zero agreed negatives as operating. Contested filings ruled operating experience an 8.05 per cent notification

rate (19 of 236) against 2.05 per cent (5 of 244) for ruled non-operators (Fisher's exact $p = 0.003$) and 7.64 per cent (22 of 288) for unclear cases.

TABLE A8. Blinded Third-Model Adjudication of Contested and Agreement Filings

		Third-model classification on operation		
Primary and second reader stratum	Filings	Operation	Not established	Unclear
Panel A: All 768 contested main-sample filings				
Primary yes, second not established (2008–2023)	166	106	9	51
Primary yes, second not established (2024–2025)	277	97	58	122
Primary not established, second yes (2008–2023)	185	21	113	51
Primary not established, second yes (2024–2025)	135	11	60	64
Second reader coded unclear	5	1	4	0
All contested filings	768	236	244	288
Subsequent 12-month breach rate (%)[a]	–	8.05	2.05	7.64
Panel B: Blinded agreement benchmark filings (80 randomly mixed draws)				
Both readers yes (40 drawn of 1,991)	40	38	1	1
Both readers not established (40 drawn of 25,393)	40	0	39	1

Notes: Anthropic Claude evaluated all 768 contested sample filings mixed blindly with 80 agreement filings (20 per agreement cell in each period), seeing neither earlier model labels nor breach outcomes. Every positive classification required a verbatim quotation (≤ 40 words) verified in the filing text.
[a] Corresponds to 19/236, 5/244, and 22/288 notified filings (Fisher's exact $p = 0.003$ comparing 8.05% vs. 2.05%).

Table A9 and Figure A1 refit the main model across all nine reading rules and the within-firm model across eight rules. Pooled estimates span 0.83 to 2.15 and reach their lowest point when the second reader decides operation by shifting large, frequently notified contested filings out of the operating group. Inside 2024–2025, every reading produces an estimate between 1.65 and 2.98 with every interval above zero. Pre-2024 slopes lie between 0.08 and 2.04 and span zero. Within-firm estimates run from -0.85 to 0.76 and all include zero.

B.3 Masked rereading, risk-section passages, and firm size

Because readers saw issuer names and filing dates, an external model could in principle draw on reputation or later breaches despite explicit instructions banning outside knowledge. To test how far classifications depend on identity, Anthropic Claude reread 40 filings after stripping identifiers and dates and replacing firm names with "the Company". Those 40 draws comprise 8 from each of the four contested cells across both eras and 8 agreement filings (all non-notified), split across two 20-filing sessions under standard third-reading rules.

Masked labels match the unmasked readings closely in Table A10. Across all 40 filings, exact three-class agreement stands at 82.5 per cent (33 of 40) and binary yes-versus-not-yes agreement reaches 90.0 per cent (36 of 40). Agreement is 25 of 32 on contested filings and 8 of 8 on controls. Every change involves an unclear label (four yes moved to unclear, two unclear moved to not established, and one not established moved to unclear), with zero direct reversals between yes and not established.

TABLE A9. Operating Coefficient Across All Nine Multi-Model Classification Rules

Classification rule	Filings [Op.]	Pooled	2008–2023	2024–2025	Within firms
1. Primary reader, all sample filings (baseline)	28,266	1.99***	1.84	2.55***	0.74
	[2,441]	(0.70)	(1.27)	(0.80)	(1.00)
2. Primary reader, filings read by second model	28,152	2.01***	1.84	2.57***	0.76
	[2,435]	(0.71)	(1.27)	(0.81)	(1.01)
3. Both models agree on operation	27,384	1.30*	0.72	2.37**	−0.85
	[1,991]	(0.78)	(1.31)	(0.94)	(1.11)
4. Both models agree on both roles	26,769	1.81**	1.16	2.98***	–
	[1,727]	(0.85)	(1.40)	(1.06)	
5. Second model for both roles	28,054	0.99	0.08	2.00**	−0.52
	[2,291]	(0.68)	(1.08)	(0.85)	(0.86)
6. Second model on operation, primary on supply	28,147	0.83	0.19	1.65**	−0.63
	[2,311]	(0.67)	(1.07)	(0.83)	(0.88)
7. Third model on contested, unclear omitted	27,978	1.88**	1.54	2.75***	−0.12
	[2,233]	(0.77)	(1.26)	(0.91)	(1.11)
8. Third model on contested, keep primary on unclear	28,266	2.15***	2.04	2.72***	0.73
	[2,406]	(0.72)	(1.25)	(0.85)	(1.04)
9. Third model on contested, unclear as comparison	28,266	1.75**	1.29	2.53***	−0.15
	[2,233]	(0.77)	(1.28)	(0.89)	(1.09)

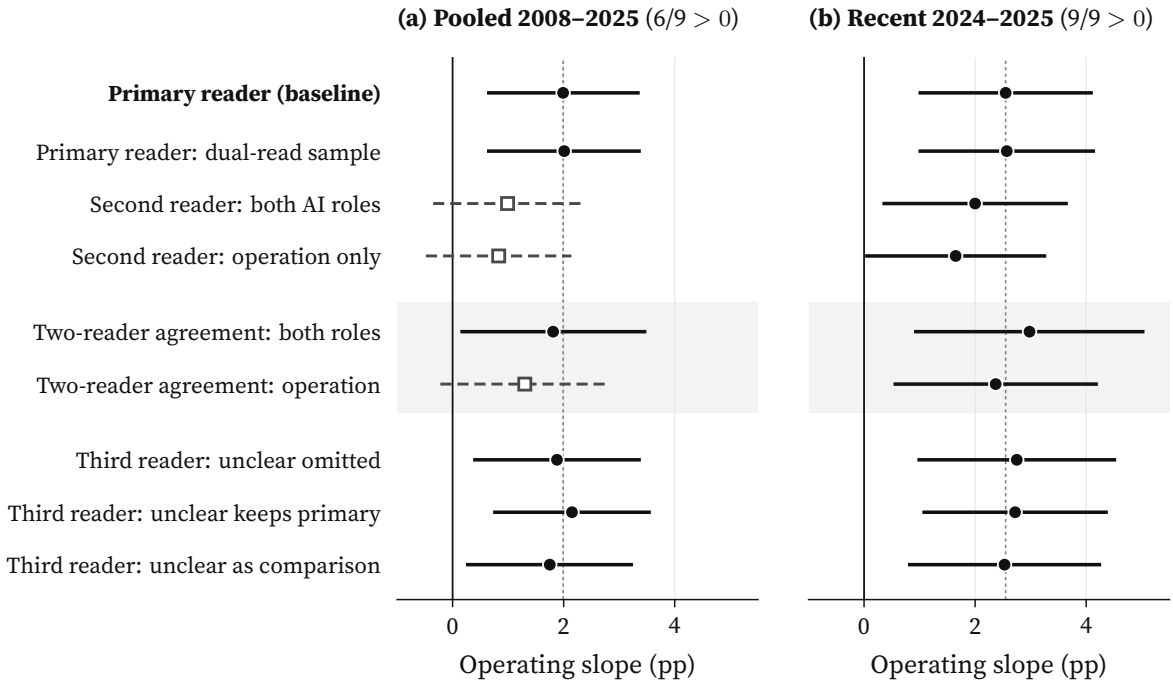
Notes: Linear probability estimates on internal AI operation (β, in percentage points), with firm-clustered (CR1) standard errors in parentheses and operating filing counts in brackets under Filings. Pooled estimates equation (1) on the main sample; 2008–2023 and 2024–2025 come from the fully interacted period model (3); Within firms estimates firm and year fixed effects (2) on the full sample ($27{,}567$–$28{,}536$ filings; omitted in row 4). In rows 7–9, the third-family model adjudicates the 768 contested filings while supply retains the primary label. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A10. Identity-Masked Rereading Against Unmasked Third-Model Classifications

	Identity-masked rereading			
Unmasked third-model class	Operation	Not established	Unclear	Total
Operation	6	0	4	10
Not established	0	12	1	13
Unclear	0	2	15	17
Total	6	14	20	40

Notes: Cross-tabulates 40 filings (8 drawn with a fixed seed from each of the four contested cells plus 8 agreement filings; 0 breach-notified) re-evaluated with company names replaced by "the Company" and identifiers and dates stripped. Three-class agreement is 33/40 (82.5%); binary yes-versus-not-yes agreement is 36/40 (90.0%), with zero reversals between `Operation` and `Not established`.

FIGURE A1. Operating associations across all nine multi-reader coding rules



Notes: Operating coefficients and 95 per cent firm-clustered intervals (in percentage points) across the nine classification rules reported in Table A9 for the pooled 2008–2025 sample in panel (a) and the 2024–2025 period specification in panel (b). Filled black circles with solid stems denote 95 per cent intervals above zero (6 of 9 pooled rules and all 9 of 9 recent rules); open squares with dashed stems denote intervals spanning zero. Vertical dashed lines mark the primary-reader estimates (1.99 and 2.55 percentage points).

Of the 2,441 operating filings, 770 at 536 firms establish current operation solely in the risk sections (699 in Item 1A alone, 58 in Item 1C alone, and 13 in both). Most appear recently, with 662 in 2024–2025 (42.7 per cent of that cohort's operating filings) against 108 in 2008–2023. These filers are larger than business-section operators or comparisons (mean log assets 8.21 versus 7.06 and 6.88, with 24.5 per cent in the top size decile versus 10.1 and 9.6 per cent). They also lean towards manufacturing (30.5 per cent) and finance (25.2 per cent) rather than technology (13.4 per cent). Agreement between primary and second readers stands at 69.0 per cent here (237 of 764 dual-read cases contested, with 236 second-read not established and one unclear) versus 87.6 per cent in business-section operators and 98.7 per cent in comparisons.

Blinded passage reading asks whether these risk-section texts declare active AI use or vague warnings. Anthropic Claude inspected the operating excerpts across all 770 risk-only filings randomly mixed with 120 business-section operators (1,068 masked packets averaging 234 words and at most 266 words). Text anonymisation replaced company names with "the Company" and removed CIKs, dates, section labels, and outcomes. Automated checks confirmed zero leaks for CIKs and filing years, though a handful retained inline domain strings or residual brand text. The

reader sorted filings into states operation (S, active internal workflows or live customer services), mentions AI only (M, risk, plan, possibility, trend, or external use), or unclear (U).

Illustrative passages from the blinded audit in Table A11 clarify that distinction. "We have incorporated, and may continue to incorporate, certain AI solutions into our operations" qualifies as stating operation (S). By contrast, the warning that "our implementation of certain new technologies, such as those related to artificial intelligence, automation and algorithms, in our business processes may have unintended consequences" mentions AI strictly as a risk (M).

Table A11 sets out the resulting distribution. Among the 770 risk-section-only filings, 664 state active operation (86.2 per cent, Wilson interval 83.6 to 88.5), 47 mention AI only, and 59 are unclear. Among the 120 business-section controls, 115 (95.8 per cent) state operation and zero mention AI only. A blinded second pass within the same model family on a random one-fifth subset (178 filings) agrees on S versus not-S for 94.9 per cent ($\kappa = 0.81$) and across all three classes for 91.6 per cent ($\kappa = 0.69$). Most risk-section-only operating disclosures thus assert active current workflows explicitly.

TABLE A11. Blinded Passage Audit of Risk-Section-Only and Business-Section Operating Filings

Operating-filing group	Filings	States op. (S)	Mentions only (M)	Unclear (U)	Share S (%)
Panel A: Blinded passage-audit counts and active-statement shares					
Business section (Item 1), random 120	120	115	0	5	95.8
Risk sections only, 2008–2023	108	92	9	7	85.2
Risk sections only, 2024–2025	662	572	38	52	86.4
Risk sections only, all years	770	664	47	59	86.2
Panel B: Subsequent 12-month breach notification rate (%)					
Risk sections only, all years[a]	770	7.53	17.02	8.47	–

Notes: Blinded audit of all 770 risk-section-only operating filings mixed with 120 randomly drawn Item 1 operating filings (issuer names replaced with "the Company" and dates, section headers, and outcomes suppressed). Class S requires an explicit statement of current AI operation; M denotes prospective, risk-only, industry, or external-actor mentions; U denotes unresolved passages.

[a] Corresponds to 50/664 (S), 8/47 (M), and 5/59 (U) notified filings, compared with 4.73% across all 1,671 Item 1 operating filings and 2.81% across all 25,825 comparison filings.

Notification rates rise from business-section operating filings (4.73 per cent of 1,671) to risk-section statements of operation (7.53 per cent of 664) and to risk-section mentions or unclear passages (12.26 per cent of 106), against 2.81 per cent for comparison filings. In the location model with operation split three ways, risk-section statements carry 2.92 [0.81, 5.03] and mentions or unclear passages 6.20 [0.25, 12.16], against 0.93 [−0.71, 2.58] for business-section operation. The difference between the two risk-section classes is −3.29 [−9.67, 3.09]. Because the interval on the 106 mention-only or unclear passages is wide, this comparison confirms explicit statements of current operation across most risk-section-only filings without separating operational wording from risk-section placement. In 2024–2025 alone, statements carry 3.60 [1.28, 5.92] and mentions or unclear passages 5.42 [−1.05, 11.90], and statements exceed business-section operation by 2.95 [0.26, 5.63].

To test whether nonlinear firm size accounts for risk-section sorting, we add nine size-decile indicators alongside linear log total assets to the three-way location specification while absorbing three-digit industry-by-year fixed effects. Risk-section statements carry 2.63 [0.53, 4.72] across all years and 3.50 [1.19, 5.80] in 2024–2025, against 0.92 [−0.63, 2.47] for business-section operation and 5.33 [−0.59, 11.26] for mentions or unclear passages.

Raw notification rates by fifth of firm size display the same ordering (Table A12, panel B). Risk-section-only operating filings show higher breach rates than comparison filings in every fifth and climb from 1.59 versus 0.51 per cent in the smallest fifth to 14.24 versus 8.51 per cent in the largest. Those filings also exceed business-section operating filings in four of the five fifths, with only the middle fifth reversed (2.34 versus 3.71 per cent).

TABLE A12. Disclosure-Location Regressions With Size Deciles and Breach Rates by Asset Quintile

Panel A: Three-way disclosure-location regressions ($100 \times Y_{if}$, percentage points)

	Baseline controls	Plus asset-decile indicators	
Disclosure category	All years	All years	2024–2025
Business-section operation (Item 1)	0.93	0.92	0.87
	(0.84)	(0.79)	(0.89)
Risk-section active statement (S)	2.92***	2.63**	3.50***
	(1.08)	(1.07)	(1.18)
Risk-section mention or unclear (M/U)	6.20**	5.33*	4.42
	(3.04)	(3.02)	(3.26)

Panel B: Raw 12-month breach notification rate (%) by sample quintile of log total assets

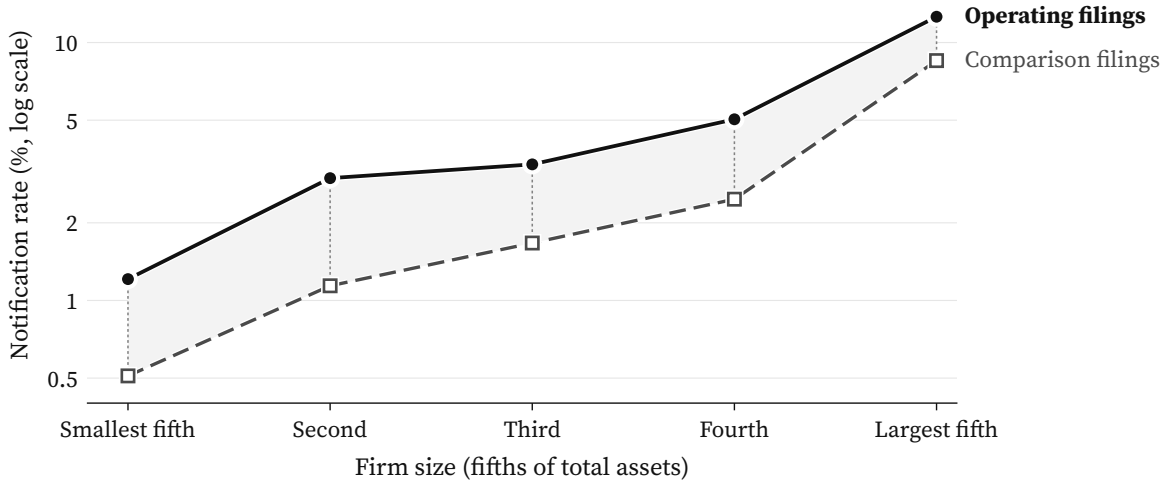
Filing group	Q1 (smallest)	Q2	Q3	Q4	Q5 (largest)
Comparison filings ($A_{if} = 0$)	0.51	1.14	1.67	2.47	8.51
	[5,323]	[5,183]	[5,148]	[5,177]	[4,994]
Business-section operating (Item 1)	1.12	2.18	3.71	4.64	11.20
	[268]	[367]	[377]	[302]	[357]
Risk-section-only operating (Items 1A/1C)	1.59	5.83	2.34	5.75	14.24
	[63]	[103]	[128]	[174]	[302]

Notes: Panel A reports linear probability coefficients ($N = 28{,}266$ filings; 4,451 firms; three-digit SIC × year fixed effects) with firm-clustered (CR1) standard errors in parentheses; column 3 comes from interacting all covariates and nine asset-decile indicators with 2008–2023 and 2024–2025. Panel B reports raw twelve-month notification percentages across pooled sample quintiles of log total assets, with cell filing counts in brackets underneath. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B.4 Human validation design and correction estimators

Two independent research assistants blinded to breach outcomes and automated labels coded a thirteen-stratum probability sample of 350 masked filings drawn from the 28,266-filing main sample. Once one worked training example in stratum S01 is excluded, $n_{\text{eval}} = 349$ evaluation filings remain (22 notified, with 9 in 2008–2023 and 13 in 2024–2025 across 19 automated operating and 3 automated comparison draws). Exact sampling probabilities deliver positive inclusion support for 28,158 of 28,266 filings (99.62 per cent) and omit 108 comparison filings in S13 outside the secondary pool.

FIGURE A2. Breach notifications in the twelve months after a filing, by firm size



Notes: Share of filings followed within twelve months by a notification naming the firm for operating filings (solid black, filled circles) and comparison filings (dashed dark grey, open squares), by fifth of total assets across the 28,266 filings. Raw rates on a logarithmic vertical scale; conditional estimates with three-digit industry by year effects and controls are in Table 4 and Table A12.

Across the 349 evaluation filings, 46 inter-coder disagreements between Coder 1 and Coder 2 were resolved uniformly to Coder 2. Coder 1 alone returns nearly identical calibrated slopes (Table 6). That resolution yields unweighted counts of 124 A (internal operation only), 64 B (commercial supply only), 31 C (both roles), 76 D (neither role), and 54 E (nominalised AI risk warning only). Design-weighted prevalences equal 6.79 per cent for strict human operation ($A^{\text{strict}} \in \{\text{A}, \text{C}\}$), 10.26 per cent for inclusive operation ($A^{\text{incl}} \in \{\text{A}, \text{C}, \text{E}\}$), and 5.17 per cent for human supply ($S^{\text{hum}} \in \{\text{B}, \text{C}\}$). Strict operating agreement between machine and human labels stays balanced across notified draws (97.93 per cent weighted and 72.73 per cent unweighted across 22 draws) and non-notified draws (97.95 per cent weighted and 76.45 per cent unweighted across 327 draws). Joint four-state agreement likewise sits at 96.37 versus 94.62 per cent weighted and 59.09 versus 60.55 per cent unweighted.

Across the first seven validation strata, post-stratified sampling weights (N_h/n_h) equal S01 contested 2024–2025 risk-section operation ($N = 206, n = 49, w = 4.20$), S02 contested 2008–2023 risk-section operation ($37, 20, 1.85$), S03 agreed 2024–2025 risk-section operation ($456, 50, 9.12$), S04 agreed 2008–2023 risk-section operation ($71, 20, 3.55$), S05 contested business-section operation only ($42, 15, 2.80$), S06 agreed business-section operation only ($277, 25, 11.08$), and S07 contested business-section dual role ($165, 20, 8.25$). Across the remaining six strata, weights equal S08 agreed business-section dual role ($1{,}187, 25, 47.48$), S09 contested supply only ($195, 30, 6.50$), S10 contested comparison ($129, 35, 3.69$), S11 agreed supply only ($575, 23, 25.00$), S12 agreed 2008–2023 comparison ($21{,}312, 18, 1{,}184.00$), and S13 agreed 2024–2025 comparison ($3{,}614, 19, 190.21$). Inside S12 (75.4 per cent of main filings and a 2.94 per cent notification rate), all 18 draws were coded D.

That count implies a 15.33 per cent one-sided 95 per cent exact binomial upper bound above the 10.27 per cent unadjusted tipping share (2,188.5 unrecognised non-notified operators).

Across the full frame, Category E totals 980.2 design-weighted filings, of which 82.9 per cent lie in machine-negative comparison strata. Specifically, S13 holds 77.6 per cent of weighted E mass (lifting human operation share from 0.000 strict to 0.211 inclusive at a 1.94 per cent stratum breach rate). Stratum S10 moves from 0.171 to 0.571 at 6.98 per cent. By contrast, contested risk stratum S01 rises from 0.306 to 0.694 (8.25 per cent breach rate). Agreed business-section strata S06 (1.000, breach rate 6.14 per cent) and S08 (0.840, 4.30 per cent) stay unchanged. Weighted validation-sample breach rates stand at 6.44 per cent (unweighted 13/155) on A/C, 2.22 per cent (unweighted 4/54) on E, and 4.72 per cent (unweighted 5/140) on B/D pooled (7.16, 2.25, and 0.25 per cent in 2024–2025). That stratum composition explains why our pre-specified expectation that strict calibration would equal or exceed inclusive calibration failed.

Under two-phase regression calibration (Chen, Hong, and Tamer 2005), outcome surrogacy requires $E[Y_{if} \mid H_{if}, W_{if}, X_{if}] = E[Y_{if} \mid H_{if}, X_{if}]$. The first-stage model for $E[H_{if} \mid W_{if}, X_{if}]$ is approximated by stratum means $\hat{H}_{if}$ at baseline and augmented with X_{if} in covariate-conditional calibration. Under the one-step orthogonal score correction (Robins, Rotnitzky, and Zhao 1994, Egami et al. 2023), let $\tilde{y}$ and $\tilde{Z}_{\text{cal}}$ denote the twelve-month outcome and stratum-calibrated role vector $(\hat{A}_{if}, \hat{S}_{if})$ residualised against X_{if} and three-digit industry-by-year fixed effects, with residual-maker diagonal m_i and within-stratum role covariance matrix $\Sigma_{\text{within}} = \sum_i m_i \widehat{\text{Cov}}(H_{if} \mid \text{stratum})$. At the initial estimate $\hat{\beta}_0 = J^{-1}\tilde{Z}_{\text{cal}}^{\top}\tilde{y}$ with positive-semidefinite variance-adjusted Jacobian $J = \tilde{Z}_{\text{cal}}^{\top}\tilde{Z}_{\text{cal}} + \Sigma_{\text{within}}$, let $\Delta_s = H_s - \hat{H}_s$ denote the human-minus-calibrated role residual on validated draw s carrying stratum-rescaled weight w_s. Exact inverse inclusion probabilities enter on the 28,158-filing positive-support frame in group 3 of Table 6. Post-stratified weights N_h/n_h extrapolate across the 108 zero-support S13 filings in groups 1 and 2. Table 6 reports $\hat{\beta}_{\text{DSL}} = \hat{\beta}_0 + J^{-1}\Delta_{\text{score}}$ together with the direct full-system solution in table note c, where

$$\Delta_{\text{score}} = \sum_s w_s \Delta_s \tilde{y}_s - \left[\sum_s w_s(\Delta_s \tilde{Z}_{\text{cal},s}^{\top} + \tilde{Z}_{\text{cal},s}\Delta_s^{\top}) + \sum_s w_s m_s(\Delta_s\Delta_s^{\top} - \widehat{\text{Cov}}(H_s \mid \text{stratum}))\right]\hat{\beta}_0.$$

When strict two-phase regression calibration is applied jointly to the source-audited 2024–2025 outcomes in Table 5, operating slopes (and operation-minus-supply contrasts) equal 2.76 (3.86) on all 157 entity-verified windows, 2.89 (3.54) without vendor breaches, and 1.98 (2.59) on the 82 confirmed own-system windows. Under strict post-onset filters, calibrated slopes equal 1.45 (2.50) on the 87 entity-verified post-onset windows and 1.22 (2.30) on the 51 confirmed own-system post-onset windows.

C. Full estimates

This appendix presents full slope estimates for the models in Section 4, period interactions, multiplicity families, connected-firm clustering, and rare-event logit checks. All estimates are expressed

in percentage points of the twelve-month notification rate, with firm-clustered standard errors in parentheses and 95 per cent confidence intervals in brackets unless a table note specifies otherwise.

C.1 Main and period regression estimates

Table A13 places the main specification beside two full-sample regressions. Column (1) absorbs broad industry and year effects across all 28,536 full-sample filings, column (2) adds firm fixed effects across those same filings, and column (3) reports the baseline main sample of 28,266 filings. Prior breach notification carries the largest slope across both pooled models at 24.55 and 21.97 points against roughly two points for internal AI operation. Firm size adds roughly one point per log unit of total assets. Section lengths, digital and data prose, and cybersecurity discussion stay within half a point of zero in every specification.

The control sequence indicates that prior-notification and cyber-discussion controls drive the covariate adjustment. On the full sample with broad industry and year effects, operation sits at 2.61 [1.15, 4.08] before cyber discussion and prior notification enter and 2.02 [0.77, 3.26] after they enter. On the main sample, the same controls with broad effects yield 2.00 [0.75, 3.25]. Three-digit industry-by-year effects produce the baseline estimate of 1.99 [0.62, 3.37]. Finer industry cells therefore move the slope by 0.01 points. Prior notifications and cyber discussion move it by 0.59 points. Within firms, those two controls shift operation from 0.73 [−1.22, 2.67] to 0.74 [−1.23, 2.71].

Inside the within-firm specification, 820 firms change operating status, 464 change notification outcomes, and 188 change both. Firm and year effects retain 38.9 per cent of centred variation in internal AI operation. Conditional variance inflation between operation and commercial supply equals 1.49 in the pooled model and 1.27 within firms, which confirms that role collinearity does not explain the wide within-firm interval. Inside firms, the prior-notification slope sits at −2.99 [−6.66, 0.68] and the size slope equals 0.50 [0.02, 0.98], against 24.55 and 1.04 in the pooled regression on the exact same filings.

The period specification lets every slope vary between 2008–2023 and 2024–2025. That joint model retains all main-sample filings with three-digit industry-by-year effects so that a single firm-clustered covariance accounts for firms that file in both regimes. Table A14 reports the eighteen era-specific slopes in panel A and linear contrasts in panel B. Internal AI operation moves from 1.84 [−0.65, 4.33] in 2008–2023 to 2.55 [0.98, 4.12] in 2024–2025, a noisy difference of 0.71 [−2.14, 3.55] ($p = 0.63$). Operation minus supply equals 1.87 [−1.87, 5.62] in 2008–2023 and 3.05 [0.17, 5.93] in 2024–2025.

Covariate slopes also shift between eras. We therefore interact every control across regimes. The prior-notification slope moves from 23.54 to 17.26 points and the size slope falls from 1.18 to 0.91. Both section-length coefficients carry intervals at least three times wider in 2024–2025, when the cohort holds 5,507 filings against 22,759 before. The interacted regression has 24,891 residual degrees of freedom.

TABLE A13. All Covariate Slopes of the Baseline Model and Full-Sample Companions

Covariate	Full sample: Broad industry & year FE	Full sample: Firm & year FE	Main sample: Industry × year FE
Internal AI operation	2.02***	0.74	1.99***
	(0.63)	(1.00)	(0.70)
AI development or supply	−0.07	0.51	−0.23
	(0.55)	(0.89)	(0.60)
Firm size (log total assets)	1.04***	0.50**	1.12***
	(0.09)	(0.24)	(0.11)
Business-section length (log words)	0.06	−0.05	−0.08
	(0.18)	(0.19)	(0.17)
Risk-section length (log Item 1A words)	0.00	−0.03	−0.10
	(0.07)	(0.09)	(0.07)
Digital wording (s.d.)	−0.05	−0.05	−0.12
	(0.09)	(0.24)	(0.11)
Customer-data wording (s.d.)	0.33*	−0.20	0.05
	(0.19)	(0.27)	(0.15)
Cybersecurity discussion (s.d.)	0.23	−0.07	0.14
	(0.18)	(0.26)	(0.19)
Prior 12-month breach notice	24.55***	−2.99	21.97***
	(2.85)	(1.87)	(2.76)
Observations (filings)	28,536	28,536	28,266
Firms (clusters)	4,484	4,484	4,451
Mean of dependent variable (%)	3.05	3.05	3.07
Adjusted R^2	0.097	0.195	0.131

Notes: Linear probability coefficients ($100 \times Y_{if}$, percentage points) with firm-clustered (CR1) standard errors in parentheses (t_{G-1} inference). Digital wording, customer-data wording, and cybersecurity discussion are standardised on the 29,692 complete-control filings prior to singleton removal. Intercepts and absorbed fixed effects are omitted. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

C.2 Clustering, multiplicity, and logit comparisons

We cluster by firm because a firm's filings share systems, providers, and reporting practices across years. Let $\widetilde{X}$ and $\widetilde{R}$ denote the regressors and outcome demeaned within three-digit industry-by-year cells, K the number of slopes, L the number of absorbed cells, N the number of filings, and G the number of clusters. With score $s_g = \sum_{f \in g} \widetilde{X}_f \hat{\varepsilon}_f$ for cluster g, the covariance is

$$\hat{V} = \frac{G}{G-1}\,\frac{N-1}{N-K-L}\,(\widetilde{X}'\widetilde{X})^{-1}\Big(\sum_{g} s_g s_g'\Big)(\widetilde{X}'\widetilde{X})^{-1}, \tag{5}$$

and intervals use a t distribution with $G-1$ degrees of freedom. The main model has nine slopes, 3,357 absorbed cells, 24,900 residual degrees of freedom, and 4,451 firm clusters. Those dimensions give a finite-sample factor of 1.135. A contrast such as operation minus supply uses $\hat{V}_{\beta\beta} + \hat{V}_{\theta\theta} - 2\hat{V}_{\beta\theta}$.

Firms can also share a breach. We therefore link firms that primary records tie to a common episode or corporate group and treat each connected component as a cluster. That grouping leaves 4,439 connected clusters rather than 4,451 firms. Connected-group clustering keeps the primary

TABLE A14. Period-Interacted Model: Era-Specific Slopes and Cross-Role / Cross-Period Contrasts

Panel A: Era-specific covariate slopes ($N = 28{,}266$ filings; 4,451 firms; Adjusted $R^2 = 0.132$)

Covariate	2008–2023	2024–2025
Internal AI operation	1.84	2.55***
	(1.27)	(0.80)
AI development or supply	−0.03	−0.50
	(0.83)	(0.89)
Firm size (log total assets)	1.18***	0.91***
	(0.12)	(0.15)
Business-section length (log words)	−0.06	−0.17
	(0.17)	(0.60)
Risk-section length (log Item 1A words)	−0.11	−0.27
	(0.07)	(0.57)
Digital wording (s.d.)	−0.12	−0.15
	(0.13)	(0.18)
Customer-data wording (s.d.)	0.04	0.04
	(0.19)	(0.22)
Cybersecurity discussion (s.d.)	0.18	0.10
	(0.25)	(0.25)
Prior 12-month breach notice	23.54***	17.26***
	(3.13)	(3.30)

Panel B: Linear contrasts from the joint firm-clustered covariance matrix

Contrast	Estimate (pp)	Std. error	95% CI	p-value
Operation minus supply, 2008–2023	1.87	(1.91)	[−1.87, 5.62]	0.327
Operation minus supply, 2024–2025	3.05**	(1.47)	[0.17, 5.93]	0.038
Operation, 2024–2025 minus 2008–2023	0.71	(1.45)	[−2.14, 3.55]	0.626
Supply, 2024–2025 minus 2008–2023	−0.46	(1.16)	[−2.74, 1.81]	0.688

Notes: Single linear probability regression on all 28,266 main-sample filings (4,451 firms; 3,357 three-digit SIC × filing-year cells) in which every covariate is interacted with indicators for 2008–2023 (22,759 filings) and 2024–2025 (5,507 filings). Firm-clustered (CR1) standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

estimate at 1.99 and moves its interval to [0.61, 3.37], with a standard error of 0.704 against 0.702 under firm clustering.

Two step-down Holm families (Holm 1979) adjust for multiple hypothesis tests across our focal breach-notification models (Table A15). Panel A adjusts across the two focal contrasts in equation (1), internal operation (β) and internal operation minus commercial supply ($\beta - \theta$). Operation carries a Holm-adjusted p-value of 0.009. The operation-minus-supply contrast carries 0.053. Panel B adjusts across the four operating specifications on the 28,536-filing full sample that combine pooled broad-industry and within-firm regressions before and after cybersecurity discussion and prior-year breach controls enter. Both pooled operating slopes remain statistically distinguishable from zero after adjustment (Holm p of 0.002 and 0.004). Both within-firm slopes stay inconclusive (0.922). Period, location, reading-variant, and sensitivity regressions in Appendix D and Appendix E report pointwise 95 per cent intervals.

An exact logit with the main model's three-digit industry-by-year intercepts has no finite estimate. Of the 3,357 cells, 2,857 contain only filings without a notification and 27 contain only filings

TABLE A15. Holm Step-Down Multiplicity Adjustments for Focal Breach-Notification Contrasts

Specification	Focal contrast	Estimate	Std. error	95% CI	Raw p	Holm p
Panel A: Focal role contrasts in the main sample ($N = 28{,}266$; three-digit SIC × year FE)						
Baseline model	Internal AI operation	1.99***	(0.70)	[0.62, 3.37]	0.0045	0.009
Baseline model	Operation minus supply	2.23*	(1.15)	[−0.03, 4.48]	0.0527	0.053
Panel B: Operating slopes across full-sample specifications ($N = 28{,}536$; 4,484 firms)						
Pooled broad FE, base controls	Internal AI operation	2.61***	(0.75)	[1.15, 4.08]	0.0005	0.002
Within-firm FE, base controls	Internal AI operation	0.73	(0.99)	[−1.22, 2.67]	0.4633	0.922
Pooled broad FE, added controls	Internal AI operation	2.02***	(0.63)	[0.77, 3.26]	0.0015	0.004
Within-firm FE, added controls	Internal AI operation	0.74	(1.00)	[−1.23, 2.71]	0.4609	0.922

Notes: Estimates in percentage points ($100 \times Y_{if}$), with firm-clustered (CR1) standard errors in parentheses and 95 per cent t_{G-1} confidence intervals in brackets. "Added controls" enters cybersecurity discussion and prior-year breach notification alongside log total assets, section lengths, digital wording, and customer-data wording. Holm p-values apply Holm (1979) step-down family-wise error adjustment separately within Panel A (2 tests) and Panel B (4 tests). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$ (raw two-sided).

with one, together spanning 13,398 filings whose unconditional intercepts diverge. Appendix E (Table A22) pairs the broad-industry logit below with concentrated three-digit industry-by-year exponential pseudo-maximum-likelihood models across the 473 non-separated cells (14,868 filings) and a full-sample complementary log-log check. On the full sample with broad industry and year effects, logit converges with an average discrete difference of 1.63 [0.55, 2.71] for internal operation under a firm-clustered delta-method interval. The linear model on the same 28,536 filings yields 2.02, with 5,707 fitted values below zero and none above one.

D. Sensitivity

D.1 Omissions and defensive security-AI exclusions

We ranked every firm by the absolute first-order influence of its filings on the operation coefficient and refitted the main model without each of the three most influential firms individually and together (Table A16). Omitted one at a time, those three firms yield estimates of 1.69, 1.87, and 1.86. When all three are omitted together, the sample loses 41 filings (18 operating and 32 of the sample's 867 notified filings) and the slope falls to 1.42 [0.23, 2.61]. Together the three firms account for 0.57 of the 1.99 points. Without them, the remaining slope stays well above zero.

Only one breach episode reaches multiple sample firms under our shared-episode rule. We define an episode as shared when archived records link two or more sample firms to the same event and a notification sits wholly inside at least one window. Only the 2024 Cencora (Lash Group) episode qualifies. Its nine firms account for 90 filings (12 operating and 15 notified windows, representing 4 operating and 5 comparison positives among 117 distinct 2024–2025 operating-breach episodes). When those nine firms exit estimation, Table A16 shows that the slope stands at 1.90 [0.53, 3.26].

Sequential single-year exclusions appear in Table A17 and Figure A3. Seventeen of the eighteen intervals exclude zero, and sixteen of the point estimates sit between 1.77 and 2.44. The two years outside that narrow band hold the largest cohorts of operating filings. Without 2024 (652 operating filings), the estimate falls to 1.29 [−0.29, 2.88]. Without 2025 (899 operating filings), it rises to 2.66 [0.79, 4.52].

TABLE A16. Connected-Firm Clustering and Omission of Influential Firms or Shared Episodes

Sample or clustering modification	Filings / clusters	Operation (β)	Supply (θ)	Op. − Supply
None (baseline firm-clustered specification)	28,266 / 4,451	1.99***	−0.23	2.23*
		(0.70) [0.62, 3.37]	(0.60)	(1.15)
None, clustered by episode/group-connected firms	28,266 / 4,439	1.99***	−0.23	2.23*
		(0.70) [0.61, 3.37]	(0.60)	(1.15)
Omit nine firms linked to shared Cencora episode	28,176 / 4,442	1.90***	−0.18	2.08*
		(0.70) [0.53, 3.26]	(0.60)	(1.14)
Omit top-influence firm 1 (GVKEY 7647)	28,249 / 4,450	1.69***	−0.10	1.79*
		(0.63) [0.45, 2.93]	(0.59)	(1.07)
Omit top-influence firm 2 (GVKEY 10035)	28,259 / 4,450	1.87***	−0.14	2.01*
		(0.69) [0.51, 3.22]	(0.60)	(1.13)
Omit top-influence firm 3 (GVKEY 177376)	28,249 / 4,450	1.86***	−0.11	1.97*
		(0.69) [0.50, 3.22]	(0.60)	(1.13)
Omit all three most influential firms	28,225 / 4,448	1.42**	0.12	1.30
		(0.61) [0.23, 2.61]	(0.57)	(1.01)

Notes: Baseline specification (Table 4, column 3) refitted after the indicated clustering or sample change ($100 \times Y_{if}$, percentage points). Firm-clustered (CR1) standard errors appear in parentheses (with 95 per cent confidence intervals also shown in brackets under Operation). Row 2 clusters on 4,439 connected sets of firms linked through shared breach episodes or corporate groups. Rows 4–7 rank all 4,451 firms by absolute first-order influence on $\hat{\beta}$ and omit the top three individually and jointly. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Firms hit by a breach might adopt AI to defend themselves and report operation afterwards because of that earlier intrusion. We test this reverse-selection channel on two sets of operating filings whose passages link AI use to cybersecurity or fraud prevention. The broader census holds all 224 sample operating filings at 112 firms whose passages establish an explicit or contextually supported link to cyber defence (114 filings) or to fraud or crime detection (120 filings, with 10 in both). The narrower core subset holds 97 filings at 43 firms drawn from the 750 operating filings where all six coded application fields are unambiguously resolved. Within that subset, 49 filings link to cyber defence and 52 link to fraud detection, with 4 in both. Each set is either dropped from the estimation sample or recoded as comparison filings ($A_{if} = 0$).

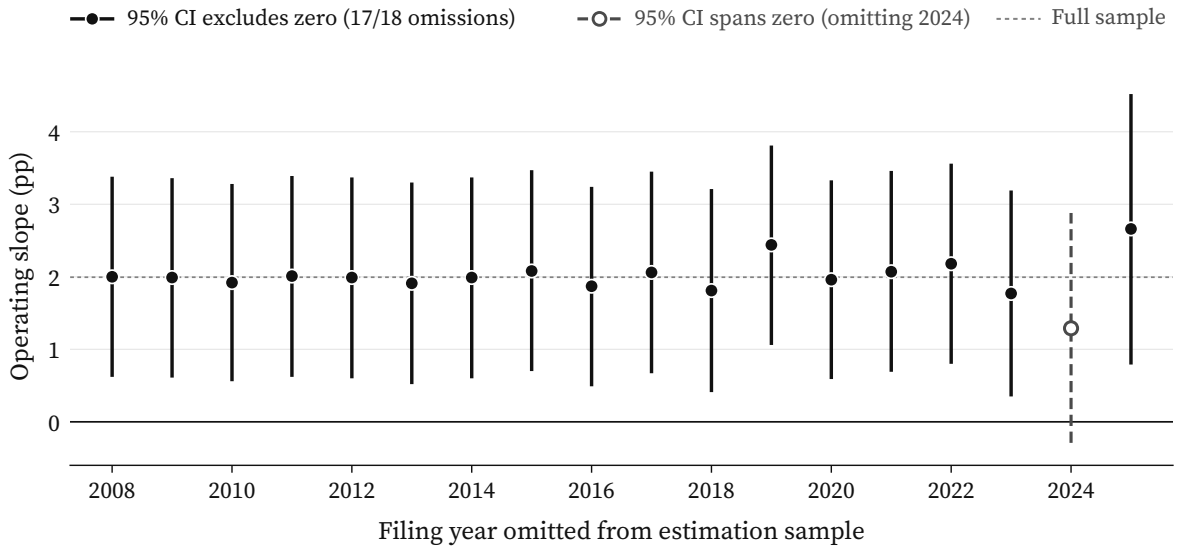
Across all eight defensive and fraud-AI checks in Table A18, the operating estimate stays between 1.65 and 1.94 percentage points with every interval above zero. When the broader census is recoded as comparison filings, the slope reaches its lowest point at 1.65 [0.25, 3.05]. The eight exclusions show that security or fraud-detection AI accounts for at most one-sixth of the primary estimate.

TABLE A17. Leave-One-Filing-Year-Out Stability of Role Estimates (2008–2025)

Year	Filings	Firms	Operation (β)	Op. 95% CI	Supply (θ)	Op. − Supply
2008	27,174	4,451	2.00*** (0.70)	[0.62, 3.38]	−0.23 (0.61)	2.23* (1.15)
2009	27,004	4,447	1.99*** (0.70)	[0.61, 3.36]	−0.24 (0.61)	2.23* (1.16)
2010	27,015	4,448	1.92*** (0.69)	[0.56, 3.28]	−0.20 (0.60)	2.12* (1.14)
2011	27,050	4,451	2.01*** (0.71)	[0.62, 3.39]	−0.23 (0.61)	2.23* (1.16)
2012	27,035	4,451	1.99*** (0.71)	[0.60, 3.37]	−0.20 (0.61)	2.19* (1.16)
2013	26,979	4,451	1.91*** (0.71)	[0.52, 3.30]	−0.23 (0.61)	2.14* (1.16)
2014	26,944	4,451	1.99*** (0.71)	[0.60, 3.37]	−0.24 (0.61)	2.22* (1.16)
2015	26,911	4,451	2.08*** (0.71)	[0.70, 3.47]	−0.28 (0.60)	2.36** (1.15)
2016	26,855	4,451	1.87*** (0.70)	[0.49, 3.24]	−0.14 (0.61)	2.01* (1.16)
2017	26,808	4,450	2.06*** (0.71)	[0.67, 3.45]	−0.27 (0.61)	2.33** (1.16)
2018	26,747	4,451	1.81** (0.71)	[0.41, 3.21]	−0.36 (0.61)	2.17* (1.18)
2019	26,733	4,451	2.44*** (0.70)	[1.06, 3.81]	−0.41 (0.62)	2.84** (1.16)
2020	26,669	4,451	1.96*** (0.70)	[0.59, 3.33]	−0.17 (0.62)	2.13* (1.16)
2021	26,588	4,451	2.07*** (0.71)	[0.69, 3.46]	−0.14 (0.63)	2.21* (1.18)
2022	26,580	4,451	2.18*** (0.70)	[0.80, 3.56]	−0.43 (0.60)	2.61** (1.14)
2023	26,405	4,451	1.77** (0.72)	[0.35, 3.19]	−0.12 (0.65)	1.89 (1.20)
2024	25,452	4,446	1.29 (0.81)	[−0.29, 2.88]	0.28 (0.63)	1.01 (1.26)
2025	25,573	4,440	2.66*** (0.95)	[0.79, 4.52]	−0.66 (0.74)	3.32** (1.53)

Notes: Baseline specification (Table 4, column 3) refitted omitting one filing year at a time ($100 \times Y_{if}$, percentage points). Three-digit SIC × filing-year fixed effects are re-absorbed on each subsample while covariate standardisation constants remain fixed. Firm-clustered (CR1) standard errors are in parentheses; 95 per cent confidence intervals for β are in brackets. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

FIGURE A3. Leave-one-filing-year-out stability of the operating coefficient



Notes: Operating coefficients and 95 per cent firm-clustered intervals (in percentage points) when the main model is refitted omitting one filing year at a time across 2008–2025 (Table A17). Filled black circles with solid stems denote subsamples whose 95 per cent intervals exclude zero (17 of 18 omissions); the open circle with a dashed stem denotes omitting 2024 (1.29 percentage points, $[-0.29, 2.88]$). The horizontal dashed line marks the full-sample baseline estimate of 1.99 percentage points.

TABLE A18. Operating Coefficient When Specific Filing Subsets Are Dropped or Recoded

Targeted subset of filings	Subset N / firms	Treatment	Model N	Operation (β)	Std. error	95% CI
None (baseline main model)	–	–	28,266	1.99***	(0.70)	[0.62, 3.37]
Security or fraud-detection AI, census	224 / 112	Dropped	28,042	1.81**	(0.72)	[0.41, 3.22]
		Recoded	28,266	1.65**	(0.71)	[0.25, 3.05]
Security or fraud-detection AI, core	97 / 43	Dropped	28,169	1.84***	(0.70)	[0.47, 3.21]
		Recoded	28,266	1.82***	(0.69)	[0.47, 3.17]
Cyber-defence AI, core	49 / 23	Dropped	28,217	1.92***	(0.71)	[0.53, 3.31]
		Recoded	28,266	1.86***	(0.71)	[0.46, 3.25]
Fraud-detection AI, core	52 / 21	Dropped	28,214	1.91***	(0.69)	[0.56, 3.26]
		Recoded	28,266	1.94***	(0.68)	[0.60, 3.27]
Risk-section-only operating filings	770 / 536	Dropped	27,496	1.06	(0.82)	[−0.54, 2.67]
		Recoded	28,266	0.23	(0.83)	[−1.40, 1.86]
Filings with AI passages in Item 1C only	114 / 106	Dropped	28,152	2.01***	(0.71)	[0.62, 3.39]

Notes: Operating coefficient (β, in percentage points) from the baseline specification (Table 4, column 3). "Dropped" removes the subset prior to absorbing three-digit SIC × year fixed effects; "Recoded" retains all 28,266 filings while setting $A_{if} = 0$ on the subset. The 224-filing census covers all operating filings linking AI to cybersecurity or fraud prevention; the 97-filing core restricts to filings with all six application fields unambiguously resolved (4 filings appear in both cyber defence and fraud detection). The final row drops the 114 sample filings with AI passages confined to Item 1C (not read by the second model). Firm-clustered (CR1) standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D.2 Dual-role overlap and disclosure location

If firms that build AI can also secure it, the gap should be smaller where the operating firm also develops or supplies AI. We add the product of operation and development or supply to the main model. Operation without development is then the operation coefficient against filings with neither role. Operation with development is operation plus the interaction term against filings that report development or supply only. Both estimates, their difference, and the raw notification rates across the four role cells appear in Table A19.

Pooled over all years, operation without development is 2.62 [0.59, 4.66] and operation with development is 1.07 [−0.48, 2.63], a difference of −1.55 [−4.08, 0.98] ($p = 0.23$). That negative pooled difference reflects 2008–2023, where it sits at −3.69 [−8.85, 1.47] across 197 operation-only filings. In 2024–2025, when 1,551 of the 2,441 operating filings appear, the two estimates equal 2.58 and 2.46 with a difference of −0.12 [−2.98, 2.73]. Under the third reading with unclear filings omitted, the pooled difference equals −1.26 [−4.08, 1.55]. Raw cell averages confirm that 7.38 per cent of operation-only filings are followed by a notification against 5.02 per cent for both roles, 2.47 per cent for development or supply only, and 2.82 per cent for neither.

Of the 2,441 operating filings, 770 at 536 firms have their only evidence of current operation in the risk sections. Among those 770 filings, 699 rest on Item 1A alone, 58 on Item 1C alone, and 13 on both, with 108 filed in 2008–2023 against 662 in 2024–2025. Section location is recorded less reliably in filings before 2024. Some of those 108 filings may therefore contain business-section evidence as well. When all 770 filings exit the sample, the estimate falls to 1.06 [−0.54, 2.67]. When

TABLE A19. Internal AI Operation With and Without Commercial AI Development or Supply

Role cell or interaction contrast	All years	2008–2023	2024–2025
Panel A: Primary classification ($N = 28{,}266$ filings; 4,451 firms)			
Operation without development or supply ($A_{if} = 1, S_{if} = 0$)	2.62**	4.10*	2.58**
	(1.04)	(2.49)	(1.05)
Operation with development or supply ($A_{if} = 1, S_{if} = 1$ vs. $S_{if} = 1$ only)	1.07	0.41	2.46**
	(0.79)	(1.16)	(0.97)
Interaction difference (η on $A_{if} \times S_{if}$)	−1.55	−3.69	−0.12
	(1.29)	(2.63)	(1.46)
Development or supply only ($A_{if} = 0, S_{if} = 1$)	0.40	0.82	−0.43
	(0.57)	(0.78)	(0.69)
Panel B: Third-model adjudication on contested filings, unclear omitted ($N = 27{,}978$ filings)			
Operation without development or supply	2.42**	3.07	2.86**
	(1.17)	(2.42)	(1.23)
Operation with development or supply	1.16	0.58	2.48**
	(0.86)	(1.20)	(1.06)
Interaction difference (η on $A_{if} \times S_{if}$)	−1.26	−2.50	−0.39
	(1.44)	(2.58)	(1.64)
Panel C: Raw 12-month breach notification rate (%) [filing count] by role cell			
Operation only ($A_{if} = 1, S_{if} = 0$)	7.38 [827]	10.15 [197]	6.51 [630]
Both operation and supply ($A_{if} = 1, S_{if} = 1$)	5.02 [1,614]	5.05 [693]	4.99 [921]
Development or supply only ($A_{if} = 0, S_{if} = 1$)	2.47 [770]	3.43 [496]	0.73 [274]
Neither role ($A_{if} = 0, S_{if} = 0$)	2.82 [25,055]	2.96 [21,373]	1.98 [3,682]

Notes: Panels A and B augment the baseline and period-interacted specifications with $A_{if} \times S_{if}$ ($100 \times Y_{if}$, percentage points). Firm-clustered (CR1) standard errors are in parentheses; the cross-era shift in η under the primary reading is 3.57 (2.88), $[-2.08, 9.22]$. Panel C reports unadjusted notification rates with cell filing counts in brackets. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

those 770 filings are recoded as comparison filings, it falls to 0.23 [−1.40, 1.86] (Table A18). The location model with two terms yields 0.97 [−0.67, 2.62] for business-section operating filings and 3.39 [1.42, 5.37] for risk-section-only operating filings, a difference of 2.42 [0.09, 4.74]. When the risk-section indicator splits between Item 1A and Item 1C, the 699 pure Item 1A filings carry 3.59 [1.49, 5.69] against an imprecise 1.49 [−3.41, 6.38] for the 71 filings that touch Item 1C. The 1,671 business-section operating filings stay at 1.00 [−0.65, 2.65] in that three-way specification.

The blinded passage audit of Appendix B.3 splits those 770 filings into 664 whose passage states operation and 106 whose passage mentions AI only or is unclear (Table A20). Risk-section passages that state operation carry 2.92 [0.81, 5.03] against 0.93 [−0.71, 2.58] for business-section operating filings. In 2024–2025, the difference between the two reaches 2.95 [0.26, 5.63]. The 106 mention-only or unclear filings carry a noisy point estimate of 6.20 [0.25, 12.16].

Firm size offers the first rival explanation for that split. Risk-section-only operating filings come from larger firms, with mean log total assets of 8.21 against 7.06 for business-section operating filings and 6.88 for comparison filings. Nearly a quarter (24.5 per cent) sit in the top size decile against 10.1 and 9.6 per cent. When nine asset-decile indicators join linear log total assets, the risk-section statement term remains at 2.63 [0.53, 4.72] pooled and 3.50 [1.19, 5.80] in 2024–2025.

Panel B of Table A12 shows that risk-section-only operating filings post the highest raw rate in four of the five size fifths. In the top fifth, 14.24 per cent are followed by a notification against 11.20 per cent of business-section operating filings and 8.51 per cent of comparison filings.

TABLE A20. Disclosure-Location Regressions by Blinded Passage Class, With and Without Asset Deciles

Specification and period	Business-section operating (Item 1)	Risk-section-only operating filings: States op. (S)	Risk-section-only operating filings: Mentions / unclear (M/U)
Panel A: Baseline controls (three-digit SIC × year FE)			
All years (2008–2025)	0.93	2.92***	6.20**
	(0.84)	(1.08)	(3.04)
2008–2023	1.47	1.92	13.13
	(1.38)	(2.37)	(9.37)
2024–2025	0.65	3.60***	5.42
	(0.88)	(1.18)	(3.30)
Panel B: Baseline controls plus nine asset-decile indicators			
All years (2008–2025)	0.92	2.63**	5.33*
	(0.79)	(1.07)	(3.02)
2008–2023	1.20	1.02	12.85
	(1.27)	(2.45)	(9.73)
2024–2025	0.87	3.50***	4.42
	(0.89)	(1.18)	(3.26)
Operating filings, all years	1,671	664	106
Operating filings, 2008–2023	782	92	16
Operating filings, 2024–2025	889	572	90
Unadjusted 12-month breach rate, all years (%)	4.73	7.53	12.26

Notes: Linear probability estimates ($N = 28{,}266$ filings; 4,451 firms; comparison breach rate 2.81%) partitioning A_{if} into three mutually exclusive indicators ($100 \times Y_{if}$, percentage points). Firm-clustered (CR1) standard errors are in parentheses. In Panel A, the contrast between Risk-section states operation (S) and Business-section operation equals $1.98\ (1.25)$, $[-0.47, 4.44]$ across all years and $2.95^{**}\ (1.37)$, $[0.26, 5.63]$ in 2024–2025. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D.3 Recording delay and unread no-cue filings

Breach outcomes are dated by the report date on which a notification enters the Privacy Rights Clearinghouse. When we rebuild every twelve-month window from report dates alone, the same 867 filings register a notification together with two filings that the outcome rule treats as unnotified. The delay from a breach to its report does not move a filing's window. Of the 1,785 episodes at sample firms with a report date, 1,152 also carry a breach start date. For the 537 breaches that began in 2015–2021, the median time to report is 78 days.

Archive ingestion delay can still hide recent reports. To measure it, we compared the Clearinghouse file of 13 March 2026 against the file of 5 August 2026. Of the August rows reported within two months before 13 March, 47.3 per cent were absent from the March file against 9.2 per cent for rows reported three or more years earlier. We applied the excess absence by months before the

August file to every week of each twelve-month window. Expected shares of in-window reports not yet recorded equal 1.7 per cent for filings of 2023, 6.7 per cent for 2024, and 9.2 per cent for 2025 (Table A21).

That shortfall is too small to explain the fall in the comparison rate from 3.95 per cent in 2023 to 2.27 in 2024 and 1.45 in 2025. Even if all 9.2 per cent of unrecorded reports were restored, the 2025 rate would rise from 1.45 only to about 1.60, still far below 3.95. When comparison filings are reweighted to the 2023 two-digit SIC mix, rates equal 2.39 and 1.51. Reweighted to the 2023 size fifths, they equal 2.48 and 1.58. As Table A21 confirms, neither archive ingestion delay nor the industry or size composition of comparison filings accounts for the fall. The operating rate moves less and changes from 6.64 per cent in 2023 to 6.75 in 2024 and 4.78 per cent in 2025.

TABLE A21. Archive Recording Delay and Reweighted Comparison Breach Rates (2023–2025)

Metric	2023	2024	2025
Comparison filings ($A_{if} = 0$)	1,620	2,162	1,794
Comparison 12-month notification rate (%)	3.95	2.27	1.45
Reweighted to 2023 two-digit SIC mix (%)	3.95	2.39	1.51
Reweighted to 2023 log-asset quintile mix (%)	3.95	2.48	1.58
Operating filings ($A_{if} = 1$)	241	652	899
Operating 12-month notification rate (%)	6.64	6.75	4.78
Expected share of in-window notices not yet recorded (%)[a]	1.7	6.7	9.2

Notes: Evaluates whether archive ingestion lag or compositional shifts among comparison filings explain the post-2023 decline in unadjusted comparison notification rates.
[a] Computed by applying the excess absence curve between the 13 March 2026 and 5 August 2026 Privacy Rights Clearinghouse snapshots, indexed by months elapsed prior to snapshot capture, across every calendar week of each filing's twelve-month post-filing window.

Across the raw archive, the 44,569 filings without a retrieved AI cue were never read. Text controls, cyber discussion, prior notification, and window exclusions were never constructed for them. Unretrieved filings without a cue cannot enter the main model as zeros. Our comparison group is therefore the set of read filings that do not establish internal operation. Appendix B.1 reports the negative second screen across initial no-cue filings.

E. Econometric diagnostics for clustering, rare events, and covariate balance

Table A22 collects four econometric diagnostics across alternative clustering assumptions, nonlinear rare-event specifications, proportional-selection bounds, and covariate balance via entropy weights. Each check tests a distinct limitation of our principal estimates without claiming causal identification. Agreement across models is informative only when changes in functional form, estimation support, and comparison populations remain explicit. Intervals are 95 per cent confidence intervals.

TABLE A22. Multi-Way Clustering, Rare-Event Nonlinear Models, Proportional-Selection Bounds, and Entropy Balancing

Panel A: Multi-way clustering and wild cluster bootstraps ($N = 28{,}266$; 4,451 firms; 252 SIC3; 18 years)

Focal indicator	Est. (pp)	Firm CR1 [p]	Firm × SIC3 [p]	Firm × yr [p]
Internal AI operation	1.99***	(0.702) [0.0045]	(0.731) [0.0140]	(0.982) [0.0850]
AI development or supply	−0.23	(0.604) [0.6981]	(0.609) [0.6687]	(0.648) [0.7501]
Item 1 operation	0.97	(0.841) [0.2469]	(0.765) [0.2009]	(1.145) [0.3698]
Item 1A-only operation	3.39***	(1.008) [0.0008]	(1.080) [0.0040]	(1.194) [0.1604]
Item 1A statement (S)	2.92***	(1.076) [0.0067]	(1.189) [0.0110]	(1.663) [0.1689]

Panel B: Concentrated SIC3 × year exponential QMLE ($N = 14{,}868$; 473 cells) and full-sample hazard links

Specification	Rel. risk [95% CI][a]	AME [95% CI]	Cell LPM (SE)	Model p
Internal AI operation	1.585*** [1.208, 2.079]	3.10*** [0.93, 5.27]	2.41** (0.94)	0.0009
AI development or supply	1.020 [0.760, 1.370]	0.12 [−1.57, 1.80]	−0.42 (0.89)	0.8925
Item 1 operation	1.372* [0.993, 1.895]	1.98* [−0.32, 4.27]	1.46 (1.15)	0.0549
Item 1A-only operation	1.847*** [1.326, 2.572]	4.50*** [1.38, 7.63]	3.61*** (1.30)	0.0003
Item 1A statement (S)	1.810*** [1.271, 2.576]	4.30*** [1.02, 7.59]	3.02** (1.43)	0.0010
Cloglog ($N = 28{,}536$)	$\hat{\beta} = 0.4555$	1.45*** [0.47, 2.42]	2.02*** (0.63)	0.0035
Logit ($N = 28{,}536$)	$\hat{\beta} = 0.5368$	1.63*** [0.55, 2.71]	2.02*** (0.63)	0.0031

Panel C: Proportional-selection bounds within SIC3 × year cells (Oster 2019; $R_{\max} = 1.3\tilde{R}^2$)

Focal indicator	Short $\dot{\beta}$	Controlled $\tilde{\beta}$	Bias-adj. β^*	δ_0 $(1.3/2.0\tilde{R}^2)$
Operation, all years	3.70	1.99	1.46	3.63 / 1.11
Operation, 2024–2025	4.30	2.55	1.91	3.48 / 1.15
Item 1A-only, all years	5.96	3.39	2.58	4.03 / 1.24
Statement (S), all years	5.50	2.92	2.11	3.53 / 1.08

Panel D: Entropy-balanced SIC3 × year WLS contrasts (Hainmueller 2012; all $p \leq 0.0001$)

Balanced contrast	Treated N_1 / ESS	Max |SMD|	Contrast (SE)	95% CI
Operation vs. comp., all yrs	2,441 / 588.1	2.27 → 0.00	2.64*** (0.66)	[1.34, 3.94]
Operation vs. comp., 2024–25	1,551 / 352.2	1.27 → 0.00	3.10*** (0.70)	[1.72, 4.47]
Item 1A-only vs. comp., all yrs	770 / 983.0	1.79 → 0.00	4.34*** (1.08)	[2.23, 6.46]
Item 1A-only vs. comp., 2024–25	662 / 857.7	0.78 → 0.00	4.72*** (1.18)	[2.40, 7.04]
Stmt. (S) vs. comp., all yrs	664 / 912.0	1.82 → 0.00	4.24*** (1.08)	[2.13, 6.35]
Stmt. (S) vs. comp., 2024–25	572 / 824.2	0.86 → 0.00	4.77*** (1.19)	[2.44, 7.10]

Notes: Standard errors in parentheses. Panel A brackets report p-values (CR1 t_{G-1} in col. 3; 2,000-draw Rademacher one-way wild cluster bootstrap p by 252 SIC3 industries in col. 4 and 18 years in col. 5 with Cameron, Gelbach, and Miller 2011 two-way SEs; Roodman et al. 2019). Panel B estimates concentrated exponential QMLE (Wooldridge 1999, Santos Silva and Tenreyro 2006) on 473 interior cells (5.65% breach rate) with firm-clustered delta-method AMEs (pp), plus full-sample cloglog and logit. Panel C reports Oster (2019) $\beta^*(\delta = 1)$ and breakdown δ_0 ($\tilde{R}^2 = 0.0735$ pooled). Panel D reweights comparison filings via entropy balancing (Hainmueller 2012) prior to SIC3 × year WLS ($p = 0.0001$ in rows 1, 3–6; $p < 0.0001$ in row 2). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

[a] Rows 6–7 report link-scale $\hat{\beta}$; Model p tests log RR in rows 1–5 and AME in rows 6–7.

E.1 Two-way clustering and wild cluster bootstrap

We calculate two-way cluster-adjusted standard errors (Cameron, Gelbach, and Miller 2011) for firm-by-industry and firm-by-year combinations together with one-way 2,000-draw restricted Rademacher wild cluster bootstrap p-values clustered by 252 three-digit SIC industries and by 18 filing years (Roodman et al. 2019, Table A22, panel A). Industry dependence leaves main inference intact. The pooled operation slope stays at 1.99 percentage points with a firm-by-SIC3 standard error of 0.731 and an industry-bootstrap p-value of 0.0140. For risk-section-only operation, the slope is 3.39, the standard error is 1.080, and the bootstrap p-value is 0.0040. For risk-section statements S, the estimates are 2.92, 1.189, and 0.0110.

Calendar clustering yields a more qualified inferential picture that reflects late-panel concentration. Firm-by-year standard errors rise to 0.982 for pooled operation and 1.194 for risk-section-only operation, with one-way year-bootstrap p-values at 0.0850 and 0.1604. Only 18 calendar clusters enter. Of the $2,441$ operating filings, $1,551$ arrive in 2024 and 2025. Those two years also hold 662 of the 770 risk-section-only filings. Calendar concentration limits nominal clusters. That pattern matches Appendix D.1, where a sample without 2024 lowers the operation slope to 1.29 percentage points $[-0.29, 2.88]$ and confirms that precision depends on which calendar years supply identifying comparisons.

E.2 Concentrated industry-by-year exponential models and hazard links

Unconditional three-digit industry-by-year logit has no finite maximum because $2,857$ all-zero cells ($13,371$ filings) and 27 single-filing all-one cells (27 filings) perfectly separate outcomes through fixed effects. Separation drives cell intercepts to infinity. Linear models retain all observations at the cost of negative fitted probabilities in low-risk cells. We therefore add a nonnegative check with explicit support.

We estimate concentrated fixed-effects exponential pseudo-maximum-likelihood models (Wooldridge 1999, Santos Silva and Tenreyro 2006) on the 473 interior cells that contain $14,868$ filings and 840 notifications, a 5.65 per cent breach rate (Table A22, panel B). Under exponential concentration, all-zero cells drop out because cell totals equal zero. Single-filing cells also contribute zero within-cell covariate variation to the concentrated log-likelihood regardless of outcome. Those 473 cells form the exact support.

Operation carries a relative risk of 1.585 $[1.208, 2.079]$ ($p = 0.0009$) and an average marginal difference of 3.10 percentage points $[0.93, 5.27]$ against 2.41 $[0.57, 4.25]$ from the linear model on the exact same $14,868$ filings. Commercial supply stays near unity at 1.020 $[0.760, 1.370]$ ($p = 0.8925$). Against the non-operating comparison baseline, business-section operation yields a relative risk of 1.372 $[0.993, 1.895]$ and an average marginal difference of 1.98 points $[-0.32, 4.27]$. Risk-section-only operation reaches 1.847 $[1.326, 2.572]$ and 4.50 points $[1.38, 7.63]$ against 1.810 $[1.271, 2.576]$ and 4.30 points $[1.02, 7.59]$ for active risk-section statements S. The location pattern survives here.

In panel B of Table A22, exponential concentration profiles out cell intercepts from slope optimisation and avoids the incidental-parameter problem under conditional-mean assumptions without guarding against misspecification. Positive means rule out negative predictions. Exponential means stay unconstrained above unity. That specification therefore acts as a multiplicative conditional-mean diagnostic rather than a bounded binary probability model.

On the full sample of $28,536$ filings, a complementary log-log model built on a 12-month proportional-hazards link yields an operation average marginal difference of 1.45 points $[0.47, 2.42]$ ($p = 0.0035$) against 1.63 points $[0.55, 2.71]$ ($p = 0.0031$) under broad logit. Jeffreys-prior penalised logit produces nearly identical estimates ($\hat{\beta} = 0.5356$, or 1.65 percentage points). These estimates complement Appendix C.2. Because their sample support and fixed effects are broader, they cannot isolate link-function curvature alone.

E.3 Proportional-selection bounds and entropy-balanced contrasts

We apply the proportional-selection framework of Oster (2019) within three-digit industry-by-year cells to relate slope movements to residualised within-cell explanatory power with fixed effects and sibling role indicators held constant (Table A22, panel C). For pooled operation, the seven controls lift residualised within-cell R^2 from $\dot{R}^2 = 0.0021$ to $\tilde{R}^2 = 0.0735$ and move the slope from 3.70 to 1.99 percentage points. Under $R_{\max} = 1.3\tilde{R}^2$, the bias-adjusted coefficient $\beta^*(\delta = 1)$ equals 1.46 points with a nullifying selection ratio $\delta_0 = 3.63$. Across the remaining rows, adjusted coefficients equal 1.91 for 2024–2025 operation ($\delta_0 = 3.48$), 2.58 for risk-section-only operation ($\delta_0 = 4.03$), and 2.11 for risk-section statements S ($\delta_0 = 3.53$).

Assumed R^2 ceilings alter those ratios. A higher ceiling of $2.0\tilde{R}^2$ lowers pooled δ_0 to 1.11 and shows that distance from zero depends strongly on stipulated omitted explanatory power. The bounds discipline confounders that covary like firm size, disclosure prose, and prior notifications under proportional selection without bounding unrestricted unobserved processes. Slow-moving personal-record holdings remain an active between-firm rival orthogonal to filing controls. The inconclusive within-firm estimate of 0.74 points $[-1.23, 2.71]$ in Section 5.3 therefore remains an essential qualification.

As in Hainmueller (2012), we combine entropy balance with industry-by-year weighted least squares to build fixed-weight contrasts that target the operating population (Table A22, panel D). Entropy weights lower maximum absolute standardised mean differences from 0.78–2.27 to 0.00 at reported precision on the positive-weight active-year support, where early years without treated filings receive exact zero weight. The operation contrast equals 2.64 points $[1.34, 3.94]$ across all years and 3.10 $[1.72, 4.47]$ in 2024–2025. Across risk-section disclosures, risk-section-only operation yields 4.34 points $[2.23, 6.46]$ across active years and 4.72 $[2.40, 7.04]$ in 2024–2025 against 4.24 $[2.13, 6.35]$ and 4.77 $[2.44, 7.10]$ for active statements S.

Mean balance cannot guarantee overlap. The pooled comparison group's effective sample size is 588.1. That concentrated effective size reveals sharp weight concentration after reweighting despite

the larger nominal comparison pool. Fixed-weight regressions still assume that the conditional mean is correctly specified given balanced covariates. Sample-mean balance removes neither omitted record stocks nor unmeasured disclosure selection.

F. Post-filing onset timing, reporting lags, and 2025 temporal prediction

Two tables collect our timing and prediction checks. Across both periods, Table A23 reports the six post-filing breach-onset and delayed-reporting specifications, raw onset-to-notification lag medians, event-study leads, and within-firm versus year-clustered inference diagnostics discussed in Sections 5.2 and 5.3. Table A24 presents the retrospective temporal-split 2025 prediction models and 2,000-resample paired bootstrap contrasts examined in Section 5.3.

G. Bank hosting check

The exposure account depends in part on how operational AI handles internal customer records and whether systems run in-house or on external vendor infrastructure. To evaluate whether public disclosures reveal that hosting boundary inside a homogeneous industry, we examined customer and account servicing across 24 listed US commercial banks (historical SIC codes 602 and 603). For each bank, our review checked the latest Form 10-K business section (Item 1), risk factors (Item 1A), and cybersecurity governance section (Item 1C, across 23 filings for fiscal 2024 and one for fiscal 2023). Where available, we also read one supplemental public bank document such as an annual shareholder report, investor presentation, or product release (12 supplemental documents across the 24 banks).

Classification required active AI use in customer or account servicing rather than credit underwriting, fraud scoring, marketing, or general staff productivity. External-hosting attribution required a documented link among the servicing AI application, nonpublic customer inputs, and an outside data processor. In-house execution required comparable inputs, bank-operated processing, and an explicit restriction on outside handling along the same data path. Across the 24 banks, four (Citizens, Bank of America, Cullen/Frost, and Princeton) establish that they operate customer-servicing AI in dated public disclosures. Two establish that workflows handle nonpublic customer inputs (including live customer-service conversations processed by an agent-assist tool). Zero disclose in-house versus external hosting.

Research assistants verified all 24 bank assessments against the full Form 10-K texts and supplemental bank documents with structured candidate packets. Across the five substantive fields coded per bank (120 fields in total), full-text verification matched 118 entries and updated two entries for Cullen/Frost. That check revised nonpublic customer-input handling from unclear to supported and outside-handling restriction to not established. Public disclosures therefore leave hosting

TABLE A23. Post-Filing Breach Onset Timing, Delayed Reporting Windows, and Panel Inference Diagnostics

	Pooled 2008–2025 ($N = 28{,}266$)			Recent 2024–2025 ($N_{\text{recent}} = 5{,}507$)		
Timing restriction	Windows	Operation (β)	Op. − Supply	Windows	Operation (β)	Op. − Supply
Panel A: Post-filing onset and delayed reporting-window regressions ($100 \times Y_{if}$, percentage points)						
1. Baseline $(0, 365]$-day window	867	1.99***	2.23*	162	2.55***	3.05**
	3.07%	(0.70)	(1.15)	2.94%	(0.80)	(1.47)
2. Drop episodes with onset $\leq d_f$	719	1.37**	1.62	118	1.85**	2.33*
	2.54%	(0.63)	(1.01)	2.14%	(0.73)	(1.32)
3. Drop window if any onset $\leq d_f$	659	1.06*	1.24	103	1.29**	1.62
	2.33%	(0.54)	(0.89)	1.87%	(0.65)	(1.16)
4. Strict confirmed onset $d_{\text{breach}} > d_f$	431	1.04**	1.12	92	1.26**	1.28
	1.52%	(0.51)	(0.81)	1.67%	(0.61)	(1.11)
5. Delayed $(90, 365]$-day window	685	1.71***	2.10**	131	2.31***	3.41**
	2.42%	(0.64)	(1.07)	2.38%	(0.75)	(1.39)
6. Delayed $(90, 365]$d, drop onset $\leq d_f$	616	1.39**	1.73*	107	1.94***	2.56*
	2.18%	(0.62)	(1.02)	1.94%	(0.72)	(1.32)
7. Delayed $(180, 365]$-day window	496	1.25**	1.48	89	1.94***	3.26**
	1.75%	(0.57)	(0.95)	1.62%	(0.70)	(1.29)

Panel B: Onset-to-report lag, event-study pre-trends, within-firm power, and resampling diagnostics	
Diagnostic test	Summary statistic and inference
Post-filing lag ($d_{\text{report}} - d_{\text{breach}}$)	Median 69 vs. 50 days pooled ($p = 0.013$; controlled $+19.5$ d, $p = 0.181$); 67 vs. 58 days in 2024–2025 ($p = 0.470$; controlled $+23.0$ d, $p = 0.195$)
Event-time pre-trends (817 switchers)	Joint Wald $F(2, 4450) = 0.69$ ($p = 0.501$, ref. $t = -1$); $t \leq -4$: -0.31 [−2.96, 2.34]; $t = -3$: -1.41 [−4.13, 1.32]; $t = -2$: 0.14 [−2.48, 2.76]; $t = 0$: 1.37 [−1.36, 4.11]; $t = +1$: -0.70 [−3.55, 2.16]
Within-firm fixed effects & power	Firm + year FE: 0.74 (1.00), [−1.23, 2.71]; firm + SIC3 × year FE: 1.43 (1.02), [−0.57, 3.44] ($\text{MDE}_{80\%} = 2.86$ pp; Mundlak $\hat{\beta}_B - \hat{\beta}_W = 1.55$ pp, $p = 0.295$; Hausman $p = 0.476$)
Year-cluster & permutation tests	Year clusters ($G = 18$, $G^* = 4.74$): six-point Webb WCB $p = 0.082$; within-cell 1,000-permutation firm-persistent RI $p = 0.008$ pooled (ordinary studentised $p = 0.009$) and $p \leq 0.001$ in 2024–2025

Notes: Panel A estimates equations (1) and (3) across 28,266 filings (4,451 firms; 3,357 three-digit SIC × year cells). Firm-clustered (CR1) standard errors are in parentheses; smaller figures under Windows report positive outcome rates (%). Standalone $N = 5{,}507$ subsample intervals are slightly tighter (e.g. $[1.01, 4.09]$ on row 1 operation and $[0.04, 5.09]$ on row 6 contrast). Commercial supply slopes (θ) stay small and statistically indistinguishable from zero across all seven rows (-0.39 to -0.08 pp pooled; -1.32 to -0.02 pp recent). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

TABLE A24. Retrospective Temporal-Split 2025 Prediction Diagnostic ($N_{\text{test}} = 2{,}693$; 69 Breaches)

Prediction model or paired contrast	Log-Loss ($\times 100$)	Brier ($\times 100$)	PR-AUC (%)	ROC-AUC	Top-10% capture (%)
Panel A: Fixed 1 January 2025 origin ($N_{train} = 22{,}759$ pre-2024 settled filings; ridge logit L2)					
M0: Scale, text lengths, prior breach & SIC3	10.18	2.24	22.75	0.742	44.93
M1: M0 + cybersecurity & filing text[a]	10.42	2.31	21.79	0.745	43.48
M2: M1 + any AI role (A_{if} or S_{if})	10.55	2.35	22.62	0.749	42.03
M3: M1 + separate operation & supply	10.51	2.33	23.08	0.750	43.48
M4: M1 + section-split op. & supply	10.49	2.33	22.81	0.749	44.93
Panel B: Monthly expanding origin ($N_{train} = 22{,}759$ to $25{,}539$ settled filings; ridge logit L2)					
M0: Scale, text lengths, prior breach & SIC3	10.13	2.24	22.81	0.750	44.93
M1: M0 + cybersecurity & filing text[a]	10.34	2.31	21.11	0.753	43.48
M2: M1 + any AI role (A_{if} or S_{if})	10.44	2.33	22.83	0.756	43.48
M3: M1 + separate operation & supply	10.48	2.33	23.00	0.751	44.93
M4: M1 + section-split op. & supply	10.50	2.33	22.73	0.750	44.93
Panel C: Paired firm-clustered bootstrap differences (Δ and 95% CI across 2,000 firm resamples)					
Fixed MLE: M3 − M1 (vs. cyber text)	$+0.086$ [-0.028, 0.197]	$+0.030^{*}$ [-0.001, 0.062]	$+1.17$ [-0.24, 3.21]	$+0.006$ [-0.003, 0.015]	$+1.45$ [-3.51, 4.55]
Fixed L2: M3 − M1 (vs. cyber text)	$+0.084$ [-0.025, 0.190]	$+0.024$ [-0.005, 0.054]	$+1.29^{*}$ [-0.05, 3.15]	$+0.005$ [-0.003, 0.014]	0.00 [-4.29, 4.48]
Monthly L2: M3 − M1 (vs. cyber text)	$+0.135^{**}$ [0.020, 0.254]	$+0.026$ [-0.005, 0.057]	$+1.89$ [-0.20, 4.25]	-0.002 [-0.013, 0.009]	$+1.45$ [-3.90, 4.55]
Monthly LPM: M3 − M1 (vs. cyber text)	-0.006 [-0.240, 0.197]	$+0.013^{**}$ [0.002, 0.023]	$+0.58^{*}$ [-0.07, 1.47]	$+0.008$ [-0.005, 0.022]	0.00 [-4.23, 4.48]
Fixed MLE: M3 − M2 (role separation)	-0.052^{**} [-0.099, -0.009]	-0.017^{***} [-0.032, -0.004]	$+0.33$ [-0.47, 1.28]	$+0.001$ [-0.003, 0.005]	$+2.90$ [0.00, 5.56]
Fixed L2: M3 − M2 (role separation)	-0.042^{**} [-0.085, -0.001]	-0.014^{**} [-0.027, -0.001]	$+0.47$ [-0.34, 1.54]	$+0.001$ [-0.003, 0.005]	$+1.45$ [0.00, 5.41]
Monthly MLE: M3 − M2 (role separation)	-0.034^{*} [-0.074, 0.003]	-0.011^{*} [-0.024, 0.002]	$+0.19$ [-0.50, 0.94]	$+0.001$ [-0.002, 0.004]	$+1.45$ [0.00, 4.69]

Notes: Out-of-sample evaluation across all 2,693 Form 10-K filings in 2025 (69 notified windows; 2.56% base rate). At every forecast origin t_0, only earlier filings whose 365-day breach window closed on or before t_0 enter estimation; continuous predictors and empirical-Bayes three-digit SIC historical breach rates are refit strictly inside each training sample. Lower values indicate better probability accuracy for Log-Loss $\times 100$ and Brier $\times 100$; higher values indicate stronger risk ranking for PR-AUC, ROC-AUC, and Top-10% breach capture. Under fixed-origin MLE, M1–M4 attain PR-AUC of 21.86, 22.71, 23.04, and 23.22% and Log-Loss $\times 100$ of 10.41, 10.55, 10.49, and 10.53. Predictions use archival Form 10-K filings and breach records observed as of the 5 August 2026 Clearinghouse snapshot rather than live ingestion feeds. Paired two-sided bootstrap significance: $^{*}p_{\text{boot}} < 0.10$, $^{**}p_{\text{boot}} < 0.05$, $^{***}p_{\text{boot}} < 0.01$.

[a] M1 adds standardised cybersecurity discussion, digital-technology wording, and customer-data wording motivated by Wang, Kannan, and Ulmer (2013).

arrangements unobservable across all 24 banks and prevent direct estimation of hosting-specific breach comparisons.

H. Data and replication

From the root of the replication package, `python3 run_all.py` executes six core modular scripts (`code/01_sample_and_descriptives.py` through `code/06_figures.py`) that reproduce the foundational exhibits and write CSV and JSON outputs to `output/tables/` and `output/figures/`. Companion validation scripts in the replication archive reproduce Tables 5, 6, A23,

and A24 from the primary-source breach audit, human-validation, and main estimation datasets. Table A25 maps each exhibit to its dataset and script.

Primary public sources comprise SEC EDGAR Form 10-K filings (1994–2026 archive) and the 5 August 2026 Privacy Rights Clearinghouse Data Breach Chronology snapshot covering notices through 31 July 2026, with a 6 July 2026 window cutoff so that 2025 filings complete full twelve-month windows. Historical firm total assets and dated three-digit SIC codes come from S&P Compustat via Wharton Research Data Services (WRDS). Researchers with WRDS access can rebuild raw upstream Compustat extracts from the variable definitions and dating rules in Appendix A.1. Derived covariates sit directly inside `data/main_sample.csv` and `data/full_sample.csv`.

TABLE A25. Replication Datasets and Scripts by Exhibit

Exhibits	Analysis dataset(s)	Replication script
Tables 2 (Panel B), 3, A3	`main_sample.csv`, `risk_passage_audit.csv`	`01_sample_and_descriptives.py`
Tables 4, A13, A14, A15; Section C.2 logit check	`main_sample.csv`, `full_sample.csv`	`02_main_regressions.py`
Tables A7, A8, A9, A10, A11	`main_sample.csv`, `full_sample.csv`, `archive_agreement.csv`, `third_reading_audit.csv`, `masked_rereading_40.csv`, `risk_passage_audit.csv`	`03_reader_reliability.py`
Tables A12, A16, A17, A18, A19, A20, A21; Appendix G	`main_sample.csv`, `recording_delay_lag_bands.csv`, `recording_delay_expected_missing.csv`, `bank_hosting_check.csv`	`04_sensitivity_checks.py`
Table A22 (Panels A–D)	`main_sample.csv`, `full_sample.csv`	`05_econometric_diagnostics.py`
Figures 1, 2, A1, A2, A3	`main_sample.csv`, `full_sample.csv`	`06_figures.py`
Tables 5, 6, A23, A24	Human-validation dataset, 2024–2025 primary-source breach audit, `main_sample.csv`	Companion validation and audit scripts (see `README.md`)

Notes: Paths in rows 1–6 are relative to the root of the core replication package (`run_all.py`, `DATA_DICTIONARY.md`, `README.md`); shared fixed-effects demeaning, cluster-robust covariance, Cohen's κ, and Holm step-down routines reside in `code/_estimators.py`. Tables 1, A1, A2, A4, A5, and A6 summarise theoretical predictions, variable definitions, sample flow counts, and illustrative excerpts documented in Appendices A and B.

References for the Online Appendix

Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2011. "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics* 29 (2): 238–249.

Chen, Xiaohong, Han Hong, and Elie Tamer. 2005. "Measurement Error Models with Auxiliary Data." *Review of Economic Studies* 72 (2): 343–366.

Egami, Naoki, Musashi Hinck, Brandon Stewart, and Hanying Wei. 2023. "Using Imperfect Surrogates for Downstream Inference: Design-Based Supervised Learning for Social Science Applications of Large Language Models." In *Advances in Neural Information Processing Systems*, vol. 36: 68589–68601.

Florackis, C., C. Louca, R. Michaely, and M. Weber. 2023. “Cybersecurity Risk.” *Review of Financial Studies* 36 (1): 351–407.

Hainmueller, J. 2012. “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies.” *Political Analysis* 20 (1): 25–46.

Holm, Sture. 1979. “A Simple Sequentially Rejective Multiple Test Procedure.” *Scandinavian Journal of Statistics* 6 (2): 65–70.

Oster, E. 2019. “Unobservable Selection and Coefficient Stability: Theory and Evidence.” *Journal of Business & Economic Statistics* 37 (2): 187–204.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. “Estimation of Regression Coefficients When Some Regressors Are Not Always Observed.” *Journal of the American Statistical Association* 89 (427): 846–866.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb. 2019. “Fast and Wild: Bootstrap Inference in Stata Using boottest.” *Stata Journal* 19 (1): 4–60.

Santos Silva, J. M. C., and S. Tenreyro. 2006. “The Log of Gravity.” *Review of Economics and Statistics* 88 (4): 641–658.

Securities and Exchange Commission. 2023. “SEC Adopts Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies.” Press release 2023-139, 26 July. https://www.sec.gov/newsroom/press-releases/2023-139.

Wang, T., K. N. Kannan, and J. R. Ulmer. 2013. “The Association Between the Disclosure and the Realization of Information Security Risk Factors.” *Information Systems Research* 24 (2): 201–218.

Wooldridge, J. M. 1999. “Distribution-Free Estimation of Some Nonlinear Panel Data Models.” *Journal of Econometrics* 90 (1): 77–97.