Measuring Data Breach Severity with Large Language Models: Evidence from Market Reactions

Fearghal Kearney[a], Tren Ma[b], Hadi Movaghari[c], Georgios Sermpinis[d,*]

[a]*Queen's University Belfast, Belfast, UK*
[b]*University of Nottingham, Nottingham, UK*
[c]*University of Leicester, Leicester, UK*
[d]*University of Glasgow, Glasgow, UK*

Abstract

Data breaches create material but difficult-to-measure risks for listed firms, yet previous studies often find only modest stock-market reactions to breach announcements. We use ChatGPT to transform narrative breach descriptions into standardised impact scores and test whether these scores explain short-window abnormal returns around breach announcements. Higher ChatGPT-assessed impact is associated with more negative returns, robust to alternative event-study specifications and repeated scoring across random seeds. The scores capture information beyond breach size, sentiment and disclosure length, and the association is strongest among financial services firms. Weak average market reactions can therefore conceal meaningful differences across breaches.

Keywords: Large Language Models, Cyber Risk, Data Breaches, ChatGPT

JEL Codes: G14, G32, C45, M15

*Corresponding author

Email addresses: f.kearney@qub.ac.uk (Fearghal Kearney), tren.ma@nottingham.ac.uk (Tren Ma), hm479@leicester.ac.uk (Hadi Movaghari), georgios.sermpinis@glasgow.ac.uk (Georgios Sermpinis)

1. Introduction

Data breaches are a specific subset of cyber risk that occur when there is unauthorized access or exposure of sensitive information.[1] Despite growing institutional and regulatory emphasis on cyber resilience and investment planning (Dacorogna et al., 2023), there is a notable paradox in the reaction to these events. On the one hand, academic studies suggest that market responses to data breaches are often negligible or minor (see, for example, Campbell et al., 2003; Kannan et al., 2007; Goel and Shawky, 2009; Foerderer and Schuetz, 2022).[2] On the other hand, policymakers have introduced increasingly stringent regulations related to cybersecurity, privacy and data breaches (see, for example, the EU's GDPR or the SEC's cybersecurity disclosure rules), suggesting that policymakers view such events as materially significant. Consistent with this, the passage of mandatory breach disclosure laws in the United States is associated with higher corporate cash holdings, while firms that experience data breaches subsequently increase their cash holdings and reduce external financing and investment (Boasiako and Keefe, 2021). Mandatory disclosure laws are also associated with distortions in firms' real activities (Liu and Ni, 2024). Breaches therefore carry material consequences for firms even where the average announcement return might appear small. This paradox likely stems from the difficulty of quantifying the full scope of data breaches, which are intangible, infrequent, context-dependent, and subject to heavy-tailed risk dynamics (Cavusoglu et al., 2004; Kamiya et al., 2021). We seek to address this challenge by leveraging advances in Generative AI to assess the severity of data breaches and offer a new tool for financial risk managers and practitioners dealing with this new type of risk.

Recently, Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLM) such as ChatGPT, has seen applications across a wide range of domains, including being proposed as tools for evaluating market sentiment using news articles (Kirtac and Germano, 2024; Lopez-Lira and

[1] Numerous major cybersecurity incidents have resulted in large-scale data breaches with significant consequences. Notable examples include the Yahoo breaches in 2016, which affected more than one billion user accounts, and the JP Morgan Chase breach, in which hackers gained access to the personal information of millions of customers.

[2] Foerderer and Schuetz (2022) suggest that the market's muted reaction to data breaches may be due to firms strategically timing their announcements to coincide with busy news periods. Closely related, Kannan et al. (2007) find that the market's reaction to security breaches depends on the presence of larger concurrent events, such as the September 11 terrorist attacks, that may distract investors. Some studies link the mitigation of negative market reactions following incident disclosures to breached firms' efforts to communicate their remedial actions and reassure investors about the restoration of normal operations (Lin and She, 2025).

Tang, 2025).[3] We study the extent to which ChatGPT can extract economically meaningful signals from data breach disclosures, and we examine the conditions under which such assessments are more or less informative. Using textual disclosures of data breaches, we generate ChatGPT-based severity assessments (expressed as scores) and examine their association with short-term stock market reactions. Rather than interpreting alignment with market reactions as direct evidence of model accuracy, we view it as indicating the extent to which an LLM can extract economically relevant information from unstructured disclosure texts. We further investigate how this relationship varies with disclosure characteristics, including textual length, numeric content, and technical sophistication, as well as with firm attributes and breach types. By analysing how ChatGPT's assessments vary across random seeds, disclosure characteristics, and contexts, this study also sheds light on the practical limitations of deploying LLMs in operational risk management.

Concerning our findings, we show that ChatGPT's assessment of a data breach is associated with firms' short-term market reactions: a one-standard-deviation increase in the score is associated with a decrease of approximately 0.38 percentage points in CAR$(-1,+1)$. The result is robust to alternative sample constructions and abnormal-return models, and to estimating a single conventional regression on the score averaged across runs. The score is not associated with pre-announcement returns, and post-announcement returns show, if anything, a partial reversal rather than continued drift. It also captures information beyond simple proxies such as breach size, negative sentiment, and disclosure length, and it is distinct from existing measures of firms' ex ante cyber-risk exposure. The association is stronger for more technically sophisticated disclosures, as measured by the Flesch–Kincaid grade level, and for firms in financial services and healthcare, driven mainly by financial firms. Differences by disclosure length, interpretability, breach type, and firm size are in the predicted direction but not statistically significant in formal tests. Overall, our study introduces a new tool in operational risk management that is an easily accessible, inexpensive way of assessing a risk that can have a significant impact on publicly traded

[3]However, despite their promise, researchers must exercise caution when applying LLMs, due to limitations highlighted by de Kok (2025). We explicitly evaluate these limitations by examining when ChatGPT-generated severity assessments align with short-term market reactions and when they fail to do so.

firms.

This study contributes to the finance literature in three ways. First, it adds to research on the textual analysis of corporate disclosures. Prior work links textual attributes of disclosures to firm outcomes (Kearney and Liu, 2014; Loughran and McDonald, 2016; Caglio et al., 2020), and a newer stream applies LLMs to extract price-relevant information from financial text (Lopez-Lira and Tang, 2025; Kirtac and Germano, 2024; de Kok, 2025; Siano, 2025). We extend this line of inquiry by constructing an LLM-based measure of data breach severity from narrative disclosures and showing that it is associated with short-window abnormal returns around breach announcements. Unlike existing text-based measures of firms' ex ante cyber-risk exposure (Florackis et al., 2023; Jamilov et al., 2026), our measure captures the severity of realised breaches. We also show how to conduct inference with a stochastic LLM-generated measure, by scoring each event repeatedly and confirming the results with a conventional regression on the averaged score. Hoang and Wiegratz (2023) produce a banking and finance survey that finds that outside asset pricing, the payoff from machine learning is in constructing novel measures that are otherwise unavailable. Second, it adds to the event-study literature on the market valuation of data breaches. Existing evidence on average price reactions is weak or mixed (Campbell et al., 2003; Cavusoglu et al., 2004; Kannan et al., 2007; Goel and Shawky, 2009; Foerderer and Schuetz, 2022), while severe incidents can generate substantial losses (Kamiya et al., 2021). Further, combined with breach disclosure, breach incidents alter firms' liquidity, financing, investment, and operating decisions (Boasiako and Keefe, 2021; Liu and Ni, 2024). A text-based severity measure explains cross-sectional variation in announcement returns, which indicates that weak average reactions mask systematic differences across events. Third, we show that the informativeness of the severity measure is conditional. The association with returns is stronger when disclosures are more technical, and for firms in financial services and healthcare. The informativeness of LLM-based severity assessments therefore depends on the quality of narrative reporting and on context, not on model capability alone.

Overall, the paper positions generative AI not as a replacement for traditional risk assessment, but as a complementary informational mechanism that transforms unstructured disclosures into structured sig-

nals under uncertainty. This framing highlights both the potential and the contextual constraints associated with deploying LLM-based analytics in information-rich, yet imperfect, information environments, where decision makers must rely on incomplete textual information to form timely judgments.

Our paper is structured as follows. Section 2 develops our conceptual framework and hypotheses. Section 3 outlines our data, while Section 4 reports the empirical findings and discusses their implications, with Section 5 concluding.

2. Conceptual Framework and Hypotheses

In this section, we develop a conceptual framework in which ChatGPT acts as an informational agent that assesses the severity of a data breach event using contemporaneously available event-related information supplied to the model, including the breach description, the explanation of breach type, and, where available, information on media coverage. The framework builds on the efficient market hypothesis (Fama, 1970), according to which publicly available information is incorporated into stock prices as investors revise their expectations of firms' future cash flows and risks. Accordingly, information about data breaches should influence market valuations to the extent that it reveals economically relevant consequences for affected firms (Cavusoglu et al., 2004; Kannan et al., 2007; Simon and Omar, 2020). However, these consequences are often difficult to quantify, and their severity may not be fully captured by conventional breach classifications or observable characteristics.

The true economic severity of a data breach is unobserved and can be viewed as a latent construct. Let S_i^* denote the true severity of breach event i for the affected firm. We observe two imperfect manifestations of this latent severity. First, we observe a ChatGPT-assessed impact score, denoted by S_i, which is intended to measure S_i^*. Second, we observe a market reaction measure, MR_i, which captures investors' responses to the breach disclosure, such as the cumulative abnormal return (CAR) over a pre-specified event window.

We posit the following data-generating processes:

$$S_i = S_i^* + u_i, \tag{1}$$

$$MR_i = \alpha + \beta S_i^* + \varepsilon_i, \tag{2}$$

where u_i captures measurement noise in the ChatGPT-generated score arising from model stochasticity and information extraction error, and ε_i captures idiosyncratic components of market reactions unrelated to breach severity, such as concurrent news, firm-specific frictions, liquidity effects, and event-study estimation error. The parameter β governs the sensitivity of investor responses to breach severity.

Since S_i^* in Equation (2) is unobserved, we replace it with S_i in practice. By additionally considering control variables, we estimate the following regression specification:

$$MR_i = \alpha + \beta\, Impact_Score_i + \gamma' Controls_i + \varepsilon_i, \tag{3}$$

where MR_i is proxied by the Cumulative Abnormal Return (CAR), $Impact_Score_i$ represents the realisation of S_i, and $Controls_i$ denotes a vector of control variables with associated coefficient vector γ. Under the proposed framework, we expect β to be negative, reflecting larger stock price declines following more severe breaches.

If ChatGPT extracts economically meaningful information from breach disclosures, its severity assessments should exhibit a systematic negative association with announcement returns. Importantly, such an association does not imply that ChatGPT generates information unavailable to investors or accurately measures the full economic consequences of a breach. Rather, it indicates that the model can extract information from public disclosures that is relevant to the cross-sectional variation in market reactions. This constitutes our first hypothesis:

H1a: ChatGPT-based breach severity scores are negatively associated with short-window abnormal stock returns around data breach events.

The informativeness of ChatGPT-based assessments may nevertheless depend on the characteristics of the information supplied to the model. Data breach disclosures vary considerably in length, numerical detail, and technical complexity. Previous research suggests that additional contextual information

can improve the performance of LLMs in financial applications (Maarouf et al., 2025; Wu et al., 2025). More informative disclosures may enable ChatGPT to identify the nature and potential consequences of a breach more precisely, reducing measurement error in Equation (1). Although richer disclosures may also facilitate investors' interpretation of breach information, these effects are not necessarily contradictory. If both ChatGPT and investors extract more precise signals about the same underlying severity, the association between the score and market reactions should become stronger. However, richer disclosures may also characterize inherently more severe incidents. We therefore examine whether the relationship between *Impact_Score* and abnormal returns varies with disclosure length, numerical content, technical sophistication, and interpretability. This leads to our second hypothesis:

H1b: The negative association between ChatGPT-based severity scores and announcement-period abnormal returns is stronger for richer and more interpretable disclosures.

Beyond disclosure characteristics, the relationship between ChatGPT-assessed severity and market reactions may vary with the nature of the breach and the characteristics of the affected firm. Differences in measurement precision and the economic consequences of breaches may affect the estimated association between severity scores and abnormal returns.

First, breach type may influence the precision with which severity can be assessed from textual descriptions. Internal incidents, such as employee errors or accidental disclosures, may have more readily identifiable consequences, whereas external attacks can involve greater uncertainty about the extent of compromised information and the vulnerabilities exposed. The severity of internal breaches may be easier to assess from the available information, potentially reducing measurement error in the ChatGPT-generated score. This does not imply that internal breaches generate larger market reactions. External attacks may reveal more substantial vulnerabilities and trigger larger valuation losses (Kamiya et al., 2021; Akey et al., 2026). Rather, we expect the association between assessed severity and market reactions to be stronger when breach consequences can be inferred more precisely from the disclosure. Thus, we hypothesize that:

H2a: The association between the ChatGPT-based severity score and market reactions is stronger

for internal breaches (non-hacking) than for external breaches (hacking).

Second, firm size may affect the economic consequences of data breaches. Previous studies provide mixed evidence, with some documenting larger stock price declines for larger firms and others suggesting that smaller firms are more vulnerable to breach-related losses (Cavusoglu et al., 2004; Gatzlaff and McCullough, 2010; Akyildirim et al., 2024). Larger firms may experience greater absolute losses and attract more public attention. However, announcement returns measure changes relative to firm value, and smaller firms may have fewer financial resources and reputational buffers to absorb breach costs. In other words, a breach of comparable economic severity may have a greater proportional impact on smaller firms. We hypothesize that:

H2b: The association between the ChatGPT-based severity score and market reactions is stronger for smaller firms than for larger firms.

Finally, the economic consequences of data breaches may differ across industries. Firms operating in sectors that handle sensitive customer information or depend heavily on data security face potentially substantial regulatory, operational, and reputational costs following security incidents (Das et al., 2012; Bamiatzi et al., 2023). In a dedicated study, Shevchenko et al. (2023) show that monetary losses and the number of records affected vary across business sectors. These considerations are particularly relevant for financial services and healthcare firms, which hold the most sensitive customer data and face breach-specific regulation, such as the Gramm–Leach–Bliley Act for financial institutions and HIPAA and the HITECH Act's breach notification rule for healthcare providers. A breach of given severity may have a greater effect on firm value in these sectors. This leads to our final hypothesis:

H2c: The association between the ChatGPT-based severity score and market reactions is stronger for firms in financial services and healthcare than for firms in other sectors.

We evaluate our hypotheses using interaction specifications that formally test whether the relationship between *Impact_Score* and abnormal returns differs across breach types, firm sizes, and industries. These tests distinguish differences in estimated associations from differences in statistical significance across subsamples. Moreover, the underlying mechanisms are not mutually exclusive: heterogeneity

may arise from differences in the precision of ChatGPT-based severity assessments, differences in the economic consequences of breaches, or both.

Specifically, to test whether the association differs between two groups, we estimate the fully interacted regression

$$\begin{aligned} CAR_i = \alpha + \beta_1 Impact_Score_i + \beta_2 D_i + \beta_3 (Impact_Score_i \times D_i) \\ + \gamma' Controls_i + \delta' (Controls_i \times D_i) + \varepsilon_i, \end{aligned} \tag{4}$$

where D_i is a group indicator, so that β_3 measures the difference in the association between the two groups. For continuous characteristics Z_i, we estimate

$$CAR_i = \alpha + \beta_1 Impact_Score_i + \beta_2 Z_i + \beta_3 (Impact_Score_i \times Z_i) + \gamma' Controls_i + \varepsilon_i, \tag{5}$$

so that the marginal association between *Impact_Score* and CAR is $\beta_1 + \beta_3 Z_i$. When D_i or Z_i is based on firm size, *Size* is omitted from the controls.

Event studies often use CAR as a dependent variable to examine the effect of an incident on stock markets (Brown and Warner, 1985; Kamiya et al., 2021). Following Kamiya et al. (2021), we use the three-day cumulative abnormal return around the breach date as our dependent variable:

$$CAR(-1, +1) = \sum_{t=-1}^{1} AR_{it} \tag{6}$$

where AR_{it} is the abnormal return for firm i on day t, which is defined as the difference between the actual and expected return. More specifically, we utilise the market model to estimate the expected return.[4] We calculate the value for the data breach event using the U.S. Daily Event Study portal of Wharton Research Data Services (WRDS). To calculate the expected return, we set the length of our estimation window and the gap between the end of the estimation window and the beginning of the event window to 100 and 50 trading days, respectively. These are the default values in WRDS' event study portal.

Because the ChatGPT-generated score is stochastic, as described in Equation (1), we generate *Impact_Score* for all events 100 times using different random seeds. Accordingly, we estimate Equation (3) 100 times, each corresponding to one of the random seeds. This procedure allows us to characterise the

[4]We later conduct robustness checks using an alternative specification based on the Fama–French three-factor model.

distribution of coefficient estimates and assess the robustness of the inferred relationship. All p-values are based on heteroskedasticity-robust (HC1) standard errors.

3. Data

3.1. Data Sample

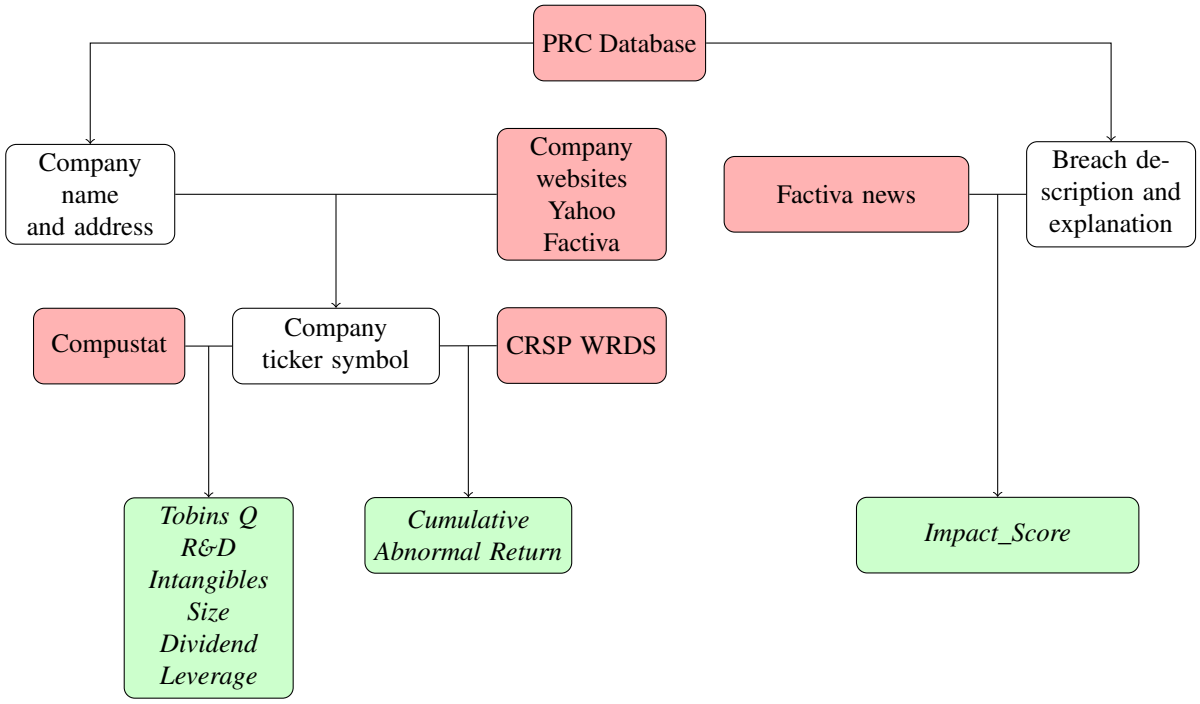
The data for this study come from multiple sources. We collect data breach events from Privacy Rights Clearinghouse (PRC) spanning 2005 to 2021. PRC is a non-profit corporation headquartered in California, U.S.A., with the mission of improving consumer awareness of privacy protection (Rosati et al., 2017; Eling and Wirfs, 2019; Bamiatzi et al., 2023).[5] PRC provides a comprehensive list of data for data breach incidents including Entity Name, Organisation Type, Breach Type and Data Breach Descriptions and Explanations.[6] We focus on the PRC data breach descriptions and the explanation for the type of breach to construct our ChatGPT scores. PRC states that their breach descriptions are collected from publicly available sources, including governmental sources such as the "US Department of Health and Human Services and various state Attorneys General who publish data breach notices they receive under their states' data breach notification laws". We manually search a sample of the PRC data breach descriptions and find that the same wording often also appears in news outlets, including The New York Times, CNN, Los Angeles Times, and on state and government department websites.

We identify public firms and retrieve their tickers by searching for the entity name on Factiva, Yahoo Finance, and company websites. To construct control variables, we acquire the firm's associated financial data for the fiscal year ending before the breach announcement from Compustat. A visual representation of our data set is presented in Figure 1. As a proxy for media attention, we search for news articles associated with each data breach. Using data from Factiva, we modify our prompt to provide the LLM with the names of major news outlets that have covered each breach. We ask the LLM, using the set of information including the PRC's description, explanation of breach type, and the list of outlets where the breach appeared, to assess the severity score. We label the obtained score as *Impact_Score* in

[5]For more details see https://privacyrights.org.
[6]Examples of data breach descriptions are given in Table A1 of the Appendix.

our output tables. We also follow an approach outlined by Kamiya et al. (2021) and use Factiva to identify data breach incidents that could be contaminated by other confounding events. We consider major corporate events, such as earnings announcements, mergers and acquisitions, investments or disinvestments, and restructuring (i.e., senior executive layoffs, global branch closures, etc.) as confounding events. We use Factiva's search builder to query a set of keywords associated with these events for the trading day before and the trading day after each data breach. This procedure led to the identification and exclusion of 47 confounding events, leaving a final sample of 238 events (before accounting for missing scores, as discussed in footnotes 8 and 9 below).

Figure 1: This figure summarizes our variable generation process. It begins with our main database (PRC Database) and ends with three green nodes representing the main variables used in our empirical analysis. The pink nodes represent databases or software, while the white nodes represent variables extracted from the databases or created during the process.



3.2. ChatGPT Impact Score

Using the OpenAI API, we input PRC data breach disclosure texts into ChatGPT to construct the *Impact_Score*.[7] Specifically, we ask ChatGPT to assess the severity of each breach by considering seven dimensions: (i) the type of data compromised (e.g., personal information, financial data, medical records,

[7]ChatGPT's outputs are generated by a pre-trained LLM trained on a wide range of publicly available sources. We do not perform any additional training or fine-tuning prior to deploying our prompts. In this study, we employ the `GPT-5-nano` model.

or intellectual property); (ii) the number of individuals affected; (iii) the sensitivity of the exposed data; (iv) the visibility of the breach; (v) the extent of media coverage; (vi) public reactions to the incident; and (vii) potential damage to the firm's reputation.

The *Impact_Score* is generated using the breach description, the explanation of the breach type, and, when available, the list of news outlets reporting the incident. To mitigate concerns related to look-ahead bias, we explicitly instruct the model to rely only on information that would have been available prior to the breach date.[8] Details of the prompt used to generate the score are provided in Section A of the Appendix. The initial score is reported on a 1–100 scale; because our analysis focuses on the relationship between the score and market reactions, we standardise the score prior to the empirical analysis.

3.3. Control Variables

We consider a set of six variables as controls whose effects on firm performance have been documented in prior cyber risk literature (e.g., Kamiya et al., 2021; Bamiatzi et al., 2023). These variables are *Tobin's Q* (market value of assets divided by total assets), *Size* (natural logarithm of total assets), *Dividend* (dummy variable for dividend paying firms), *R&D Intensity* (ratio of R&D expenditure to total assets), *Intangibles* (30% of selling, general and administrative expenses plus R&D expenses divided by total assets), and *Leverage* (total debt to total assets). Panels A and B of Table 1 provide descriptions and summary statistics for these variables, respectively.

4. Empirical Results

In this Section, we empirically evaluate whether ChatGPT-generated severity assessments provide economically meaningful signals about the latent severity of data breach events, as formalised in Section 2. We examine whether the *Impact_Score* is negatively associated with short-term stock market reactions, as predicted by H1a.

[8]This restriction is imposed consistently in both the system-level instructions and the user-level prompt. We also allow ChatGPT to return no score for events whose severity it cannot assess; see the detailed prompts in Appendix A.

Table 1: Variable Definitions and Descriptive Statistics. Panel A reports the definitions and data sources of the main dependent and control variables used in the analysis. Panel B presents descriptive statistics for these variables, including the minimum, first quartile (Q1), median, third quartile (Q3), and maximum values of each variable, as well as the standard deviation (Std) and the number of observations (N).

Panel A: Variable explanations

Variables	Definitions	Source
$CAR(-1,+1)$	Cumulative abnormal return within three-day window around the data breach date, using a market adjusted model	CRSP/WRDS
Tobin's Q	(Total assets + Market value – Total common equity)/Total assets	Compustat
R&D Intensity	R&D expenditure/Total assets	Compustat
Intangibles	(30% of selling, general and administrative expenses + R&D expenses)/Total assets	Compustat
Size	Natural logarithm of total assets	Compustat
Dividend	Dummy variable (1 if a firm pays cash dividends on common equity, 0 otherwise)	Compustat
Leverage	(Long-term debt + Debt in current liabilities)/Total assets	Compustat

Panel B: Descriptive Statistics

Variable	Min	Q1	Median	Q3	Max	Std	N
$CAR(-1,+1)$	-11.84	-1.37	-0.08	1.32	10.37	2.64	238
Tobin's Q	0.63	1.13	1.55	2.06	9.93	1.42	238
Dividend	0.00	0.00	1.00	1.00	1.00	0.49	238
R&D Intensity	0.00	0.00	0.00	0.01	0.63	0.07	238
Intangibles	0.00	0.01	0.04	0.10	0.87	0.10	238
Firm Size	3.59	7.65	9.08	10.83	14.71	2.22	238
Leverage	0.00	0.10	0.24	0.40	3.02	0.31	238

We generate severity scores using the OpenAI API, employing the `GPT-5-nano` model. To account for the stochastic nature of large language models and to characterise the measurement noise described in Equation (1), we score each breach event 100 times using different random seeds. For each seed, we obtain realisations of the *Impact_Score* for all events and estimate regression (3). As such, we conduct 100 regressions.[9] As discussed below, we evaluate the significance of variables based on the distributions of their estimated coefficients and corresponding p-values across the 100 regressions rather than relying on a single realisation. Using alternative sets of random seeds yields very similar distributions and does not affect our conclusions.

[9]Specifically, we use random seeds ranging from 0 to 99. Because ChatGPT is allowed to return no score when it cannot assess the severity of an event, the number of observations with a valid *Impact_Score* may vary across random seeds. In the baseline sample of 238 breach events, the effective sample size therefore ranges from 193 to 209 across runs (median 201).

4.1. ChatGPT Severity Score and Market Reaction

We first focus on the variable of interest, the ChatGPT impact score (*Impact_Score*). We present the histograms of its coefficient and corresponding p-value in Panel A of Figure 2. Our first observation is that all regressions yield a negative coefficient, ranging from -0.77 to -0.21, suggesting a negative association between the impact score and the abnormal return. The p-value histogram confirms this, with a concentration near zero and a majority less than 0.1. We note that if the scores are random, one should expect a uniform distribution of the p-value, i.e., a relatively flat histogram spread from 0 to 1.[10] To make this clearer, we conduct a simulation in which we replace ChatGPT's generated impact score with scores generated randomly (with replacement) between 1 and 100. We standardise the random score, conduct regressions, and depict the distributions of the coefficient of the random score and its p-value in Figure 3.

The simulation indicates that if the impact score is random and does not reflect the economic impact of the breach event on the firm, the estimated coefficient on the score would take both positive and negative values and be centred around zero, with corresponding p-values uniformly distributed over the interval $[0, 1]$. This stands in contrast to the distributions shown in Figure 2, where the ChatGPT-generated impact score yields: i) a consistently negative set of coefficient estimates; and, ii) a distribution of p-values that is highly concentrated at low values. These indicate an association between ChatGPT's generated impact score and the market reaction. Our findings are broadly consistent with the literature suggesting that data breaches have a negative impact on firms by reducing their market value (Bose and Leung, 2014).

Because *Impact_Score* is a stochastic, text-derived proxy for latent breach severity (and/or investor perceived severity), each run produces a slightly different realisation of the regressor. Under the null hy-

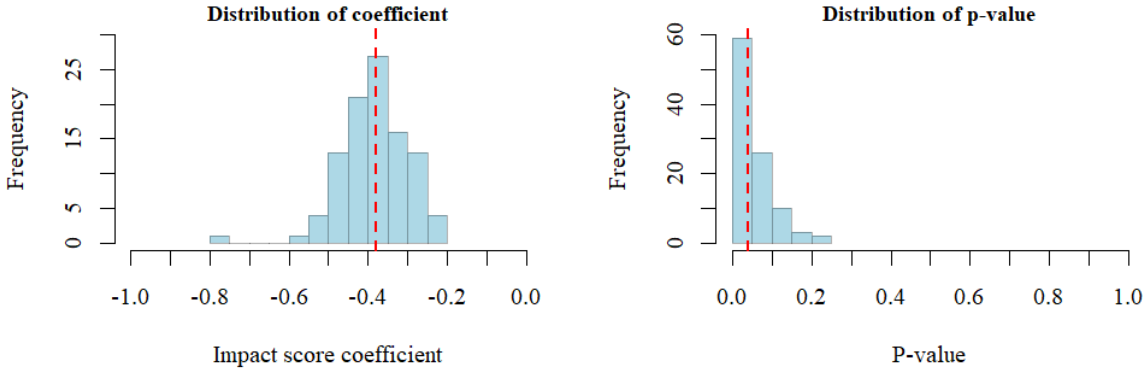
[10] Formally, the p-value of a test is a random variable defined as $P = 1 - F(|t|)$, where $F(|t|) = \mathbb{P}(|\mathcal{T}| < |t| \mid \beta = 0)$ and $\mathcal{T}$ denotes the Student's t-distribution, and t is the random variable representing the t-statistic of testing the null $\beta = 0$. Under the null,

$$\text{Prob}(P < p) = \text{Prob}(1 - F(|\mathcal{T}|) < p) = \text{Prob}(|\mathcal{T}| > F^{-1}(1-p)) = 1 - F(F^{-1}(1-p)) = p,$$

hence, the p-value is uniformly distributed on $[0, 1]$. If ChatGPT completely fails to assess the breach severity, i.e., if the scores are random, then the true $\beta = 0$ and the tests (each corresponding to a regression) are independent. In this case, the uniform distribution of realised p-values under the true null hypothesis implies a flat histogram over the interval $[0, 1]$.

Figure 2: Baseline Sample Results. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated p-values across 100 stochastic regressions. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for each coefficient from regression (3) across the 100 runs. The corresponding summary statistics for the p-values testing each coefficient against zero are displayed in italics. The final column reports the p-value from the Fisher–Brown test of the null hypothesis that the p-values follow a uniform distribution.

Panel A: Estimated *Impact_Score* Coefficient and p-value distributions

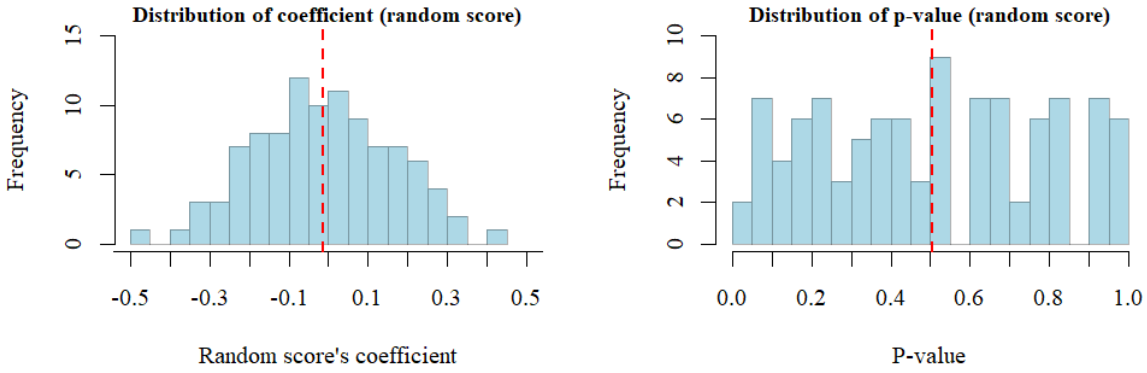


Panel B: Summary Statistics and Fisher–Brown Test Results

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact_Score	-0.77	-0.43	-0.38	-0.38	-0.33	-0.21	
	0	*0.02*	*0.04*	*0.06*	*0.08*	*0.25*	*0.000*
Tobin's Q	-0.15	-0.11	-0.08	-0.08	-0.05	0.07	
	0.36	*0.51*	*0.65*	*0.66*	*0.76*	*1*	*0.891*
Dividend	-0.39	-0.11	0	0	0.09	0.43	
	0.34	*0.69*	*0.81*	*0.79*	*0.92*	*1*	*0.980*
R&D Intensity	-1.48	0.76	1.48	1.35	2.1	4.19	
	0.57	*0.78*	*0.85*	*0.83*	*0.91*	*1*	*0.993*
Intangibles	1.68	3.26	4.04	4.1	4.79	6.85	
	0.12	*0.26*	*0.35*	*0.36*	*0.45*	*0.69*	*0.369*
Firm Size	0.15	0.21	0.23	0.23	0.26	0.29	
	0	*0.01*	*0.02*	*0.02*	*0.03*	*0.12*	*0.000*
Leverage	-0.93	-0.72	-0.64	-0.64	-0.54	-0.4	
	0.17	*0.29*	*0.34*	*0.35*	*0.41*	*0.54*	*0.369*
Constant	-3.12	-2.48	-2.24	-2.21	-1.9	-1.24	
	0.01	*0.04*	*0.06*	*0.08*	*0.11*	*0.31*	*0.003*
Observations	193	200	201	201	203	209	
R^2	0.059	0.071	0.08	0.081	0.09	0.153	

pothesis of no association between the score and market reactions ($\beta = 0$), the p-values from tests of the null $\beta = 0$ are uniformly distributed on $[0, 1]$. To aggregate evidence across stochastic realisations while accounting for cross-run dependence, we apply a dependence-adjusted Fisher–Brown combination test (Brown, 1975), hereafter referred to as the Fisher–Brown test. The resulting test rejects the uniformity implied by the null $\beta = 0$, providing aggregate evidence that the negative association between *Impact_Score* and CAR$(-1, +1)$ is not an artefact of LLM randomness. Economically, the median estimate

Figure 3: Simulation: Random Score Scenario. The figure presents the distribution of the estimated coefficient on the score across 100 regressions and the associated p-values when the score is randomly generated.



across runs implies that a one-standard-deviation increase in *Impact_Score* is associated with a decrease of approximately 0.38 percentage points in CAR$(-1,+1)$.

As a conventional alternative to combining evidence across runs, we standardise *Impact_Score* within each run, average it across the 100 runs for each breach event, and estimate equation (3) once using the averaged score for the 234 events with at least one valid score. The coefficient is -0.352 and statistically significant at the 5% level, and it remains significant at that level with heteroskedasticity-robust, firm-clustered, and firm-level cluster-bootstrap standard errors (Internet Appendix Table XVI). Because the averaged score is measured in units of within-run standard deviations, this implies a decrease of approximately 0.35 percentage points in CAR$(-1,+1)$ per standard deviation, close to the median estimate across runs. The two approaches are complementary: the averaged-score regression provides a conventional estimate of the economic magnitude and its sampling uncertainty, while the repeated-run analysis shows that the relationship is stable across stochastic realisations of the score.

4.2. Robustness Checks

Now we assess whether our main conclusion in Section 4.1, that the ChatGPT-based *Impact_Score* is negatively associated with short-window market reactions, is sensitive to (i) sample construction choices and (ii) the model used to compute abnormal returns. For each robustness exercise, we replicate the same procedure as in the baseline analysis and summarise the resulting distributions of coefficient estimates and p-values. We then examine the timing of the association and longer-horizon outcomes.

First, firms may experience multiple data breach events over time, and investor reactions may differ for initial versus subsequent breaches due to learning effects on a firm's cybersecurity posture. To account for this possibility, we restrict the sample to each firm's first observed breach and repeat the analysis. The results, as presented in Figure 4, are qualitatively and quantitatively similar to those obtained for the full sample in Figure 2. Relative to the baseline, the distribution of p-values shifts upward, consistent with reduced statistical power in the smaller first-breach sub-sample; nevertheless, the aggregate evidence continues to reject the null implied by no association. Overall, the first-breach results indicate that the negative relationship between *Impact_Score* and CAR$(-1,+1)$ is not driven by repeated-breach dynamics.

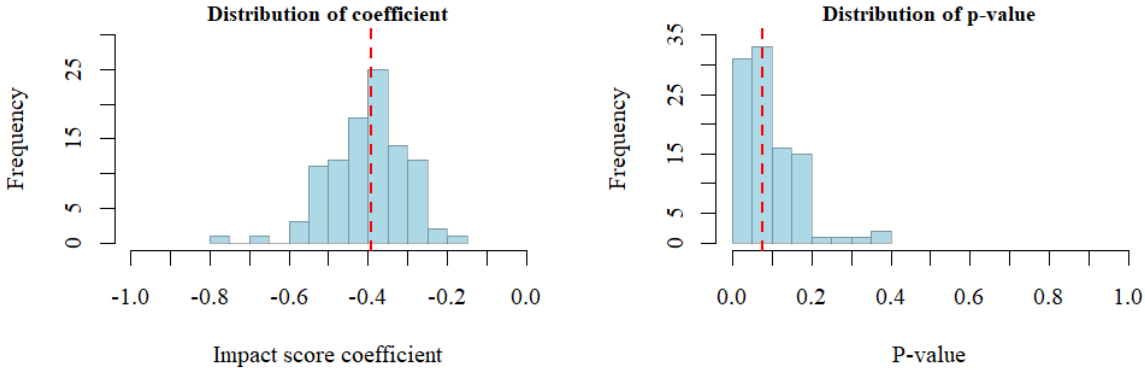
Second, our baseline sample excludes breach observations that coincide with potentially confounding firm news in the event window, in order to better isolate the market response to breach disclosures. As a robustness check, we relax this restriction and re-estimate (3) on the expanded sample that includes these potentially contaminated events. Figure 5 shows that the *Impact_Score* coefficient remains negative across runs, although the evidence is attenuated relative to the baseline, as expected when additional noise from overlapping announcements is introduced. This pattern suggests that any remaining confounds, if present, tend to bias the estimated association toward zero rather than generate a spurious negative relationship.

Last, we verify that our findings are not an artefact of the abnormal-return model used to compute CAR$(-1,+1)$. In the baseline analysis, expected returns are computed using the market model in the WRDS event study framework. We re-compute abnormal returns using the Fama–French three-factor model and repeat the stochastic regression procedure. Figure 6 confirms that the *Impact_Score* coefficient remains negative and the aggregate evidence continues to reject the null of no association, indicating that our conclusions are robust to standard alternative expected-return specifications.

We next examine whether the association reflects price movements that begin before the announcement or persist after it. Using pre-announcement abnormal returns, CAR(-10,-2), the *Impact_Score* coefficient is close to zero (median -0.055; Fisher–Brown p-value 0.942; Internet Appendix Table I),

Figure 4: First-Breach sub-sample. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions for the sub-sample of first-time breach events. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for each coefficient from regression (3) across the 100 runs. The corresponding summary statistics for the *p*-values testing each coefficient against zero are displayed in italics. The final column reports the *p*-value from the Fisher–Brown test of the null hypothesis that the *p*-values follow a uniform distribution.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions



Panel B: Summary Statistics and Fisher–Brown Test Results

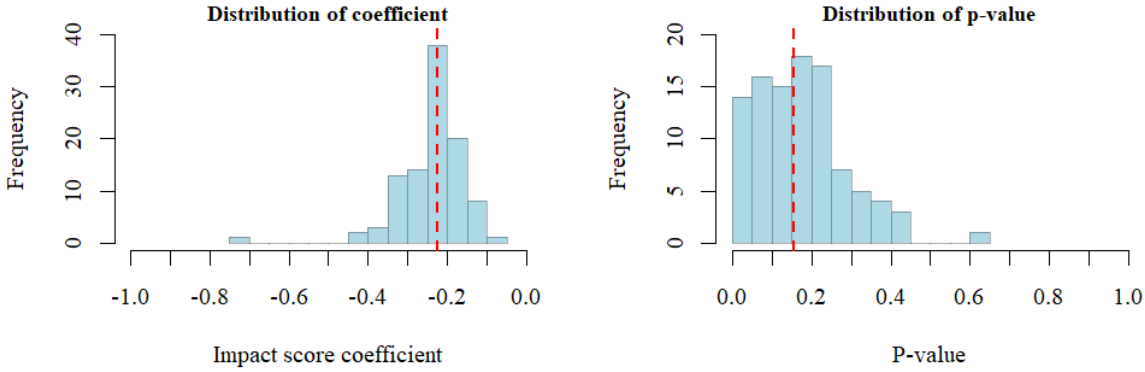
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.76	-0.45	-0.39	-0.4	-0.34	-0.18	
	0	*0.04*	*0.08*	*0.09*	*0.12*	*0.37*	*0.001*
Tobin's Q	-0.17	-0.11	-0.08	-0.07	-0.04	0.05	
	0.33	*0.55*	*0.66*	*0.68*	*0.81*	*0.99*	*0.913*
Dividend	-0.49	-0.2	-0.06	-0.07	0.05	0.36	
	0.36	*0.66*	*0.78*	*0.77*	*0.92*	*0.99*	*0.975*
R&D Intensity	-2.36	0	0.82	0.68	1.54	3.72	
	0.63	*0.82*	*0.89*	*0.87*	*0.93*	*1*	*0.998*
Intangibles	2.07	3.8	4.72	4.69	5.44	7.66	
	0.11	*0.24*	*0.32*	*0.34*	*0.42*	*0.65*	*0.312*
Firm Size	0.15	0.22	0.24	0.25	0.27	0.32	
	0.01	*0.02*	*0.04*	*0.05*	*0.06*	*0.2*	*0.001*
Leverage	-0.94	-0.75	-0.66	-0.66	-0.56	-0.44	
	0.18	*0.28*	*0.33*	*0.34*	*0.41*	*0.51*	*0.353*
Constant	-3.33	-2.6	-2.38	-2.34	-2.06	-1.17	
	0.01	*0.05*	*0.07*	*0.1*	*0.13*	*0.4*	*0.008*
Observations	165	171	172	173	174	178	
R^2	0.05	0.07	0.07	0.08	0.09	0.15	

so the announcement-window association is not a continuation of earlier price movements. Using post-announcement abnormal returns, CAR(2,20), the coefficient is positive (median 1.240; Fisher–Brown *p*-value 0.034; Internet Appendix Table XXIX), consistent with a partial reversal of the initial reaction for breaches the model assesses as more severe.

We also examine longer-horizon firm outcomes. More severe breaches are followed by a weak de-

Figure 5: Sample including Contaminations. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions for the sample that includes potentially confounding events. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for each coefficient from regression (3) across the 100 runs. The corresponding summary statistics for the *p*-values testing each coefficient against zero are displayed in italics. The final column reports the *p*-value from the Fisher–Brown test of the null hypothesis that the *p*-values follow a uniform distribution.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions



Panel B: Summary Statistics and Fisher–Brown Test Results

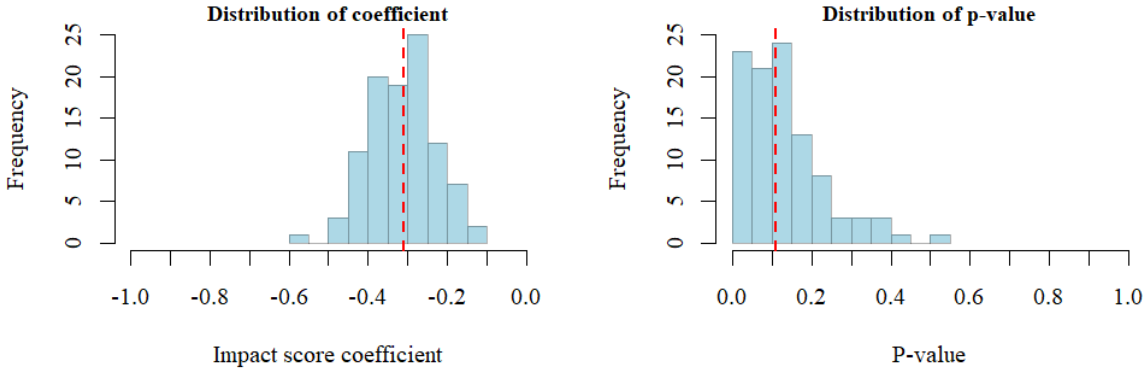
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.73	-0.28	-0.23	-0.24	-0.2	-0.08	
	0	*0.08*	*0.16*	*0.17*	*0.24*	*0.6*	*0.007*
Tobin's Q	-0.19	-0.15	-0.12	-0.12	-0.1	-0.01	
	0.21	*0.32*	*0.42*	*0.44*	*0.51*	*0.97*	*0.535*
Dividend	-0.46	-0.21	-0.13	-0.12	-0.05	0.2	
	0.27	*0.61*	*0.74*	*0.72*	*0.84*	*1*	*0.944*
R&D Intensity	-0.87	1.17	1.97	1.87	2.52	4.36	
	0.55	*0.73*	*0.79*	*0.8*	*0.87*	*1*	*0.985*
Intangibles	1.2	2.78	3.37	3.4	3.96	5.77	
	0.15	*0.31*	*0.4*	*0.4*	*0.49*	*0.76*	*0.46*
Firm Size	0.17	0.22	0.24	0.24	0.26	0.29	
	0	*0*	*0.01*	*0.01*	*0.01*	*0.06*	*0.000*
Leverage	-0.78	-0.61	-0.52	-0.53	-0.46	-0.32	
	0.23	*0.34*	*0.4*	*0.41*	*0.46*	*0.6*	*0.494*
Constant	-2.99	-2.38	-2.18	-2.16	-1.9	-1.34	
	0.01	*0.02*	*0.04*	*0.05*	*0.08*	*0.22*	*0.001*
Observations	234	241	242	243	245	252	
R^2	0.05	0.06	0.07	0.07	0.08	0.14	

cline in ROE in the year after the breach (Fisher–Brown *p*-value 0.077 with firm-level controls; Internet Appendix Tables II–VII), but we find no significant association with subsequent changes in ROA or sales growth (Internet Appendix Tables II–VII), or with the predicted securities litigation risk measure of Kim and Skinner (2012) (Internet Appendix Tables VIII and IX).

Our robustness tests confirm that the negative association between the ChatGPT-derived severity

Figure 6: Alternative Market Reaction Measure: Fama-French Three-Factor CAR. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, where cumulative abnormal returns are computed using the Fama-French three-factor model. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for each coefficient from regression (3) across the 100 runs. The corresponding summary statistics for the *p*-values testing each coefficient against zero are displayed in italics. The final column reports the *p*-value from the Fisher–Brown test of the null hypothesis that the *p*-values follow a uniform distribution.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions



Panel B: Summary Statistics and Fisher–Brown Test Results

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.59	-0.37	-0.31	-0.31	-0.26	-0.15	
	0	*0.06*	*0.11*	*0.13*	*0.18*	*0.5*	*0.004*
Tobin's Q	-0.24	-0.21	-0.18	-0.17	-0.14	-0.07	
	0.12	*0.2*	*0.25*	*0.29*	*0.38*	*0.67*	*0.227*
Dividend	-0.71	-0.48	-0.36	-0.36	-0.26	0.01	
	0.11	*0.28*	*0.42*	*0.43*	*0.55*	*0.99*	*0.485*
R&D Intensity	-4.54	-2.39	-1.49	-1.61	-0.66	1.14	
	0.63	*0.8*	*0.87*	*0.86*	*0.94*	*1*	*0.996*
Intangibles	4.35	5.89	6.74	6.72	7.52	9.49	
	0.04	*0.1*	*0.15*	*0.16*	*0.21*	*0.36*	*0.051*
Firm Size	0.08	0.15	0.17	0.18	0.2	0.24	
	0.02	*0.04*	*0.08*	*0.09*	*0.12*	*0.44*	*0.007*
Leverage	-0.54	-0.22	-0.15	-0.15	-0.07	0.16	
	0.36	*0.67*	*0.76*	*0.76*	*0.88*	*1*	*0.971*
Constant	-2.33	-1.85	-1.59	-1.57	-1.3	-0.36	
	0.05	*0.12*	*0.17*	*0.2*	*0.27*	*0.76*	*0.077*
Observations	193	200	201	201	203	209	
R^2	0.049	0.06	0.066	0.068	0.074	0.11	

proxy and short-window market reactions is not specific to repeated-breach firms, to the exclusion of potentially confounding events,or to a particular abnormal-return model, and it does not reflect pre-announcement price movements. Before examining when this signal strengthens or weakens, we next assess what information the score captures.

4.3. Validating the Impact Score

Because the *Impact_Score* is a new measure, we examine what information it captures beyond simpler proxies and how it relates to existing measures of cybersecurity risk.

We first compare the score with three transparent proxies for breach severity: the number of records affected (Rosati et al., 2017), negative sentiment measured with the Loughran and McDonald (2011) dictionary, and disclosure length. The score is significantly related to disclosure length and, in the joint specification, to negative sentiment, but is unrelated to the number of records affected, and together the three proxies explain only a small share of its variation (median R^2 of 0.161; Internet Appendix Table XXI). When the proxies are added to the baseline return regression, individually or jointly, the *Impact_Score* coefficient remains negative and significant (Fisher–Brown p-value 0.011 in the joint specification), whereas none of the proxies is significant at the 5% level (Internet Appendix Tables XVII–XX). The score therefore captures return-relevant information beyond breach size, sentiment, and length.

Of the 234 events with a valid score, only 22 have recorded newspaper coverage. Controlling for a news-coverage indicator leaves the *Impact_Score* coefficient negative and significant (Fisher–Brown p-value below 0.001), while the indicator itself is insignificant (Internet Appendix Tables XXII and XXIII). Given the small number of covered events, we do not interpret this as evidence that media coverage is irrelevant.

We next compare the score with firm-level textual measures of cybersecurity risk (Florackis et al., 2023; Jamilov et al., 2026). These measures capture a firm's ex ante exposure to cyber threats, whereas the *Impact_Score* assesses the severity of a breach that has already occurred, so a strong positive correlation is not expected. In overlapping observations, the score is negatively correlated with the Florackis et al. (2023) measure (median Pearson and Spearman correlations of −0.21 and −0.24) and only weakly correlated with the Jamilov et al. (2026) measures (median correlations between −0.10 and −0.04, with only one of ten tests significant at the 5% level; Internet Appendix Tables XXIV and XXV). Part of the negative correlation may reflect the decline in breach severity over our sample period, documented below, set against rising disclosure-based exposure. The *Impact_Score* therefore complements, rather

than reproduces, existing exposure-based measures.

Following the validation tests of Florackis et al. (2023), we also examine firm vulnerability characteristics and industry and time patterns. Breach severity is positively associated with firm size, intangible assets, and leverage and negatively associated with R&D intensity, although these characteristics explain little of its variation (median R^2 of 0.061; Internet Appendix Table XXVI). Average scores do not differ significantly across sectors (Internet Appendix Tables XXVII and XII), consistent with the score capturing event-specific severity rather than broad industry differences, and they decline over the sample period (Internet Appendix Table XXVIII), which may reflect improved cybersecurity practices or changes in the composition of reported breaches. Finally, the score does not predict whether the same firm experiences another breach (Internet Appendix Tables X and XI), consistent with it measuring the severity of a realised breach rather than a firm's exposure to future attacks.

4.4. Textual Disclosure Characteristics and Information Content

Having established that the ChatGPT-based *Impact_Score* covaries with short-window market reactions, we next examine when this signal is most informative. From an information processing perspective, breach announcements differ in the degree of uncertainty and equivocality they impose on the interpreter. Disclosures that are more detailed and better organised make the information less ambiguous and help the model identify useful signals more reliably, while sparse or internally inconsistent narratives increase the measurement noise component in the extracted score (the u_i term in Equation (1)). Accordingly, we study four disclosure attributes that plausibly affect signal extraction: (i) disclosure interpretability, (ii) disclosure length, (iii) quantitative (numeric) cue density, and (iv) technical sophistication of language. Sub-sample estimates suggest that the *Impact_Score* is more informative when disclosures are assessable to the model, longer, and more technical. Formal interaction tests, however, support only the technical-sophistication dimension: the association strengthens with the Flesch–Kincaid grade level of the disclosure, whereas differences by interpretability, length, and numeric content are not statistically significant.

We begin by assessing whether the extracted severity signal depends on how coherent and assessable the disclosure is for the model. If a disclosure is semantically clear and internally coherent, then the severity assessment extracted from the text should contain more economically relevant information. Conversely, when disclosures are ambiguous, incomplete, or internally inconsistent, the noise component u_i in the extracted score in Equation (1) becomes large relative to the latent severity construct, making the realised *Impact_Score* behave like a weak or nearly random signal.

To examine this notion of interpretability, for each event we ask ChatGPT ten times whether it can assess the breach disclosure.[11] We classify a disclosure as *Fully understood* if the model indicates assessability in all ten assessments. Out of 238 events, 219 are classified as *Fully understood*. For each group, we repeat the stochastic regression experiment and summarise the distribution of the estimated *Impact_Score* coefficients and corresponding p-values in Figure 7. The association between *Impact_Score* and market reactions is concentrated in *Fully understood* disclosures. The coefficients are consistently negative and the p-values concentrate near conventional significance thresholds. In contrast, the *Not fully understood* group exhibits coefficient estimates centred near zero and p-values consistent with weak informativeness. These patterns are consistent with an information processing interpretation in which disclosure equivocality amplifies measurement noise u_i, suggesting that the extracted severity proxy may be more informative when disclosures provide more textual content.

We next examine whether the amount of information conveyed in the disclosure affects the informativeness of the extracted severity proxy. More detailed disclosures provide more opportunities for diagnostic cues (e.g., scope, systems affected, data types, remediation actions) to be articulated, which should facilitate more reliable signal extraction. We proxy information quantity by the *length* of the breach disclosure, measured as the number of characters in the breach description, and partition the sample into *Long* and *Short* descriptions using a median split. Figure 8 reports results for the two subsamples.

For events with *Short* descriptions, the *Impact_Score* coefficient is not statistically different from

[11] We emphasise that this procedure is not intended to claim a psychological notion of "understanding". Rather, it provides a practical indicator of whether the disclosure is sufficiently coherent for the model to produce a stable, assessable output.

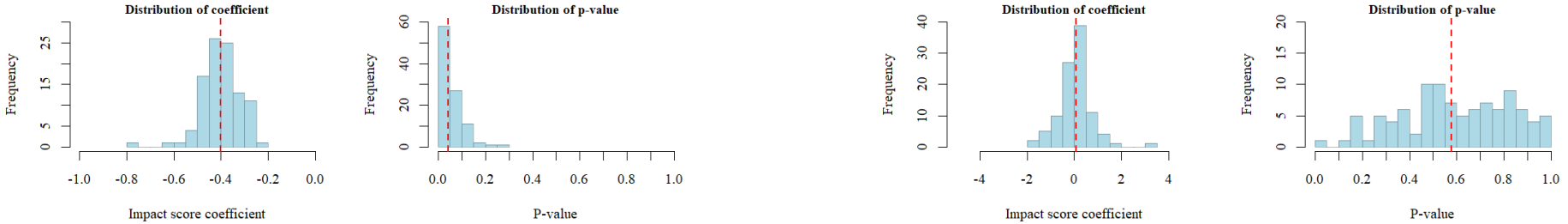
zero and p-values are well above conventional thresholds. In contrast, the coefficients for *Long* descriptions are consistently negative (centred around approximately -0.45) and the associated p-values concentrate near zero, suggesting that the extracted severity proxy may be more informative when disclosures provide more textual content. However, the difference between the two groups is not statistically significant in a fully interacted regression (median interaction 0.404 for *Short* descriptions; Fisher–Brown p-value 0.161), nor when the logarithm of disclosure length enters as a continuous interaction (Fisher–Brown p-value 0.479; Internet Appendix Tables XIV and XV). As an alternative proxy for information quantity, we also consider the number of quantitative cues in the disclosure, measured by the number of numeric tokens in the text. This measure is highly correlated with disclosure length (see Table A2 of the Appendix). Consistent with the length-based results, Figure A2 of the Appendix shows that the coefficients are similar for data-light and data-heavy disclosures (medians of -0.39 and -0.36), although the Fisher–Brown test rejects the null only for data-heavy disclosures (p-values 0.115 and 0.007), and the difference is not statistically significant in formal interaction tests (Fisher–Brown p-values of 0.954 for the sub-sample split and 0.898 for the logarithm of the numeric count).

Length captures the quantity of disclosure but does not directly reflect information quality. To better capture informational and conceptual content, we use the Flesch–Kincaid Grade Level (FKG) score (Kincaid et al., 1975). Although originally developed as a readability metric, the FKG has been widely used in the finance and accounting literature on professional, financial, and legal disclosures to capture linguistic and technical sophistication (see, e.g., Li, 2008; Lehavy et al., 2011; Loughran and McDonald, 2014). In such settings, higher FKG values typically reflect the use of longer sentences and more complex technical or legal terminology, which may convey richer informational content. In the context of data breach disclosures, we interpret higher FKG as indicating more technically sophisticated narratives that may contain more diagnostic details about the breach and its potential economic consequences. We note that the FKG score is nearly orthogonal to disclosure length (with a correlation coefficient of 0.06, see Table A2 of the Appendix), suggesting that it captures a distinct information dimension not reflected in text quantity.

We partition the sample based on the median FKG score and re-estimate the stochastic regressions for low- and high-FKG disclosures. Figure 9 reports the corresponding results. The relationship between *Impact_Score* and market reactions is more pronounced for the *High FKG* group, consistent with the view that technically specific disclosures provide more economically diagnostic cues for extracting severity. The binary split is not statistically significant in a fully interacted regression (median interaction −0.258; Fisher–Brown p-value 0.407). When the FKG score enters as a continuous interaction, however, the association becomes significantly more negative as technical sophistication increases (median interaction −0.149; Fisher–Brown p-value 0.001; Internet Appendix Table XV). Because the marginal effect of *Impact_Score* on CAR in Equation (5) equals $\beta_1 + \beta_3 \times FKG$, each additional grade level makes the association between a one-standard-deviation increase in *Impact_Score* and CAR(-1,+1) about 0.15 percentage points more negative. As a robustness check, we alternatively use the Flesch Reading Ease (FRE) score (Flesch, 1948). The results are qualitatively unchanged and are presented in Figure A1 of the Appendix. The continuous interaction confirms this: the association becomes significantly weaker as readability increases (median interaction 0.020; Fisher–Brown p-value 0.033; Internet Appendix Table XV).

Figure 7: Results by Disclosure Interpretability. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for disclosures classified as *Fully understood* and *Not fully understood* based on repeated ChatGPT assessments. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Fully understood (left) and Not fully understood (right) sub-sample

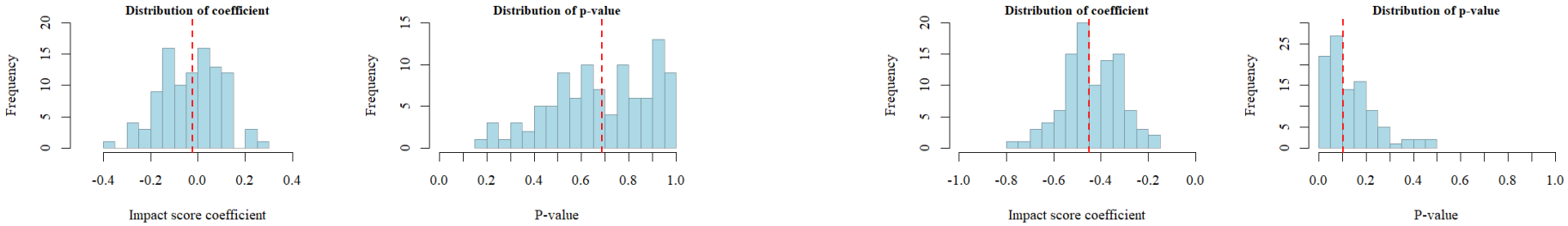


Panel B: Summary Statistics and Fisher–Brown Test Results

	Fully understood							Not fully understood						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.78	-0.45	-0.4	-0.4	-0.35	-0.21		-1.86	-0.26	0.07	0.06	0.39	3.41	
	0	*0.02*	*0.04*	*0.06*	*0.07*	*0.28*	*0.000*	*0*	*0.45*	*0.58*	*0.59*	*0.79*	*0.98*	*0.739*
Tobin's Q	-0.2	-0.15	-0.11	-0.11	-0.08	0.03		0.44	0.89	0.92	0.92	0.96	1.77	
	0.25	*0.4*	*0.53*	*0.55*	*0.64*	*0.95*	*0.739*	*0.06*	*0.09*	*0.11*	*0.12*	*0.13*	*0.47*	*0.021*
Dividend	-0.35	-0.01	0.11	0.11	0.21	0.6		-1.46	-0.69	-0.61	-0.58	-0.43	1.37	
	0.22	*0.66*	*0.79*	*0.75*	*0.9*	*1*	*0.959*	*0.32*	*0.56*	*0.64*	*0.64*	*0.72*	*0.95*	*0.898*
R&D Intensity	-1.78	0.41	1.15	1.05	1.77	4.04		-36.69	-18.73	-11.86	-12.22	-6.85	24.91	
	0.59	*0.81*	*0.87*	*0.85*	*0.93*	*0.99*	*0.995*	*0.4*	*0.58*	*0.77*	*0.72*	*0.86*	*0.98*	*0.954*
Intangibles	1.93	3.64	4.5	4.46	5.24	7.17		-9.94	4.36	8.05	7.66	10.57	23.9	
	0.1	*0.24*	*0.32*	*0.34*	*0.42*	*0.66*	*0.314*	*0.06*	*0.22*	*0.34*	*0.42*	*0.59*	*0.99*	*0.385*
Firm Size	0.14	0.2	0.22	0.22	0.25	0.29		-0.47	0.07	0.15	0.14	0.19	0.54	
	0	*0.01*	*0.03*	*0.04*	*0.05*	*0.17*	*0.000*	*0.14*	*0.6*	*0.67*	*0.69*	*0.82*	*1*	*0.930*
Leverage	-0.92	-0.73	-0.64	-0.64	-0.55	-0.39		-2.63	-0.56	0.17	0.11	0.51	3.11	
	0.19	*0.31*	*0.37*	*0.37*	*0.43*	*0.57*	*0.422*	*0.34*	*0.76*	*0.84*	*0.81*	*0.89*	*0.99*	*0.991*
Constant	-3.18	-2.43	-2.13	-2.13	-1.8	-1.07		-7.9	-3.76	-3.41	-3.33	-2.76	1.04	
	0.01	*0.05*	*0.08*	*0.11*	*0.15*	*0.41*	*0.01*	*0.03*	*0.23*	*0.33*	*0.36*	*0.45*	*0.87*	*0.344*
Observations	179	183	185	185	187	192		14	15	16	16	17	18	
R^2	0.06	0.08	0.08	0.09	0.1	0.16		0.24	0.36	0.4	0.39	0.42	0.56	

Figure 8: Results by Disclosure Length. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for breaches with *Short* and *Long* textual descriptions based on a median split of character counts. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Short (left) and Long (right) sub-samples



Panel B: Summary Statistics and Fisher–Brown Test Results

	Short description							Long description						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.36	-0.12	-0.02	-0.03	0.06	0.26		-0.76	-0.52	-0.45	-0.44	-0.35	-0.19	
	0.18	*0.53*	*0.69*	*0.68*	*0.87*	*0.99*	*0.895*	*0*	*0.06*	*0.1*	*0.13*	*0.18*	*0.47*	*0.002*
Tobin's Q	-0.16	-0.11	-0.1	-0.07	-0.07	0.09		-0.31	-0.22	-0.16	-0.17	-0.13	0.11	
	0.41	*0.57*	*0.61*	*0.63*	*0.69*	*0.95*	*0.872*	*0.19*	*0.37*	*0.5*	*0.49*	*0.6*	*0.99*	*0.637*
Dividend	-1.22	-1.08	-1.03	-1.03	-0.98	-0.75		-0.1	0.35	0.6	0.59	0.78	1.23	
	0.03	*0.06*	*0.08*	*0.08*	*0.09*	*0.21*	*0.008*	*0.04*	*0.22*	*0.37*	*0.39*	*0.59*	*0.89*	*0.337*
R&D Intensity	-18.71	-12.75	-12.44	-12.48	-11.87	-8.64		4.95	8.12	9.52	9.31	10.95	13.53	
	0.12	*0.19*	*0.2*	*0.21*	*0.23*	*0.39*	*0.125*	*0.01*	*0.04*	*0.07*	*0.1*	*0.13*	*0.34*	*0.006*
Intangibles	7.43	9.46	9.83	9.87	10.1	14.07		2.73	5.14	7.1	6.93	8.07	11.74	
	0.06	*0.11*	*0.12*	*0.12*	*0.14*	*0.26*	*0.031*	*0.02*	*0.12*	*0.19*	*0.23*	*0.32*	*0.59*	*0.092*
Firm Size	0.31	0.36	0.37	0.37	0.39	0.45		0.05	0.15	0.18	0.18	0.23	0.29	
	0	*0*	*0.01*	*0.01*	*0.01*	*0.02*	*0.000*	*0.03*	*0.09*	*0.17*	*0.21*	*0.28*	*0.74*	*0.065*
Leverage	-2.28	-2.03	-1.95	-1.94	-1.85	-1.59		-0.6	-0.1	0.01	0.03	0.17	0.41	
	0.02	*0.04*	*0.05*	*0.06*	*0.07*	*0.15*	*0.003*	*0.41*	*0.77*	*0.87*	*0.84*	*0.95*	*1*	*0.994*
Constant	-3.59	-2.69	-2.55	-2.58	-2.42	-2.09		-4.15	-2.99	-2.43	-2.49	-2	-0.85	
	0.02	*0.08*	*0.09*	*0.09*	*0.1*	*0.14*	*0.012*	*0.02*	*0.09*	*0.16*	*0.19*	*0.25*	*0.64*	*0.045*
Observations	95	99	100	100	101	104		98	100	101	101	102	105	
R^2	0.1	0.12	0.13	0.13	0.13	0.17		0.15	0.18	0.19	0.2	0.22	0.3	

Figure 9: Results by Technical Sophistication in Breach Descriptions. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for disclosures with *Low* and *High* technical sophistication based on a median split of the Flesch–Kincaid Grade Level (FKG). In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Low (left) and High (right) Flesch–Kincaid Grade level description



Panel B: Summary Statistics and Fisher–Brown Test Results

	Low FKG score							High FKG score						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.75	-0.33	-0.21	-0.21	-0.09	0.27		-1.23	-0.52	-0.45	-0.48	-0.4	-0.19	
	0	*0.24*	*0.43*	*0.47*	*0.72*	*0.98*	*0.165*	*0*	*0.02*	*0.04*	*0.05*	*0.06*	*0.44*	*0.000*
Tobin's Q	-0.09	0.09	0.18	0.17	0.26	0.38		-0.45	-0.38	-0.36	-0.35	-0.32	-0.17	
	0.1	*0.23*	*0.38*	*0.45*	*0.66*	*0.98*	*0.440*	*0.04*	*0.11*	*0.15*	*0.16*	*0.19*	*0.5*	*0.049*
Dividend	-1.77	-0.87	-0.6	-0.62	-0.41	0.11		-0.04	0.24	0.32	0.32	0.41	0.65	
	0.01	*0.3*	*0.45*	*0.47*	*0.61*	*0.99*	*0.518*	*0.23*	*0.43*	*0.54*	*0.54*	*0.65*	*0.96*	*0.738*
R&D Intensity	-7.35	-3.62	-2.44	-2.31	-0.63	3.65		2.23	5.25	6.17	6.05	6.91	9.13	
	0.12	*0.49*	*0.66*	*0.65*	*0.86*	*0.99*	*0.855*	*0.43*	*0.55*	*0.58*	*0.59*	*0.64*	*0.82*	*0.834*
Intangibles	3.34	6.27	8.23	8.03	9.84	13.03		-2.66	-0.65	-0.05	0.16	0.8	3.55	
	0.01	*0.04*	*0.12*	*0.15*	*0.24*	*0.54*	*0.016*	*0.63*	*0.85*	*0.92*	*0.9*	*0.95*	*1*	*0.999*
Firm Size	0.19	0.31	0.33	0.34	0.37	0.46		0.12	0.17	0.19	0.19	0.21	0.32	
	0	*0.01*	*0.01*	*0.02*	*0.03*	*0.18*	*0.000*	*0.03*	*0.12*	*0.17*	*0.18*	*0.23*	*0.39*	*0.063*
Leverage	-0.66	-0.39	-0.23	-0.23	-0.05	0.17		-3.16	-2.76	-2.6	-2.57	-2.4	-1.82	
	0.18	*0.41*	*0.65*	*0.62*	*0.82*	*0.99*	*0.813*	*0.000*	*0.01*	*0.01*	*0.01*	*0.02*	*0.08*	*0.000*
Constant	-5.54	-4.24	-3.69	-3.69	-3.18	-1.74		-2.24	-0.99	-0.78	-0.81	-0.53	-0.05	
	0	*0.01*	*0.03*	*0.05*	*0.06*	*0.35*	*0.000*	*0.24*	*0.54*	*0.64*	*0.64*	*0.74*	*0.98*	*0.869*
Observations	96	100	100	100	101	104		97	100	101	101	102	105	
R^2	0.04	0.08	0.1	0.1	0.12	0.25		0.15	0.18	0.19	0.19	0.2	0.24	

In summary, Figures 7 through 9 show that the informativeness of the ChatGPT-derived severity proxy is not uniform across breach disclosures. Sub-sample estimates suggest that the association between *Impact_Score* and short-window abnormal returns is strongest when the disclosure is assessable to the model and when the narrative is longer and more technical. Interpreting these results through the lens of bounded information processing, these disclosure attributes plausibly reduce the measurement noise component u_i in Equation (1) by increasing the availability and structure of severity relevant cues. Alternatively, when disclosures are short, ambiguous, or lack concrete quantitative/technical detail, the extracted score may behave like a weak signal, consistent with a high-u_i regime in which reliable severity inference is difficult.

Findings obtained in this section motivate the subsequent heterogeneity analysis and help clarify when LLM-based assessments can be expected to support managerial and investor decision-making under incomplete information.

4.5. Types of Breach

We next examine whether the association between *Impact_Score* and market reactions varies across breach types (H2a). The consequences of internally driven incidents, such as employee error or accidental disclosure, are typically more transparent, so their severity can be assessed more precisely from the disclosure text. Externally driven attacks involve more complex dynamics and attacker motives that are harder to evaluate at the time of disclosure. If the score measures the severity of external breaches with more error, its estimated association with returns will be attenuated for these events, even if investors respond strongly to their underlying severity.

Following Foerderer and Schuetz (2022), we partition the sample using the PRC breach-type classification into *Hack* and *Not hack* events. The *Hack* category captures externally driven incidents such as hacks and ransomware attacks, while the *Not hack* category captures incidents that are more plausibly internal or insider-related, including employee error, accidental disclosure, or theft. We then repeat the stochastic estimation procedure separately for each sub-sample.

Panel A of Figure 10 presents the corresponding coefficient and p-value distributions. The sub-sample estimates differ markedly across breach types. For *Hack* events, the estimated coefficient on *Impact_Score* is centred near zero, with a median of approximately 0.00, and the associated p-values are broadly distributed and concentrated far from conventional significance thresholds. The Fisher–Brown test fails to reject the null that the set of p-values behaves as under no association. In contrast, for *Not hack* events, the coefficient on *Impact_Score* is consistently negative across runs, with a median estimate of approximately -0.42, and the associated p-values are concentrated near zero. The Fisher–Brown test rejects the null at the 1% level, indicating aggregate evidence of a non-zero association in this sub-sample.

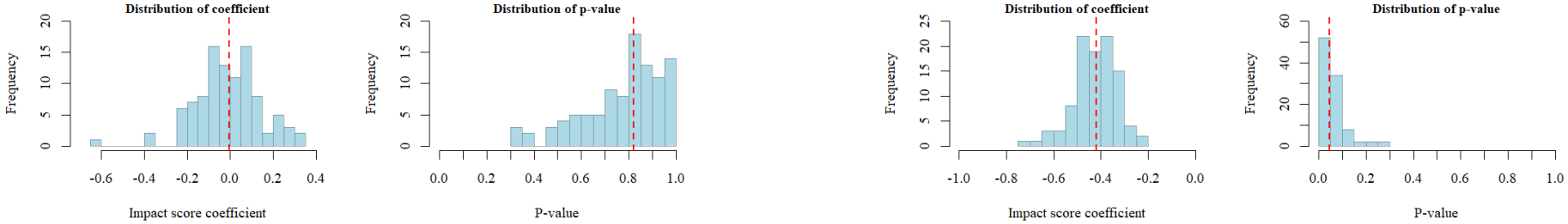
Differences in statistical significance across sub-samples do not, however, establish that the two coefficients differ. We therefore estimate a fully interacted regression in which the coefficient on *Impact_Score* and the coefficients on all firm-level controls may differ between *Hack* and *Not hack* events. The interaction coefficient is positive, consistent with a weaker association for *Hack* events, but it is not statistically significant (median 0.426; Fisher–Brown p-value 0.266; Internet Appendix Table XIV). The breach-type evidence is therefore directionally consistent with H2a but does not support it formally. Hacking incidents also do not generate larger announcement reactions on average: in a regression of CAR$(-1,+1)$ on a *Hack* indicator and the baseline controls, the coefficient is positive and statistically insignificant (0.465, standard error 0.409; Internet Appendix Table XIII).

4.6. Firm Characteristics

We next examine whether the informativeness of the ChatGPT-derived severity proxy varies with firm characteristics (H2b and H2c). From an information processing and decision relevance perspective, the economic meaning of a breach disclosure need not be constant across firms. Investors may assign different weights to otherwise similar breach information depending on firm size, industry specific data dependence, regulatory exposure, and the extent to which a breach can be interpreted as threatening the firm's core operating model. We therefore study heterogeneity along two dimensions: firm size and industry sector.

Figure 10: Results by Breach Type. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for externally driven breaches (Hack) and internally driven breaches (Not hack). In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Hack (left) and Not Hack (right) breaches



Panel B: Summary Statistics and Fisher–Brown Test Results

	Hack							Not hack						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.63	-0.08	0	-0.01	0.08	0.33		-0.74	-0.48	-0.42	-0.43	-0.38	-0.23	
	0.33	*0.68*	*0.82*	*0.77*	*0.9*	*1*	*0.976*	*0*	*0.02*	*0.05*	*0.06*	*0.08*	*0.29*	*0.000*
Tobin's Q	0	0.17	0.22	0.2	0.24	0.36		-0.34	-0.27	-0.23	-0.23	-0.2	-0.07	
	0.02	*0.09*	*0.14*	*0.26*	*0.28*	*1*	*0.067*	*0.05*	*0.13*	*0.22*	*0.23*	*0.29*	*0.7*	*0.118*
Dividend	-1.07	-0.5	-0.21	-0.21	0	0.62		0.02	0.19	0.28	0.31	0.42	0.73	
	0.27	*0.54*	*0.68*	*0.68*	*0.84*	*1*	*0.912*	*0.18*	*0.43*	*0.61*	*0.59*	*0.74*	*0.97*	*0.799*
R&D Intensity	-14.01	-7.89	-7.35	-7.32	-6.58	-2.01		-0.61	2.17	3.38	3.27	4.36	7.67	
	0.07	*0.29*	*0.34*	*0.35*	*0.39*	*0.81*	*0.348*	*0.26*	*0.53*	*0.62*	*0.64*	*0.74*	*0.99*	*0.869*
Intangibles	-0.1	2.95	3.65	3.62	4.29	8.53		3.33	6.22	7.16	7.31	8.43	11.36	
	0.11	*0.42*	*0.48*	*0.5*	*0.56*	*0.98*	*0.664*	*0.04*	*0.14*	*0.21*	*0.22*	*0.29*	*0.57*	*0.105*
Firm Size	-0.25	-0.09	0.02	0	0.08	0.2		0.21	0.28	0.3	0.3	0.32	0.36	
	0.25	*0.51*	*0.71*	*0.68*	*0.86*	*0.99*	*0.906*	*0*	*0.01*	*0.01*	*0.01*	*0.02*	*0.07*	*0.000*
Leverage	-2.16	-1.65	-1.23	-1.05	-0.71	0.64		-0.56	-0.37	-0.31	-0.3	-0.22	-0.03	
	0.13	*0.29*	*0.44*	*0.49*	*0.65*	*0.99*	*0.573*	*0.48*	*0.64*	*0.69*	*0.7*	*0.77*	*0.97*	*0.938*
Constant	-1.79	-0.66	0.12	0.17	0.92	2.75		-4.24	-3.58	-3.27	-3.25	-2.95	-2.07	
	0.18	*0.56*	*0.69*	*0.68*	*0.83*	*0.99*	*0.907*	*0.01*	*0.02*	*0.04*	*0.05*	*0.06*	*0.17*	*0.001*
Observations	43	45	46	46	47	52		150	153	155	155	156	160	
R^2	0.04	0.06	0.07	0.08	0.09	0.14		0.11	0.13	0.14	0.15	0.16	0.21	

4.6.1. Firm size

We begin with firm size. Following H2b, a breach of given severity represents a larger share of a smaller firm's value, and smaller firms have fewer financial and reputational buffers with which to absorb breach costs. The association between *Impact_Score* and short-window market reactions should therefore be stronger for smaller firms than for larger firms.

Figure 11 reports the results from splitting the sample at the median of the firm-size variable. Because the sample is stratified by size, we omit *Size* from the control set in this analysis. Across both sub-samples, the estimated coefficient on *Impact_Score* remains negative, indicating that the directional relationship documented in the baseline specification continues to hold. However, the association is stronger for smaller firms. The mean coefficient is approximately −0.45 for the small-firm sub-sample, compared with approximately −0.30 for the large-firm sub-sample. In addition, the Fisher–Brown test rejects the null of no association at the 1% level for small firms (p-value 0.006) and at the 10% level for large firms (p-value 0.063).

These sub-sample results are directionally consistent with H2b. To test whether the difference is statistically significant, we estimate a fully interacted regression that allows the *Impact_Score* coefficient and the coefficients on the remaining controls to differ between small and large firms. The interaction coefficient is negative, consistent with a stronger association for smaller firms, but it is not statistically significant (median −0.151; Fisher–Brown p-value 0.774; Internet Appendix Table XIV). When firm size enters as a continuous interaction, the interaction is positive and significant at the 10% level (median 0.131; Fisher–Brown p-value 0.066; Internet Appendix Table XV), indicating a weaker association for larger firms. The evidence for H2b is therefore directionally consistent but weak.

Figure 11: Results by Firm Size. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for small and large firms based on a median split of firm size. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results. In this experiment, the firm size is omitted from the control set.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for small (left) and large (right) firms



Panel B: Summary Statistics and Fisher–Brown Test Results

	Small firms							Large firms						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.78	-0.54	-0.46	-0.45	-0.37	-0.18		-1.75	-0.34	-0.3	-0.3	-0.22	-0.06	
	0	*0.07*	*0.12*	*0.16*	*0.2*	*0.56*	*0.006*	*0.01*	*0.11*	*0.16*	*0.2*	*0.27*	*0.8*	*0.063*
Tobin's Q	-0.25	-0.18	-0.13	-0.13	-0.09	0.01		-0.64	-0.59	-0.54	-0.54	-0.48	-0.43	
	0.26	*0.43*	*0.55*	*0.58*	*0.69*	*1*	*0.780*	*0.02*	*0.04*	*0.06*	*0.06*	*0.08*	*0.13*	*0.003*
Dividend	-0.69	-0.35	-0.17	-0.18	-0.03	0.35		0.29	0.51	0.66	0.63	0.76	0.96	
	0.26	*0.55*	*0.74*	*0.71*	*0.87*	*1*	*0.932*	*0.11*	*0.19*	*0.24*	*0.27*	*0.35*	*0.57*	*0.202*
R&D Intensity	-3.14	-0.03	1.24	1.01	2.09	4.83		13.33	16.1	17.17	17.17	18.3	20.85	
	0.57	*0.79*	*0.86*	*0.85*	*0.93*	*1*	*0.995*	*0.05*	*0.11*	*0.12*	*0.12*	*0.14*	*0.19*	*0.031*
Intangibles	-0.54	1.39	2.17	2.41	3.2	6.36		0.77	4.25	4.85	4.82	5.35	7.3	
	0.25	*0.56*	*0.69*	*0.67*	*0.81*	*0.99*	*0.900*	*0.15*	*0.3*	*0.34*	*0.36*	*0.41*	*0.87*	*0.389*
Leverage	-0.23	0.08	0.2	0.2	0.32	0.55		-2.9	-2.62	-2.52	-2.5	-2.39	-1.93	
	0.38	*0.62*	*0.75*	*0.74*	*0.88*	*1*	*0.961*	*0*	*0*	*0*	*0*	*0*	*0.01*	*0.000*
Constant	-0.94	-0.53	-0.28	-0.3	-0.1	0.22		0.41	0.68	0.77	0.77	0.87	1.02	
	0.25	*0.53*	*0.74*	*0.71*	*0.88*	*0.99*	*0.925*	*0.1*	*0.19*	*0.24*	*0.26*	*0.33*	*0.57*	*0.187*
Observations	96	100	100	100	101	104		97	100	101	101	102	105	
R^2	0.03	0.05	0.06	0.06	0.06	0.17		0.17	0.19	0.19	0.19	0.2	0.22	

4.6.2. Sector

We next examine cross-sector heterogeneity. Industry context can shape both the expected consequences of a breach and the way investors interpret disclosed breach information. Firms in sectors that rely heavily on customer trust, handle especially sensitive information, or face elevated regulatory scrutiny may experience stronger market reactions to severity-relevant breach disclosures. In such settings, a given textual cue may be more economically diagnostic, making the extracted severity proxy more informative.

The baseline sample contains five sector groupings based on the PRC organisation-type classification: *BFS*, covering Financial Services, Banking, and Insurance firms; *BSO*, including Manufacturing, Technology, and Communications firms; *BSR*, covering retail- and merchant-oriented firms; *MED*, consisting of healthcare and medical providers; and a residual *Unknown* category (UNKN). We partition the sample accordingly and repeat the stochastic estimation separately for each sector. Table 2 and Figure 12 summarise the resulting distributions of the *Impact_Score* coefficient and its associated p-values.

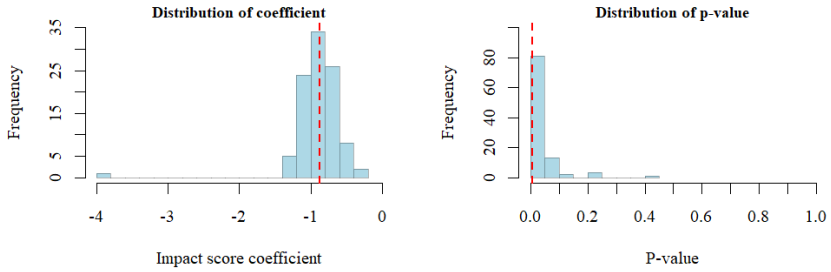
Table 2: Sector sub-sample regression results. The table reports summary statistics for the *Impact_Score* coefficient from equation (3) across 100 stochastic regressions, separately for each sector sub-sample. For each sub-sample, the first row reports descriptive statistics of the coefficient estimates, while the second row (in *italics*) reports the corresponding statistics for the p-values testing the coefficient against zero. The final column reports p-values from the Fisher–Brown test of the null hypothesis that the p-values follow a uniform distribution.

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
BFS sub-sample							
Impact_Score	-3.93	-1.02	-0.87	-0.91	-0.76	-0.34	
	0	*0*	*0.01*	*0.03*	*0.03*	*0.41*	*0.000*
BSO sub-sample							
Impact_Score	-0.33	-0.09	-0.01	-0.01	0.07	0.41	
	0.15	*0.53*	*0.74*	*0.71*	*0.9*	*1*	*0.924*
BSR sub-sample							
Impact_Score	-2.86	-0.67	-0.5	-0.56	-0.34	0.17	
	0.06	*0.21*	*0.35*	*0.39*	*0.53*	*0.97*	*0.345*
MED sub-sample							
Impact_Score	-1.12	-0.58	-0.4	-0.4	-0.19	0.38	
	0	*0.14*	*0.27*	*0.35*	*0.53*	*0.98*	*0.073*
UNKN sub-sample							
Impact_Score	-2.39	-0.87	-0.62	-0.65	-0.44	0.19	
	0	*0.15*	*0.26*	*0.3*	*0.41*	*0.8*	*0.150*

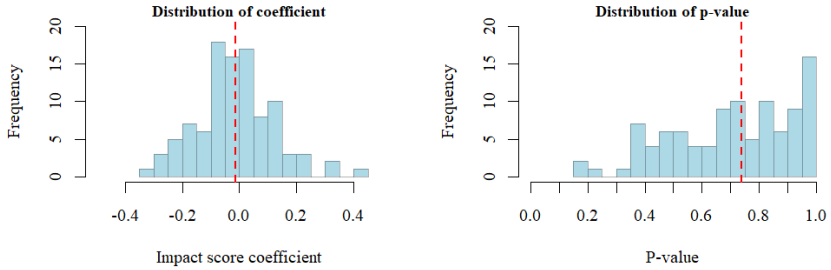
A clear sector pattern emerges. The association between *Impact_Score* and short-window market re-

Figure 12: Results by sector. The figure shows the distributions of the Impact_Score's coefficient and its *p*-value by each sector. In each histogram, the red dashed line indicates the median of the distribution.

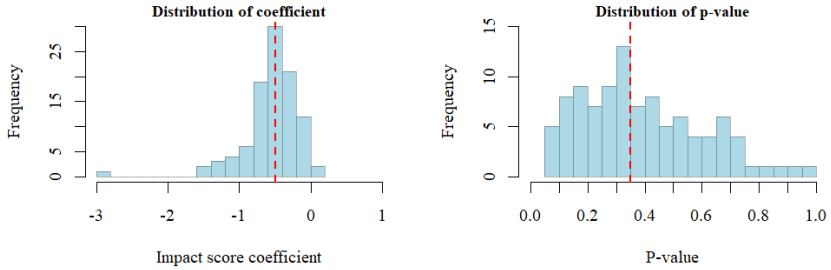
Panel A: Financial Services, Banking, Insurance Services (PRC BFS type)



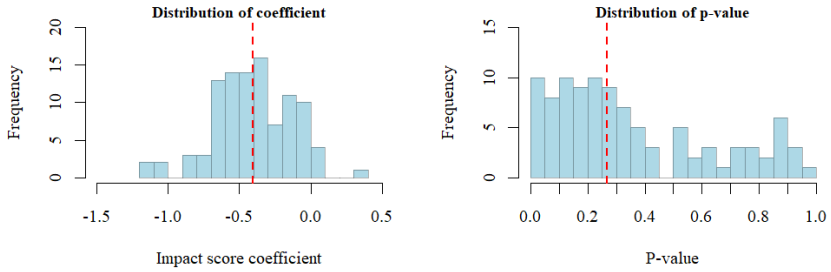
Panel B: Manufacturing, technology, and communications (PRC BSO type) firms



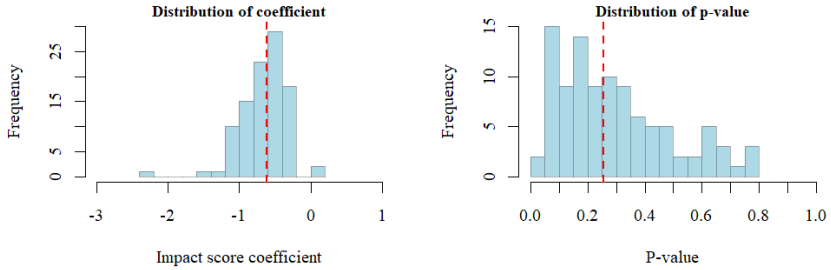
Panel C: Retail/Merchant: Grocery Stores, Online Retailers, Restaurants (PRC BSR type)



Panel D: Healthcare and Medical Providers (PRC MED type)



Panel E: Unknown sector (PRC UNKN type)



actions is strongest in the *BFS* sector. In this sub-sample, the coefficient on *Impact_Score* is consistently negative across runs, with a mean and median of approximately −0.9, and the corresponding p-values are concentrated near conventional significance thresholds. The Fisher–Brown test rejects the null of no association at the 1% level. By contrast, no comparable evidence is observed in the other sectors. For *BSO*, the coefficient distribution is centred near zero; for *MED*, the coefficients are mostly negative and the Fisher–Brown test rejects the null at the 10% level (p-value 0.073); for *BSR* and *Unknown*, the coefficient estimates are often negative but the associated p-values are too dispersed for the aggregate evidence to reject the null (p-values 0.345 and 0.150). In other words, outside the *BFS* sector the evidence is considerably weaker: only *MED* reaches statistical significance, and only at the 10% level.

To test H2c formally, we estimate a fully interacted regression in which the *Impact_Score* coefficient and the coefficients on all firm-level controls may differ for firms in financial services or healthcare (*BFS* or *MED*) relative to all other firms. The interaction coefficient is negative and statistically significant (median −0.622; Fisher–Brown p-value 0.005; Internet Appendix Table XIV), indicating a stronger association between breach severity and market reactions in these sectors. Estimating the interaction separately for each sector shows that the difference is driven mainly by financial firms (*BFS*: median −0.565, Fisher–Brown p-value 0.017), whereas the *MED* interaction is negative but not statistically significant (median −0.162, p-value 0.561).

These results support H2c, with the effect concentrated in financial services. In financial services, banking, and insurance, breach disclosures likely carry unusually high economic importance because these firms depend heavily on data integrity, customer trust, and regulatory compliance. As a result, investors can more easily link the reported details of a breach to likely operational, legal, and reputational damage. In these cases, the cues used by the model to create *Impact_Score* are closer to the cues that investors use when judging firm value. By contrast, in other sectors the same textual information may be less tightly connected to near-term valuation consequences, or the relevant consequences may be more diffuse and harder to infer from the initial disclosure.

We also examine whether the association varies with firms' exposure to securities litigation, mea-

sured by the predicted litigation probability of Kim and Skinner (2012) in the year before the breach. The association is somewhat stronger for firms at or above the median litigation risk (Fisher–Brown p-value 0.085, against 0.376 for below-median firms), but the difference is not statistically significant in a fully interacted regression (median interaction −0.121; Fisher–Brown p-value 0.818; Internet Appendix Figure I and Table XIV).

Overall, these findings reinforce our central argument that the informativeness of LLM-generated assessments is context-dependent. The extracted severity proxy is not uniformly informative across firms. Formal interaction tests confirm a stronger association for firms in financial services and healthcare, driven mainly by financial firms, where breach disclosures are likely to be interpreted as more directly relevant to core operations, governance, and stakeholder trust. The differences by breach type and firm size are directionally consistent with our predictions but not statistically significant.

4.7. The Source of Heterogeneity

In the previous sections, sub-sample estimates suggested that the association between *Impact_Score* and short-window market reactions varies with disclosure characteristics, breach type, and firm attributes. Formal interaction tests confirm the sector difference but not the differences by breach type or firm size. Now we examine whether these heterogeneous results are primarily driven by variation in the quality of the textual input itself, or whether they reflect distinct economic contexts beyond the disclosure text. Based on the framework of Equation (1), this analysis helps distinguish whether cross-sub-sample differences arise mainly from variation in the measurement noise component u_i, as some groups provide richer or more assessable disclosures, or alternatively from differences in how investors map breach severity into firm value.

To investigate this issue, Table 3 reports summary statistics for two textual attributes, the disclosure length and the FKG score, across the main sub-samples used in Sections 4.5 and 4.6. Disclosure length captures the quantity of textual information available to the model, while FKG captures the technical and linguistic complexity of the disclosure. Taken together, these two measures provide a simple way to assess whether the cross-sectional heterogeneity documented earlier could be mechanically explained

by differences in input richness or textual sophistication.

Panel A of Table 3 shows that disclosures associated with *Hack* events are, on average, substantially longer than those associated with *Not hack* events. The median description length is approximately 745 characters for *Hack* events, compared with approximately 348 characters for *Not hack* events. If the stronger association documented before was driven primarily by the amount of text available to the model, one would expect the relationship between *Impact_Score* and market reactions to be strongest in the *Hack* sub-sample. Figure 10 shows the opposite pattern, the association is present in *Not hack* events and is largely absent for *Hack* events. This contrast indicates that breach type and disclosure length capture distinct and complementary sources of variation. In other words, to the extent that the association is stronger for non-hacking breaches, this cannot be attributed to those disclosures being longer.

Table 3: Disclosure length and textual complexity across sub-samples. This table reports summary statistics for the length of data breach descriptions and the Flesch–Kincaid Grade Level (FKG) score across sub-samples defined by breach type, firm size, and industry sector. Panel A presents the distribution of description length, measured by the number of characters in the disclosure text. Panel B reports the distribution of FKG scores, capturing the technical complexity of disclosure language. The statistics include the minimum, quartiles, mean, and maximum.

Group	Min	Q1	Median	Mean	Q3	Max
Panel A: Description length by sub-samples						
Hack	81	410	745	1576	1412	9148
Not hack	42	213	348	560	574	9045
Small firm	71	251	452	992	838	9148
Large firm	42	226	382	587	715	4359
BFS	101	256	404	663	743	4359
BSO	42	236	394	854	722	8827
BSR	128	344	468	607	631	1864
MED	71	158	420	684	894	4962
UNKN	78	210	278	1277	928	9148
Panel B: FKG score by sub-samples						
Hack	8	12	15	15	16	26
Not hack	8	12	14	14	15	22
Small firm	9	12	14	14	16	26
Large firm	8	12	14	14	16	25
BFS	9	12	14	14	16	25
BSO	8	12	14	14	15	17
BSR	8	12	13	13	14	17
MED	10	12	14	14	15	26
UNKN	9	13	14	14	16	22

Panel A also shows that smaller firms tend to provide longer disclosures than larger firms. The median description length is approximately 452 characters for small firms, compared with approximately 382 characters for large firms, and the mean is likewise higher for the small firm sub-sample. This pattern is directionally consistent with the previously reported association observed for smaller firms. Accordingly, part of the size based heterogeneity may reflect the fact that smaller firms provide richer textual inputs, which reduce the noise in the model extracted severity proxy. At the same time, the difference in length is economically modest relative to the breach-type contrast, suggesting that firm size heterogeneity may reflect a combination of disclosure richness and underlying differences in the economic consequences of breaches for smaller firms.

On a different note, the sector results do not appear to be explained by simple differences in disclosure length. The *BFS* sub-sample, where the association between *Impact_Score* and short-window market reactions is strongest, does not exhibit systematically longer disclosures than the remaining sectors. Its median description length is approximately 404 characters, which is similar to *BSO* (394), below *BSR* (468), and close to *MED* (420). This evidence suggests that the pronounced sector effect documented in Section 4.6.2 is not mechanically driven by richer textual input. Instead, it is more consistent with the interpretation that breach disclosures in financial services, banking, and insurance carry greater economic relevance as they map more directly into concerns about trust, compliance, and operational continuity.

Panel B of Table 3 further shows that FKG scores are broadly similar across sub-samples. Median FKG values are tightly clustered around 14 and 15 for most groups, and no systematic pattern emerges that mirrors the heterogeneity reported in the previous sections. This result implies that the technical complexity of disclosure language is largely orthogonal to breach type, firm size, and sector in this sample. Put differently, the heterogeneity documented earlier is unlikely to be driven by systematic differences in textual sophistication alone.

Summarising our findings, Table 3 helps to clarify the source of heterogeneity in our results. Some of the sub-sample variation, especially across the firm-size groups, may be related to differences in

disclosure richness, which can improve the informativeness of an LLM-generated severity proxy by lowering the measurement noise component u_i. However, the breach-type and sector patterns cannot be explained by disclosure length or FKG differences alone. These findings suggest that the heterogeneity documented before reflects not only variation in the quality of textual inputs, but also deeper differences in the economic meaning and decision relevance of disclosed breach information across contexts. More broadly, the results reinforce the paper's central claim that the usefulness of LLM-generated assessments depends jointly on the properties of the input text, the breach type, and the organisational setting in which that text is interpreted.

5. Conclusion

This paper examines whether an LLM can extract economically meaningful signals from unstructured corporate data breach disclosures. We develop a ChatGPT-based impact score that maps narrative breach descriptions into a structured proxy for breach severity, and we evaluate whether this text-derived measure covaries with firms' short-window abnormal stock returns around breach announcements. To account for the stochastic nature of LLM outputs, we repeat the scoring process across multiple random seeds, aggregate evidence across the resulting estimates, and confirm the main result using a conventional regression on the score averaged across runs. We then investigate the boundary conditions under which the extracted signal is more or less informative by examining disclosure characteristics, breach types, and firm-level heterogeneity.

Our results show that the ChatGPT-derived impact score is negatively associated with short-window market reactions, indicating that the model is able to extract severity-relevant information from disclosure text. The score captures information beyond simple proxies for breach severity, such as breach size, sentiment, and disclosure length, and it is distinct from existing measures of firms' ex ante exposure to cyber risk. It is not associated with pre-announcement returns, post-announcement returns suggest, if anything, a partial reversal of the initial reaction, and its associations with subsequent operating performance and predicted litigation risk are weak or insignificant. At the same time, this informativeness is

not uniform. The relationship strengthens with the technical sophistication of the disclosure, consistent with the view that more diagnostic inputs reduce measurement noise in the extracted signal. Sub-sample estimates also suggest stronger associations for longer and more assessable disclosures, but these differences are not statistically significant. Across contexts, the association is significantly stronger for firms in financial services and healthcare, driven mainly by financial firms. It is also stronger for non-hacking breaches and smaller firms, but these differences are not statistically significant. Additional analyses suggest that these patterns are not explained solely by simple differences in disclosure length or textual complexity, but also reflect differences in the economic meaning and decision relevance of breach information across organisational settings.

Our findings contribute to the literature by showing how an LLM can convert unstructured operational disclosures into a structured severity signal, while also identifying the information and contextual conditions under which that signal is more or less informative. More broadly, our paper highlights that the usefulness of AI-generated assessments depends jointly on the richness of the underlying information environment and on the context in which that information is interpreted. For practice, the results suggest that firms, analysts, and risk managers may use LLM-based severity scores as an initial triage input to prioritise disclosures for further review, particularly when disclosures are technically detailed rather than sparse or ambiguous. At the same time, the findings also underscore the need for governance and human oversight: LLM outputs should be interpreted as noisy advisory signals, not as direct measures of true underlying severity. Future research can build on this approach by studying other operational events, alternative model architectures, and organisational mechanisms that improve the quality and comparability of textual disclosures.

Disclosure statement

The authors report there are no competing interests to declare.

Funding

No funding was received.

References

Akey, P., Lewellen, S., Liskovich, I., Schiller, C., 2026. Hacking corporate reputations. Review of Finance 30, 795–862.

Akyildirim, E., Conlon, T., Corbet, S., Hou, Y.G., 2024. HACKED: Understanding the stock market response to cyberattacks. Journal of International Financial Markets, Institutions and Money 97, 102082. doi:10.1016/j.intfin.2024.102082.

Bamiatzi, V., Dowling, M., Gogolin, F., Kearney, F., Vigne, S.A., 2023. Are the good spared? Corporate social responsibility as insurance against cyber security incidents. Risk Analysis 43, 2503–2518.

Boasiako, K.A., Keefe, M.O., 2021. Data breaches and corporate liquidity management. European Financial Management 27, 528–551. doi:10.1111/eufm.12289.

Bose, I., Leung, A.C.M., 2014. Do phishing alerts impact global corporations? A firm value analysis. Decision Support Systems 64, 67–78. doi:10.1016/j.dss.2014.04.006.

Brown, M.B., 1975. A method for combining non-independent, one-sided tests of significance. Biometrics 31, 987–992.

Brown, S.J., Warner, J.B., 1985. Using daily stock returns: The case of event studies. Journal of Financial Economics 14, 3–31.

Caglio, A., Melloni, G., Perego, P., 2020. Informational content and assurance of textual disclosures: Evidence on integrated reporting. European Accounting Review 29, 55–83.

Campbell, K., Gordon, L., Loeb, M., Zhou, L., 2003. The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. Journal of Computer Security 11, 431–448. doi:10.3233/JCS-2003-11308.

Cavusoglu, H., Mishra, B., Raghunathan, S., 2004. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce 9, 70–104.

Dacorogna, M., Debbabi, N., Kratz, M., 2023. Building up cyber resilience by better grasping cyber risk via a new algorithm for modelling heavy-tailed data. European Journal of Operational Research 311, 708–729. doi:10.1016/j.ejor.2023.05.012.

Das, S., Mukhopadhyay, A., Anand, M., 2012. Stock market response to information security breach: A study using firm and attack characteristics. Journal of Information Privacy and Security 8, 27–55.

Eling, M., Wirfs, J., 2019. What are the actual costs of cyber risk events? European Journal of Operational Research 272, 1109–1119.

Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25, 383–417.

Flesch, R., 1948. A new readability yardstick. Journal of Applied Psychology 32, 221–233.

Florackis, C., Louca, C., Michaely, R., Weber, M., 2023. Cybersecurity risk. The Review of Financial Studies 36, 351–407.

Foerderer, J., Schuetz, S., 2022. Data breach announcements and stock market reactions: A matter of timing? Management Science 68, 7298–7322. doi:10.1287/mnsc.2021.4264.

Gatzlaff, K., McCullough, K., 2010. The effect of data breaches on shareholder wealth. Risk Management and Insurance Review 13, 61–83.

Goel, S., Shawky, H., 2009. Estimating the market impact of security breach announcements on firm values. Information & Management 46, 404–410. doi:10.1016/j.im.2009.06.005.

Hoang, D., Wiegratz, K., 2023. Machine learning methods in finance: Recent applications and prospects. European Financial Management 29, 1657–1701. doi:10.1111/eufm.12408.

Jamilov, R., Rey, H., Tahoun, A., 2026. The anatomy of cyber risk. The Journal of Finance, forthcoming .

Kamiya, S., Kang, J.K., Kim, J., Milidonis, A., Stulz, R.M., 2021. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. Journal of Financial Economics 139, 719–749.

Kannan, K., Rees, J., Sridhar, S., 2007. Market reactions to information security breach announcements: An empirical analysis. International Journal of Electronic Commerce 12, 69–91. doi:10.2753/JEC1086-4415120103.

Kearney, C., Liu, S., 2014. Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis 33, 171–185.

Kim, I., Skinner, D.J., 2012. Measuring securities litigation risk. Journal of Accounting and Economics 53, 290–310. doi:10.1016/j.jacceco.2011.09.005.

Kincaid, J.P., Fishburne, R.P., Rogers, R.L., Chissom, B.S., 1975. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8-75. Naval Technical Training Command. Millington, TN.

Kirtac, K., Germano, G., 2024. Sentiment trading with large language models. Finance Research Letters 62, 105227. doi:10.1016/j.frl.2024.105227.

de Kok, T., 2025. ChatGPT for textual analysis? How to use generative LLMs in accounting research. Management Science 71, 7888–7906.

Lehavy, R., Li, F., Merkley, K.J., 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. The Accounting Review 86, 1087–1115.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. Journal of Accounting and Economics 45, 221–247.

Lin, X., She, G., 2025. Timely cybersecurity disclosure and information manipulation. Management Science 71, 9308–9327.

Liu, J., Ni, X., 2024. Ordeal by innocence in the big-data era: Intended data breach disclosure, unintended real activities manipulation. European Financial Management 30, 129–163. doi:10.1111/eufm.12410.

Lopez-Lira, A., Tang, Y., 2025. Can ChatGPT forecast stock price movements? Return predictability and large language models. Working Paper. doi:10.2139/ssrn.4412788.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance 66, 35–65. doi:10.1111/j.1540-6261.2010.01625.x.

Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. The Journal of Finance 69, 1643–1671.

Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: A survey. Journal of Accounting Research 54, 1187–1230.

Maarouf, A., Feuerriegel, S., Pröllochs, N., 2025. A fused large language model for predicting startup success. European Journal of Operational Research 322, 198–218.

Rosati, P., Cummins, M., Deeney, P., Gogolin, F., Van der Werff, L., Lynn, T., 2017. The effect of data breach announcements beyond the stock price: Empirical evidence on market activity. International Review of Financial Analysis 49, 146–154.

Shevchenko, P., Jang, J., Malavasi, M., Peters, G.W., Sofronov, G., Trück, S., 2023. The nature of losses from cyber-related events: Risk categories and business sectors. Journal of Cybersecurity 9, 1–12.

Siano, F., 2025. The news in earnings announcement disclosures: Capturing word context using LLM methods. Management Science 71, 9831–9855. doi:10.1287/mnsc.2024.05417.

Simon, J., Omar, A., 2020. Cybersecurity investments in the supply chain: Coordination and a strategic attacker. European Journal of Operational Research 282, 161–171. doi:10.1016/j.ejor.2019.09.017.

Wu, Z., Dong, Y., Li, Y., Shi, B., 2025. Unleashing the power of text for credit default prediction: Comparing human-written and generative AI-refined texts. European Journal of Operational Research 326, 691–706. doi:10.1016/j.ejor.2025.04.032.

Appendix

A. ChatGPT Scoring Prompt

The OpenAI API query used to produce our ChatGPT-assessed severity score is based on the prompts below.

The user prompt:

"*Forget all previous instructions. I will give you a 'Description' and an 'Explanation for Type of Breach' from Privacy Rights Clearinghouse of a data breach incident below. Using only information available prior to the {event date}, can you assess the impact on the firm using a 1-100 severity rating, with 1 being almost no impact on the firm, and 100 being the most severe impact on the firm.*

A non-exhaustive list of important factors to take into consideration when deciding on a severity rating include the:

- Type of data breached (e.g., personal data, financial data, medical records, intellectual property);

- Number of individuals affected;

- Sensitivity of the data breached;

- Visibility of the breach;

- Media coverage of the breach;

- Public reaction to the breach;

- Damage to the firm's reputation.

The 'Description of the data breach' is as following: {event description}.

The 'Explanation for Type of Breach' of the data breach is as following:

{breach explanation}.

{news prompt}

Please produce the 1-100 score of the impact of the event on the firm by using only the data that you can access directly and by ignoring any internal links in the Descriptions.

Please also give me a written explanation in a single sentence for why that score was given. Else if you can not provide an explanation, give NA.

The return must be a list consisting of the score, and the written explanation sentence, separated by a semicolon ';'. If the score is NA, Return NA;NA. Otherwise, the score part must be only the number without any text. "

The system prompt:

"*You are a financial risk model. You must only use information available before the event date which is {event date}.* "

In the prompt, for each event, its breach date is inserted into *event date*, its PRC Description is inserted into *event description* and its PRC Explanation for Type of Breach is inserted in *event explanation*.

The *news prompt* is "*News of the breach was reported in the following outlets: the outlets list.*" for those events that appeared in the news and is "" (empty) otherwise.

B. ChatGPT Comprehension Check Prompt

The user prompt is:

"*Forget all previous instructions. Please check if the description and explanation [sic] of the following data breach incident is understandable to you. If yes, return '1', if no return '0', otherwise return 'NA'.*

The 'Description';[sic] of the data breach is as following: {event description}.

The 'Explanation for Type of Breach' of the data breach is as following: {event explanation}.

The return must be only one of: 1; 0; NA "

The system prompt is: "*You are a financial risk model.*"

C. Data Breach Text Descriptions

Table A1: Data Breach Description Examples

Firm name	Breach description
Aptevo Therapeutics Corporation	A phishing attack allowed an unauthorized individual to potentially gain access to the system; Names and social security numbers.
Intel Corporation	A hacker accessed user information on an Intel website through a vulnerability. The hacker had access to credit card data, Social Security numbers, emails, passwords, and other details.
Walmart	On September 10, 2019, an associate reported that a flash drive containing patient information from the Vision Center at Walmart Store #3404 could not be located. Walmart immediately searched for the flash drive throughout the store and other relevant locations. Unfortunately, Walmart was unable to locate the flash drive. Walmart's investigation did not indicate that the situation was the result of theft or other criminal behavior. The information contained on the flash drive pertained only to information collected at the Vision Center at Walmart Store #3404, located at 1960 Twin Lakes Parkway in Roseville, Minnesota. The information included your name, date of birth, gender, retina image(s), a date associated with the image(s), and an alphanumeric code associated with the image(s). Walmart has no indication that the information has been misused in any way.

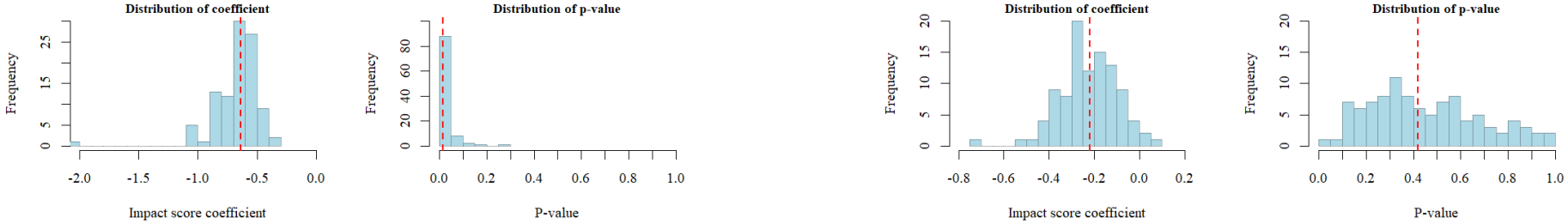
D. Additional Results

Table A2: Correlation among attributes of disclosure and information content measures

	Description Length	Count of numbers	FKG score	FRE score
Description Length	1.00	0.88	0.06	0.01
Count of numbers		1.00	-0.04	0.08
FKG score			1.00	-0.89
FRE score				1.00

Figure A1: Results by Technical Sophistication in Breach Descriptions (Flesch Reading Ease score). Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for disclosures with *Low* and *High* readability based on a median split of the Flesch Reading Ease (FRE) score. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Low (left) and High (right) Flesch Reading Ease (FRE) score descriptions

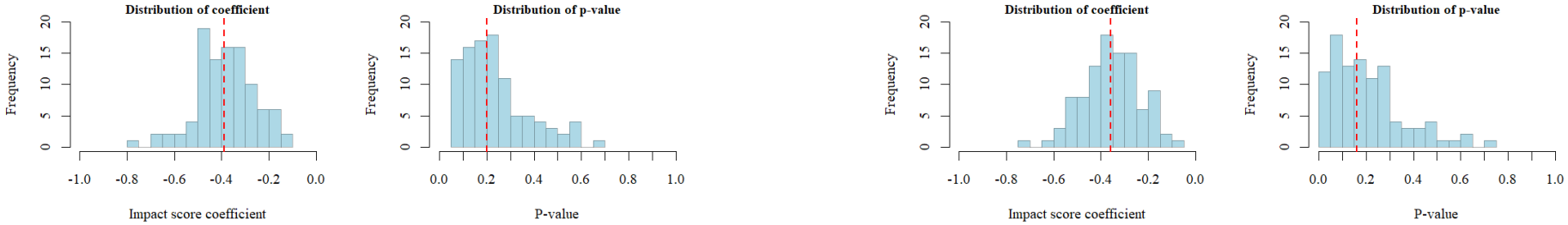


Panel B: Summary Statistics and Fisher–Brown Test Results

	Low FRE score							High FRE score						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-2.02	-0.76	-0.64	-0.68	-0.56	-0.35		-0.74	-0.29	-0.22	-0.22	-0.13	0.05	
	0	*0*	*0.02*	*0.03*	*0.03*	*0.26*	*0.000*	*0*	*0.27*	*0.42*	*0.45*	*0.59*	*0.96*	*0.148*
Tobin's Q	-0.4	-0.32	-0.3	-0.3	-0.28	-0.03		-0.15	-0.03	0.04	0.04	0.08	0.29	
	0.12	*0.21*	*0.24*	*0.26*	*0.28*	*0.87*	*0.183*	*0.17*	*0.61*	*0.74*	*0.7*	*0.87*	*1*	*0.917*
Dividend	-0.44	-0.17	-0.1	-0.1	-0.02	0.12		-0.95	-0.52	-0.39	-0.33	-0.14	0.45	
	0.42	*0.77*	*0.86*	*0.84*	*0.94*	*1*	*0.993*	*0.22*	*0.48*	*0.6*	*0.62*	*0.78*	*0.99*	*0.834*
R&D Intensity	-1.91	2.61	3.72	3.43	4.44	6.95		-6.2	-3.03	-1.51	-1.63	0.01	4.16	
	0.55	*0.69*	*0.74*	*0.76*	*0.82*	*0.99*	*0.973*	*0.27*	*0.53*	*0.74*	*0.69*	*0.9*	*1*	*0.913*
Intangibles	-1.58	1.37	2.06	2.35	3.23	6.59		3.15	6.08	7.07	7.37	8.88	11.61	
	0.33	*0.64*	*0.77*	*0.74*	*0.84*	*1*	*0.959*	*0.01*	*0.05*	*0.16*	*0.17*	*0.25*	*0.53*	*0.025*
Firm Size	0.24	0.28	0.3	0.31	0.33	0.42		0.07	0.18	0.22	0.22	0.25	0.32	
	0.01	*0.03*	*0.05*	*0.05*	*0.07*	*0.13*	*0.001*	*0.01*	*0.04*	*0.07*	*0.11*	*0.16*	*0.6*	*0.007*
Leverage	-3.47	-2.99	-2.85	-2.84	-2.68	-2.27		-0.65	-0.16	-0.02	-0.05	0.09	0.34	
	0	*0*	*0*	*0.01*	*0.01*	*0.03*	*0.000*	*0.2*	*0.66*	*0.81*	*0.76*	*0.9*	*1*	*0.966*
Constant	-2.64	-1.83	-1.57	-1.62	-1.35	-0.7		-4	-3.13	-2.59	-2.63	-2.12	-0.77	
	0.16	*0.33*	*0.41*	*0.4*	*0.47*	*0.71*	*0.456*	*0.01*	*0.04*	*0.09*	*0.12*	*0.16*	*0.64*	*0.007*
Observations	96	100	100	100	101	104		97	100	101	101	102	105	
R^2	0.15	0.18	0.19	0.2	0.21	0.27		0.03	0.05	0.07	0.07	0.08	0.24	

Figure A2: Results by Numeric Information in Breach Descriptions. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for disclosures with data-light and data-heavy content based on a median split of the number of numerical values reported in the description. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values (in italic) for each sub-sample, together with Fisher–Brown test results.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for Data-light (left) and Data-heavy (right) description



Panel B: Summary Statistics and Fisher–Brown Test Results

	Data-light description							Data-heavy description						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.78	-0.46	-0.39	-0.38	-0.3	-0.11		-0.75	-0.44	-0.36	-0.36	-0.28	-0.1	
	0.05	*0.13*	*0.2*	*0.24*	*0.3*	*0.66*	*0.115*	*0*	*0.08*	*0.16*	*0.2*	*0.28*	*0.7*	*0.007*
Tobin's Q	-0.21	-0.08	-0.06	-0.05	-0.04	0.08		-0.12	-0.02	0.04	0.03	0.07	0.42	
	0.25	*0.69*	*0.74*	*0.73*	*0.8*	*0.94*	*0.959*	*0.04*	*0.79*	*0.87*	*0.83*	*0.93*	*1*	*0.988*
Dividend	-0.79	-0.6	-0.53	-0.53	-0.46	-0.17		-0.14	0.27	0.44	0.46	0.59	1.08	
	0.21	*0.35*	*0.42*	*0.42*	*0.48*	*0.81*	*0.516*	*0.06*	*0.3*	*0.45*	*0.46*	*0.65*	*0.99*	*0.494*
R&D Intensity	4.64	6.68	7.12	7.19	7.68	9.19		-27.38	-23.14	-20.02	-19.74	-16.4	-9.39	
	0.33	*0.39*	*0.42*	*0.42*	*0.44*	*0.56*	*0.541*	*0*	*0.01*	*0.03*	*0.05*	*0.07*	*0.36*	*0.000*
Intangibles	-2.8	-1.3	-0.85	-0.87	-0.47	4.45		4.18	6.92	9.02	8.85	10.14	14.82	
	0.53	*0.83*	*0.9*	*0.88*	*0.94*	*1*	*0.998*	*0.01*	*0.08*	*0.14*	*0.17*	*0.25*	*0.51*	*0.029*
Firm Size	0.17	0.2	0.21	0.22	0.23	0.28		0.11	0.2	0.24	0.24	0.28	0.36	
	0.04	*0.11*	*0.14*	*0.14*	*0.17*	*0.26*	*0.037*	*0.01*	*0.03*	*0.06*	*0.09*	*0.13*	*0.4*	*0.003*
Leverage	-0.37	-0.06	-0.02	-0.01	0.09	0.32		-2.39	-1.92	-1.76	-1.77	-1.62	-1.32	
	0.61	*0.82*	*0.92*	*0.88*	*0.96*	*0.98*	*0.998*	*0*	*0.02*	*0.03*	*0.04*	*0.05*	*0.11*	*0.000*
Constant	-2.34	-1.91	-1.78	-1.77	-1.61	-1.33		-4.2	-3.06	-2.53	-2.48	-1.9	-0.8	
	0.16	*0.26*	*0.3*	*0.3*	*0.34*	*0.43*	*0.275*	*0.02*	*0.06*	*0.13*	*0.17*	*0.26*	*0.65*	*0.030*
Observations	89	91	92	92	93	95		103	108	109	109	110	117	
R^2	0.07	0.09	0.1	0.1	0.11	0.14		0.08	0.11	0.14	0.14	0.16	0.24	

Internet Appendix for

“Measuring Data Breach Severity with Large Language Models: Evidence from Market Reactions"

IA.1. Pre-Announcement Stock Price Movements Around Cybersecurity Breaches

Table I: This table reports the distributions of coefficient estimates from regressions of pre-announcement cumulative abnormal returns, CAR(-10,-2), on the Impact Score and firm-level controls. CAR(-10,-2) measures cumulative abnormal returns from trading day -10 to trading day -2 relative to the breach announcement, estimated using the market model. The dependent variable is expressed in percentage, and the Impact Score is standardised within each run. The regressions are estimated separately for 100 stochastic Impact Scores. The table reports the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum of the coefficient estimates across runs. Heteroskedasticity-robust p-values are reported in italics beneath the corresponding estimates. The final column reports Fisher–Brown combined p-values, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.562	-0.165	-0.055	-0.074	0.033	0.287	
	0.187	*0.613*	*0.780*	*0.732*	*0.883*	*0.998*	*0.942*
Tobin's Q	0.062	0.221	0.254	0.254	0.285	0.458	
	0.101	*0.344*	*0.398*	*0.399*	*0.457*	*0.822*	*0.472*
Dividend	1.005	1.283	1.404	1.407	1.546	1.786	
	0.042	*0.072*	*0.097*	*0.106*	*0.135*	*0.256*	*0.017*
R&D Intensity	8.835	13.008	14.664	14.718	16.008	23.010	
	0.057	*0.142*	*0.184*	*0.191*	*0.232*	*0.407*	*0.088*
Intangibles	-15.416	-9.329	-7.973	-8.170	-6.792	-1.971	
	0.060	*0.184*	*0.245*	*0.259*	*0.315*	*0.719*	*0.175*
Firm Size	-0.400	-0.318	-0.298	-0.295	-0.267	-0.192	
	0.037	*0.090*	*0.113*	*0.121*	*0.150*	*0.271*	*0.025*
Leverage	-0.729	-0.294	-0.182	-0.192	-0.093	0.280	
	0.385	*0.726*	*0.825*	*0.799*	*0.885*	*0.997*	*0.984*
Constant	0.866	1.734	1.956	1.962	2.251	2.898	
	0.124	*0.231*	*0.282*	*0.298*	*0.354*	*0.613*	*0.254*
Observations	191	198	199	199	201	207	
R^2	0.029	0.039	0.045	0.045	0.049	0.064	

IA.2. Impact Score and Subsequent ROA

Table II: This table reports summary statistics for the coefficient estimates from regressions of changes in ROA on *Impact_Score* across the 100 stochastic runs. The dependent variable is ROA at $t+h$ minus ROA at $t-1$, for $h = 1, 2$. Regressions are estimated without firm-level controls. ROA is net income divided by total assets, ROE is net income divided by common equity, and sales growth is the annual change in sales scaled by lagged sales. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year t + 1

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.012	-0.007	-0.004	-0.004	-0.002	0.005	
	0.106	*0.423*	*0.567*	*0.578*	*0.749*	*0.992*	*0.743*
Observations	183	190	192	192	193	198	
R^2	0.000	0.001	0.001	0.002	0.004	0.012	

Year t + 2

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.015	-0.009	-0.007	-0.007	-0.005	0.004	
	0.054	*0.282*	*0.404*	*0.446*	*0.541*	*0.997*	*0.477*
Observations	180	187	189	189	190	195	
R^2	0.000	0.002	0.006	0.008	0.010	0.029	

IA.3. Impact Score and Subsequent ROE

Table III: This table reports summary statistics for the coefficient estimates from regressions of changes in ROE on *Impact_Score* across the 100 stochastic runs. The dependent variable is ROE at $t+h$ minus ROE at $t-1$, for $h = 1, 2$. Regressions are estimated without firm-level controls. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year t + 1

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-1.257	-1.012	-0.896	-0.880	-0.776	-0.206	
	0.071	*0.134*	*0.161*	*0.172*	*0.201*	*0.391*	*0.070*
Observations	183	190	192	192	193	198	
R^2	0.001	0.008	0.010	0.010	0.013	0.019	

Year t + 2

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	0.057	0.552	0.666	0.642	0.756	1.051	
	0.224	*0.302*	*0.348*	*0.376*	*0.417*	*0.768*	*0.418*
Observations	180	187	189	189	190	195	
R^2	0.000	0.013	0.019	0.019	0.025	0.048	

IA.4. Impact Score and Subsequent Sales Growth

Table IV: This table reports summary statistics for the coefficient estimates from regressions of changes in Sales Growth on *Impact_Score* across the 100 stochastic runs. The dependent variable is Sales Growth at $t + h$ minus Sales Growth at $t - 1$, for $h = 1, 2$. Regressions are estimated without firm-level controls. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year $t + 1$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.033	-0.015	-0.009	-0.009	-0.002	0.011	
	< 0.001	*0.429*	*0.621*	*0.601*	*0.790*	*0.995*	*0.629*
Observations	183	190	192	192	193	198	
R^2	0.000	0.000	0.001	0.002	0.003	0.012	

Year $t + 2$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.030	-0.006	0.001	0.000	0.009	0.023	
	0.052	*0.511*	*0.719*	*0.684*	*0.887*	*1.000*	*0.886*
Observations	180	187	189	189	190	195	
R^2	0.000	0.000	0.001	0.001	0.002	0.007	

IA.5. Impact Score and Subsequent ROA

Table V: This table reports summary statistics for the coefficient estimates from regressions of changes in ROA on *Impact_Score* across the 100 stochastic runs. The dependent variable is ROA at $t + h$ minus ROA at $t - 1$, for $h = 1, 2$. **All regressions include six firm-level controls** (Tobin's Q, dividend, R&D intensity, intangibles, firm size, and leverage). Only the coefficient on *Impact_Score* is reported. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year $t + 1$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.015	-0.008	-0.006	-0.006	-0.004	0.004	
	0.061	*0.271*	*0.447*	*0.460*	*0.580*	*1.000*	*0.486*
Observations	183	190	192	192	193	198	
R^2	0.044	0.055	0.060	0.060	0.064	0.079	

Year $t + 2$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.016	-0.009	-0.007	-0.007	-0.005	0.005	
	0.044	*0.261*	*0.406*	*0.428*	*0.520*	*0.990*	*0.434*
Observations	180	187	189	189	190	195	
R^2	0.027	0.044	0.048	0.049	0.054	0.076	

IA.6. Impact Score and Subsequent ROE

Table VI: This table reports summary statistics for the coefficient estimates from regressions of changes in ROE on *Impact_Score* across the 100 stochastic runs. The dependent variable is ROE at $t + h$ minus ROE at $t - 1$, for $h = 1, 2$. **All regressions include six firm-level controls** (Tobin's Q, dividend, R&D intensity, intangibles, firm size, and leverage). Only the coefficient on *Impact_Score* is reported. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year $t + 1$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-1.291	-1.061	-0.959	-0.928	-0.815	-0.293	
	0.074	*0.140*	*0.171*	*0.177*	*0.210*	*0.323*	*0.077*
Observations	183	190	192	192	193	198	
R^2	0.016	0.023	0.026	0.026	0.029	0.036	

Year $t + 2$

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.024	0.611	0.736	0.701	0.809	1.171	
	0.189	*0.264*	*0.330*	*0.353*	*0.391*	*0.819*	*0.366*
Observations	180	187	189	189	190	195	
R^2	0.034	0.058	0.066	0.067	0.074	0.103	

IA.7. Impact Score and Subsequent Sales Growth

Table VII: This table reports summary statistics for the coefficient estimates from regressions of changes in Sales Growth on *Impact_Score* across the 100 stochastic runs. The dependent variable is Sales Growth at $t + h$ minus Sales Growth at $t - 1$, for $h = 1, 2$. **All regressions include six firm-level controls** (Tobin's Q, dividend, R&D intensity, intangibles, firm size, and leverage). Only the coefficient on *Impact_Score* is reported. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Year t + 1

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.030	-0.007	-0.000	-0.000	0.007	0.022	
	< 0.001	*0.549*	*0.724*	*0.693*	*0.864*	*0.993*	*0.841*
Observations	183	190	192	192	193	198	
R^2	0.055	0.061	0.065	0.066	0.070	0.090	

Year t + 2

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.030	-0.009	-0.002	-0.002	0.006	0.024	
	0.113	*0.505*	*0.731*	*0.691*	*0.886*	*0.991*	*0.899*
Observations	180	187	189	189	190	195	
R^2	0.026	0.047	0.048	0.049	0.051	0.058	

IA.8. Impact Score and Predicted Securities Litigation Risk in the Year Following the Breach

Table VIII: This table reports the distributions of coefficient estimates from regressions of predicted securities litigation risk in the year following the breach on the Impact Score and firm-level controls. The measure is $-7.883 + 0.566\,\text{FPS} + 0.518\,\ln(\text{Assets}) + 0.982\,\text{Sales Growth} + 0.379\,\text{Stock Return} - 0.108\,\text{Return Skewness} + 25.635\,\text{Return Volatility} + 0.00007\,\text{Share Turnover}$, where FPS indicates high-litigation-risk industries (biotechnology, computers, electronics, and retail), Sales Growth is the change in annual sales scaled by beginning-of-year total assets, Stock Return is the cumulative market-adjusted return over the preceding 12 months, Return Skewness and Return Volatility are the skewness and standard deviation of monthly returns, and Share Turnover is trading volume relative to shares outstanding; except for FPS, all inputs are measured in the preceding fiscal year. The litigation risk measure follows Kim and Skinner (2012) and is constructed from firm characteristics rather than observed litigation outcomes. The regressions are estimated separately for 100 stochastic Impact Scores. All specifications include the six firm-level controls used in the baseline analysis. The table reports the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum of the coefficient estimates across runs. Heteroskedasticity-robust HC1 p-values are reported in italics beneath the corresponding estimates. The final column reports Fisher–Brown combined p-values, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.011	-0.003	-0.001	-0.002	0.000	0.007	
	0.062	*0.524*	*0.716*	*0.679*	*0.888*	*0.998*	*0.869*
Observations	179	186	188	188	189	194	
R^2	0.597	0.610	0.613	0.614	0.617	0.678	

IA.9. Impact Score and Changes in Predicted Securities Litigation Risk

Table IX: This table reports the distributions of coefficient estimates from regressions of the change in predicted securities litigation risk between the year after and the year before the breach on the Impact Score and firm-level controls. The litigation risk measure follows Kim and Skinner (2012) and is constructed from firm characteristics rather than observed litigation outcomes. The regressions are estimated separately for 100 stochastic Impact Scores. All specifications include the six firm-level controls used in the baseline analysis. The table reports the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum of the coefficient estimates across runs. Heteroskedasticity-robust HC1 p-values are reported in italics beneath the corresponding estimates. The final column reports Fisher–Brown combined p-values, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.013	-0.007	-0.005	-0.005	-0.003	0.002	
	0.025	*0.282*	*0.461*	*0.468*	*0.635*	*0.967*	*0.484*
Observations	148	155	157	157	159	164	
R^2	0.072	0.235	0.238	0.237	0.242	0.258	

IA.10. Impact Score and Future Breach Incidence Within Three Years

Table X: This table reports summary statistics for the coefficient estimates from logistic regressions of future breach incidence on *Impact_Score* and firm-level controls across the 100 stochastic runs. The dependent variable equals one if the same firm experiences another recorded data breach in one of the three subsequent calendar years, and zero otherwise. Events without complete three-year follow-up are excluded. The dependent variable is constructed using the full available breach history before applying the original baseline sample restrictions. The corresponding heteroskedasticity-robust p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs. McFadden pseudo-R^2 is reported.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-2.509	-0.218	-0.041	-0.079	0.081	0.659	
	0.033	*0.466*	*0.652*	*0.636*	*0.867*	*0.998*	*0.796*
Tobin's Q	-0.207	-0.039	0.000	0.000	0.029	0.224	
	0.436	*0.828*	*0.898*	*0.869*	*0.950*	*1.000*	*0.997*
Dividend	-0.761	-0.095	0.006	0.076	0.102	1.077	
	0.429	*0.847*	*0.924*	*0.869*	*0.967*	*0.999*	*0.996*
R&D intensity	-43.687	-37.304	-35.953	-36.059	-34.726	-30.981	
	0.005	*0.031*	*0.043*	*0.041*	*0.050*	*0.066*	*0.001*
Intangibles	8.489	10.155	10.526	10.770	11.219	13.775	
	0.011	*0.021*	*0.026*	*0.028*	*0.033*	*0.077*	*0.000*
Firm size	0.307	0.363	0.385	0.404	0.445	0.610	
	0.000	*0.026*	*0.037*	*0.041*	*0.051*	*0.149*	*0.001*
Leverage	-4.480	-3.125	-2.745	-2.822	-2.475	-2.144	
	0.018	*0.047*	*0.058*	*0.060*	*0.068*	*0.134*	*0.003*
Observations	148	154	155	155	157	160	
Pseudo-R^2	0.110	0.117	0.133	0.134	0.145	0.196	

IA.11. Impact Score and Future Breach Incidence in All Subsequent Years

Table XI: This table reports summary statistics for the coefficient estimates from logistic regressions of future breach incidence on *Impact_Score* and firm-level controls across the 100 stochastic runs. The dependent variable equals one if the same firm experiences another recorded data breach in any subsequent calendar year through 2021, and zero otherwise. Events in the final year of the observation period are excluded because no subsequent calendar year is observed. The available follow-up period may differ across events. The dependent variable is constructed using the full available breach history before applying the original baseline sample restrictions. The corresponding heteroskedasticity-robust p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs. McFadden pseudo-R^2 is reported.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.086	0.238	0.314	0.312	0.387	0.633	
	0.032	*0.154*	*0.236*	*0.266*	*0.319*	*0.885*	*0.148*
Tobin's Q	-0.110	-0.062	-0.046	-0.044	-0.021	0.013	
	0.464	*0.667*	*0.748*	*0.760*	*0.878*	*0.998*	*0.971*
Dividend	-0.524	-0.364	-0.315	-0.308	-0.258	0.052	
	0.331	*0.505*	*0.567*	*0.580*	*0.634*	*0.936*	*0.806*
R&D intensity	-15.026	-13.405	-12.883	-12.855	-12.207	-11.446	
	0.038	*0.060*	*0.067*	*0.069*	*0.078*	*0.103*	*0.006*
Intangibles	2.879	3.924	4.231	4.228	4.577	5.747	
	0.138	*0.225*	*0.258*	*0.262*	*0.293*	*0.438*	*0.209*
Firm size	0.303	0.356	0.365	0.366	0.376	0.425	
	0.001	*0.002*	*0.003*	*0.003*	*0.004*	*0.014*	*0.000*
Leverage	-1.742	-1.489	-1.423	-1.430	-1.353	-1.176	
	0.031	*0.054*	*0.064*	*0.065*	*0.077*	*0.106*	*0.005*
Observations	188	194	196	196	198	203	
Pseudo-R^2	0.092	0.102	0.108	0.109	0.115	0.132	

IA.12. Sector Effects on Breach Severity

Table XII: This table reports summary statistics for coefficient estimates from regressions of the standardised *Impact_Score* on sector indicators across 100 stochastic runs. BFS is the reference sector. The corresponding heteroskedasticity-robust p-values are displayed in italics. The joint Wald test examines whether all four sector coefficients are jointly equal to zero. Summary statistics for the Wald statistics and their p-values are also reported. The final column reports Fisher–Brown combined p-values across the 100 stochastic runs, accounting for dependence across runs. The median number of observations and R^2 are also reported.

Sector	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
BSO	-0.244	-0.126	-0.064	-0.059	-0.001	0.221	
	0.196	*0.509*	*0.700*	*0.671*	*0.857*	*0.998*	*0.935*
BSR	-0.075	0.128	0.212	0.206	0.296	0.450	
	0.019	*0.150*	*0.302*	*0.358*	*0.507*	*0.982*	*0.178*
MED	-0.257	0.028	0.091	0.080	0.141	0.355	
	0.110	*0.492*	*0.636*	*0.637*	*0.804*	*0.986*	*0.906*
UNKN	-0.432	-0.234	-0.166	-0.158	-0.079	0.125	
	0.093	*0.356*	*0.516*	*0.537*	*0.733*	*0.977*	*0.716*
Joint Wald statistic	0.507	2.739	3.729	4.049	5.072	10.345	
	0.035	*0.280*	*0.444*	*0.461*	*0.602*	*0.973*	*0.463*
Observations	201						
R^2	0.015						

IA.13. Market Reactions to Hacking Incidents

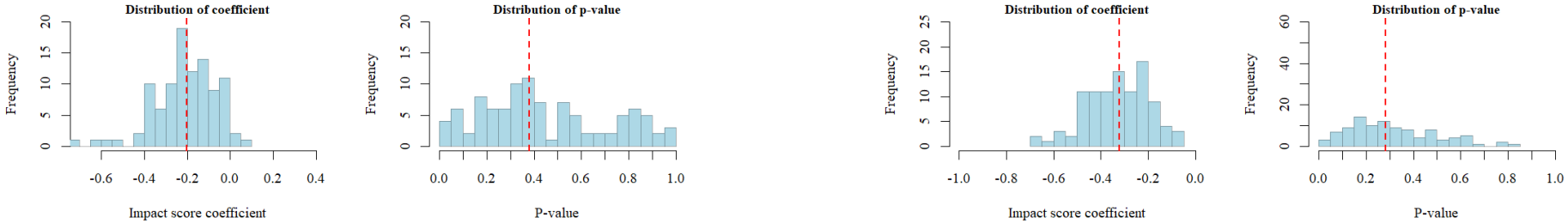
Table XIII: This table reports OLS regression results examining the relationship between cumulative abnormal returns (CARs) over the event window $(-1, +1)$ and an indicator for hacking incidents (Hack). Column (1) includes six firm-level controls: Tobin's Q, dividend, R&D intensity, intangibles, firm size, and leverage. Column (2) additionally includes three breach-description characteristics: description length, number of numerical expressions, and Flesch–Kincaid grade level. All regressions exclude contaminated observations and year fixed effects. CARs are multiplied by 100. Heteroskedasticity-robust standard errors (HC1) are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

	$CAR(-1, 1)$	
	Firm controls	Firm and description controls
	(1)	(2)
Hack	0.465	0.625
	(0.409)	(0.443)
Tobin's Q	−0.209	−0.221
	(0.153)	(0.154)
Dividend	−0.408	−0.394
	(0.426)	(0.418)
R&D Intensity	−1.333	−1.142
	(9.197)	(9.192)
Intangibles	6.615	6.008
	(4.564)	(4.670)
Firm Size	0.152	0.117
	(0.101)	(0.100)
Leverage	−0.170	−0.157
	(0.506)	(0.503)
Description Length		0.0001
		(0.0002)
Number of Numerical Expressions		−0.012
		(0.009)
Flesch–Kincaid Grade		−0.021
		(0.058)
Constant	−1.358	−0.603
	(1.199)	(1.421)
Observations	238	238
R^2	0.054	0.075
Adjusted R^2	0.026	0.035
Residual Std. Error	2.600 (df = 230)	2.588 (df = 227)
F Statistic	1.893* (df = 7; 230)	1.853* (df = 10; 227)

IA.14. Results by Prior Securities Litigation Risk

Figure I: Results by Prior Securities Litigation Risk. Panel A displays the distributions of the estimated *Impact_Score* coefficients and their associated *p*-values across 100 stochastic regressions, separately for firms with low and high prior securities litigation risk. Litigation risk is measured using the predicted probability of securities litigation, measured in the year preceding the breach. Firms are classified into low- and high-risk groups based on the sample median of predicted litigation risk. In each histogram, the red dashed line indicates the median of the distribution. Panel B reports summary statistics for the regression coefficients and corresponding *p*-values for each subsample, together with Fisher–Brown combined *p*-values accounting for dependence across stochastic runs.

Panel A: Estimated *Impact_Score* Coefficient and *p*-value distributions for low (left) and high (right) prior litigation risk sub-sample



Panel B: Summary Statistics and Fisher–Brown Test Results

	Low prior litigation risk							High prior litigation risk						
	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown *p*-value
Impact_Score	-0.74	-0.27	-0.2	-0.21	-0.11	0.07		-0.69	-0.43	-0.32	-0.33	-0.23	-0.07	
	0.01	*0.23*	*0.38*	*0.45*	*0.64*	*1*	*0.376*	*0*	*0.17*	*0.28*	*0.31*	*0.43*	*0.8*	*0.085*
Tobin's Q	-0.2	-0.08	-0.05	-0.03	0.02	0.19		-0.26	-0.16	-0.1	-0.09	-0.02	0.09	
	0.22	*0.56*	*0.72*	*0.71*	*0.89*	*1*	*0.93*	*0.26*	*0.54*	*0.73*	*0.7*	*0.91*	*1*	*0.912*
Dividend	-1.24	-0.9	-0.75	-0.76	-0.6	-0.32		0.22	0.61	0.79	0.8	0.95	1.39	
	0.02	*0.1*	*0.16*	*0.17*	*0.23*	*0.53*	*0.051*	*0.08*	*0.23*	*0.31*	*0.34*	*0.43*	*0.76*	*0.281*
R&D Intensity	-12.11	-9.24	-8.37	-8.48	-7.67	-4.77		13.81	17.6	21.92	20.78	23.91	28.44	
	0.09	*0.18*	*0.22*	*0.22*	*0.25*	*0.47*	*0.127*	*0*	*0*	*0*	*0.01*	*0.01*	*0.02*	*0*
Intangibles	-0.51	1.3	2.04	2.17	2.95	5.24		-8.43	-4.16	-2.11	-1.47	1.33	3.71	
	0.26	*0.49*	*0.64*	*0.63*	*0.76*	*1*	*0.852*	*0.38*	*0.62*	*0.72*	*0.72*	*0.81*	*0.99*	*0.951*
Firm Size	-0.62	-0.46	-0.39	-0.39	-0.3	-0.17		0.32	0.39	0.41	0.42	0.46	0.59	
	0	*0.03*	*0.08*	*0.11*	*0.17*	*0.42*	*0.004*	*0.01*	*0.04*	*0.07*	*0.07*	*0.1*	*0.16*	*0.005*
Leverage	-0.64	-0.15	0.01	-0.02	0.12	0.52		-2.74	-2.32	-2.13	-2.12	-1.91	-1.29	
	0.17	*0.68*	*0.81*	*0.76*	*0.9*	*1*	*0.962*	*0*	*0.01*	*0.01*	*0.02*	*0.03*	*0.21*	*0*
Constant	0.85	2.24	3.05	3.03	3.65	5.51		-6.87	-4.87	-4.42	-4.36	-3.64	-2.88	
	0.01	*0.07*	*0.13*	*0.17*	*0.26*	*0.67*	*0.028*	*0.03*	*0.11*	*0.14*	*0.17*	*0.24*	*0.36*	*0.052*
Observations	76	80	80	80	81	84		75	79	80	80	81	83	
R2	0.08	0.12	0.13	0.13	0.15	0.2		0.28	0.3	0.33	0.33	0.34	0.49	

IA.15. Formal Tests of Heterogeneity

Table XIV: This table reports summary statistics for the interaction coefficient on *Impact_Score* $\times D_i$ from fully interacted regressions of CAR(-1,+1) on *Impact_Score*, a group indicator D_i, their interaction, the firm-level controls, and the interactions of each control with D_i, estimated separately for each of the 100 stochastic runs. In Panel A, D_i equals one for disclosures that ChatGPT assesses as understood in all ten assessments (Fully understood), descriptions below the median length (Short description), descriptions at or above the median Flesch–Kincaid Grade Level (High FKG), descriptions at or above the median Flesch Reading Ease score (High FRE), and descriptions at or above the median number of numerical expressions (Number heavy). In Panel B, D_i equals one for hacking incidents (Hack), firms below the sample median of firm size (Small firm), financial services firms (BFS), healthcare providers (MED), firms in either financial services or healthcare (BFS or MED), or firms at or above the median of predicted securities litigation risk in the year before the breach (High litigation risk). Firm size and its interaction are omitted when D_i is based on firm size. A negative coefficient indicates a stronger negative association between *Impact_Score* and announcement returns for the group with $D_i = 1$. Heteroskedasticity-robust p-values are reported in italics. The final column reports Fisher–Brown combined p-values across the 100 stochastic runs.

D_i	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Panel A: Disclosure characteristics							
Fully understood	-3.675	-0.808	-0.473	-0.460	-0.095	1.478	
	0.001	*0.114*	*0.270*	*0.343*	*0.527*	*0.991*	*0.120*
Short description	-0.010	0.306	0.404	0.410	0.524	0.981	
	0.005	*0.142*	*0.254*	*0.309*	*0.387*	*0.980*	*0.161*
High FKG	-0.917	-0.422	-0.258	-0.264	-0.112	0.185	
	0.008	*0.206*	*0.472*	*0.478*	*0.727*	*0.974*	*0.407*
High FRE	-0.072	0.302	0.413	0.452	0.580	1.276	
	0.013	*0.131*	*0.252*	*0.301*	*0.418*	*0.971*	*0.129*
Number heavy	-0.483	-0.073	0.022	0.023	0.126	0.463	
	0.195	*0.616*	*0.796*	*0.747*	*0.911*	*0.999*	*0.954*
Panel B: Breach type and firm characteristics							
Hack	0.000	0.303	0.426	0.415	0.530	0.811	
	0.038	*0.177*	*0.318*	*0.358*	*0.519*	*0.999*	*0.266*
Small firm	-0.584	-0.268	-0.151	-0.155	-0.038	0.981	
	0.069	*0.437*	*0.639*	*0.613*	*0.839*	*0.984*	*0.774*
BFS	-3.190	-0.730	-0.565	-0.589	-0.414	0.173	
	0.000	*0.039*	*0.136*	*0.205*	*0.296*	*0.942*	*0.017*
MED	-1.125	-0.340	-0.162	-0.166	0.037	0.733	
	0.010	*0.366*	*0.548*	*0.539*	*0.740*	*0.997*	*0.561*
BFS or MED	-1.158	-0.721	-0.622	-0.606	-0.487	-0.162	
	0.000	*0.034*	*0.068*	*0.120*	*0.176*	*0.623*	*0.005*
High litigation risk	-0.518	-0.284	-0.121	-0.114	-0.014	0.435	
	0.144	*0.417*	*0.669*	*0.621*	*0.803*	*0.989*	*0.818*

IA.16. Continuous Interaction Tests

Table XV: This table reports summary statistics for the interaction coefficient on *Impact_Score* $\times X_i$ from regressions of CAR(-1,+1) on *Impact_Score*, a continuous characteristic X_i, their interaction, and the firm-level controls, estimated separately for each of the 100 stochastic runs. The marginal effect of *Impact_Score* on CAR(-1,+1) equals $\beta_1 + \beta_3 X_i$, where β_3 is the reported interaction coefficient. Firm size is omitted from the controls when it is the interacted characteristic. Description length is measured in characters, so the coefficients on its raw interaction round to zero. Heteroskedasticity-robust p-values are reported in italics. The final column reports Fisher–Brown combined p-values across the 100 stochastic runs.

X_i	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Firm size	0.040	0.102	0.131	0.129	0.154	0.242	
	0.006	*0.107*	*0.190*	*0.214*	*0.286*	*0.889*	*0.066*
Description length	-0.000	-0.000	0.000	0.000	0.000	0.000	
	0.001	*0.359*	*0.634*	*0.580*	*0.831*	*0.997*	*0.614*
log(Description length)	-0.404	-0.186	-0.129	-0.128	-0.062	0.167	
	0.019	*0.277*	*0.450*	*0.472*	*0.689*	*0.993*	*0.479*
Number of numerical expressions	-0.008	0.002	0.006	0.006	0.009	0.040	
	0.000	*0.116*	*0.310*	*0.376*	*0.629*	*0.988*	*0.087*
log(Number of numerical expressions)	-0.178	-0.004	0.035	0.034	0.081	0.184	
	0.161	*0.537*	*0.697*	*0.682*	*0.869*	*0.999*	*0.898*
Flesch–Kincaid Grade	-0.560	-0.175	-0.149	-0.154	-0.125	-0.050	
	0.001	*0.015*	*0.040*	*0.062*	*0.087*	*0.428*	*0.001*
Flesch Reading Ease	0.001	0.014	0.020	0.020	0.026	0.067	
	0.003	*0.061*	*0.142*	*0.205*	*0.321*	*0.944*	*0.033*
Times understood (out of 10)	-0.351	-0.093	-0.038	-0.048	-0.004	0.560	
	0.077	*0.351*	*0.595*	*0.575*	*0.840*	*0.996*	*0.678*

IA.17. Average Impact Score Regressions

Table XVI: This table reports OLS estimates of the relationship between CAR(-1,+1) and the average standardised *Impact_Score* across 100 stochastic ChatGPT runs. *Impact_Score* values are standardised within each run before being averaged across runs for each breach event. All columns use the same sample and firm-level controls. Column (1) reports heteroskedasticity-robust standard errors (HC1), Column (2) reports standard errors clustered at the firm level, and Column (3) reports firm-level cluster-bootstrap standard errors based on 1,000 replications. Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	Dependent variable:		
	CAR (-1,1)		
	HC1 Robust	Firm-Clustered	Firm Bootstrap
	(1)	(2)	(3)
Average Impact Score	−0.352**	−0.352**	−0.352**
	(0.174)	(0.174)	(0.175)
Tobin's Q	−0.122	−0.122	−0.122
	(0.165)	(0.166)	(0.179)
Dividend	−0.149	−0.149	−0.149
	(0.420)	(0.427)	(0.439)
R&D Intensity	0.379	0.379	0.379
	(6.986)	(6.968)	(7.543)
Intangibles	4.898	4.898	4.898
	(4.117)	(4.083)	(4.256)
Firm Size	0.215**	0.215**	0.215**
	(0.093)	(0.094)	(0.101)
Leverage	−0.642	−0.642	−0.642
	(0.634)	(0.636)	(0.715)
Constant	−1.901*	−1.901*	−1.901
	(1.136)	(1.136)	(1.213)
Observations	234	234	234
R^2	0.067	0.067	0.067
Adjusted R^2	0.038	0.038	0.038
Residual Std. Error (df = 226)	2.593	2.593	2.593
F Statistic (df = 7; 226)	2.320**	2.320**	2.320**

IA.18. Impact Score and Simple Transparent Proxies

Table XVII: The table reports summary statistics for the coefficient estimates from regressions of abnormal returns on *Impact_Score*, breach size, and firm-level controls across the 100 stochastic runs. Breach size is measured as the natural logarithm of one plus the number of records affected. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact_Score	-0.763	-0.450	-0.373	-0.375	-0.291	-0.088	
	0.000	*0.040*	*0.078*	*0.119*	*0.165*	*0.683*	*0.001*
Breach Size	-0.077	-0.059	-0.052	-0.050	-0.040	-0.020	
	0.075	*0.188*	*0.232*	*0.272*	*0.348*	*0.642*	*0.192*
Tobin's Q	-0.119	-0.053	-0.017	-0.014	0.021	0.166	
	0.256	*0.732*	*0.828*	*0.810*	*0.921*	*0.995*	*0.987*
Dividend	-0.727	-0.402	-0.290	-0.264	-0.143	0.283	
	0.197	*0.472*	*0.605*	*0.620*	*0.780*	*0.991*	*0.842*
R&D Intensity	-14.557	-10.373	-8.775	-8.989	-7.642	-5.303	
	0.049	*0.120*	*0.188*	*0.198*	*0.259*	*0.424*	*0.083*
Intangibles	4.953	6.686	7.989	7.914	8.934	12.444	
	0.007	*0.040*	*0.079*	*0.097*	*0.142*	*0.267*	*0.007*
Firm Size	0.248	0.312	0.337	0.336	0.363	0.437	
	0.000	*0.003*	*0.005*	*0.010*	*0.013*	*0.049*	*0.000*
Leverage	-0.859	-0.629	-0.562	-0.549	-0.449	-0.262	
	0.156	*0.336*	*0.384*	*0.405*	*0.483*	*0.683*	*0.487*
Constant	-4.354	-3.417	-3.064	-3.062	-2.695	-1.763	
	0.003	*0.021*	*0.039*	*0.056*	*0.079*	*0.263*	*0.001*
Observations	155	159	161	161	163	166	
R^2	0.063	0.077	0.086	0.089	0.098	0.182	

Table XVIII: The table reports summary statistics for the coefficient estimates from regressions of abnormal returns on *Impact_Score*, negative sentiment, and firm-level controls across the 100 stochastic runs. Negative sentiment is measured as the proportion of words in the breach description classified as negative using the Loughran–McDonald financial dictionary. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact_Score	-0.767	-0.424	-0.379	-0.380	-0.319	-0.199	
	0.000	*0.022*	*0.046*	*0.062*	*0.081*	*0.277*	*0.000*
Negative Sentiment	-4.262	0.001	1.119	0.829	1.760	3.683	
	0.408	*0.702*	*0.787*	*0.790*	*0.910*	*1.000*	*0.981*
Tobin's Q	-0.155	-0.110	-0.078	-0.076	-0.050	0.065	
	0.357	*0.514*	*0.653*	*0.661*	*0.761*	*0.994*	*0.894*
Dividend	-0.395	-0.112	-0.000	0.002	0.085	0.426	
	0.344	*0.698*	*0.815*	*0.790*	*0.920*	*0.999*	*0.980*
R&D Intensity	-1.548	0.786	1.476	1.401	2.134	4.374	
	0.555	*0.771*	*0.840*	*0.830*	*0.901*	*0.998*	*0.992*
Intangibles	1.741	3.288	3.981	4.068	4.760	6.854	
	0.124	*0.269*	*0.353*	*0.368*	*0.454*	*0.679*	*0.379*
Firm Size	0.154	0.207	0.234	0.230	0.257	0.292	
	0.002	*0.008*	*0.018*	*0.024*	*0.032*	*0.120*	*0.000*
Leverage	-0.926	-0.722	-0.638	-0.636	-0.541	-0.386	
	0.166	*0.285*	*0.344*	*0.351*	*0.412*	*0.555*	*0.376*
Constant	-3.200	-2.531	-2.250	-2.253	-1.952	-1.204	
	0.010	*0.039*	*0.063*	*0.082*	*0.112*	*0.350*	*0.004*
Observations	193	200	201	201	203	209	
R^2	0.059	0.071	0.080	0.082	0.090	0.153	

Table XIX: The table reports summary statistics for the coefficient estimates from regressions of abnormal returns on *Impact_Score*, disclosure length, and firm-level controls across the 100 stochastic runs. Disclosure length is measured as the natural logarithm of one plus the number of words in the breach description. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact_Score	-0.764	-0.416	-0.367	-0.363	-0.313	-0.193	
	0.000	*0.025*	*0.051*	*0.069*	*0.095*	*0.287*	*0.000*
Disclosure Length	-0.090	0.088	0.138	0.127	0.174	0.293	
	0.052	*0.249*	*0.383*	*0.434*	*0.594*	*0.974*	*0.424*
Tobin's Q	-0.152	-0.106	-0.077	-0.074	-0.047	0.072	
	0.360	*0.525*	*0.653*	*0.662*	*0.762*	*0.998*	*0.896*
Dividend	-0.390	-0.104	0.010	0.017	0.108	0.442	
	0.328	*0.705*	*0.814*	*0.790*	*0.897*	*0.997*	*0.980*
R&D Intensity	-1.416	0.786	1.489	1.389	2.079	4.160	
	0.568	*0.769*	*0.840*	*0.829*	*0.904*	*0.997*	*0.992*
Intangibles	1.981	3.449	4.229	4.280	4.936	6.904	
	0.119	*0.245*	*0.326*	*0.340*	*0.427*	*0.636*	*0.321*
Firm Size	0.158	0.211	0.236	0.234	0.262	0.301	
	0.002	*0.007*	*0.014*	*0.022*	*0.030*	*0.101*	*0.000*
Leverage	-0.932	-0.733	-0.648	-0.645	-0.547	-0.394	
	0.161	*0.277*	*0.331*	*0.339*	*0.402*	*0.548*	*0.352*
Constant	-3.776	-3.078	-2.845	-2.818	-2.543	-1.819	
	0.006	*0.020*	*0.033*	*0.040*	*0.054*	*0.164*	*0.001*
Observations	193	200	201	201	203	209	
R^2	0.061	0.074	0.082	0.084	0.092	0.153	

Table XX: The table reports summary statistics for the coefficient estimates from regressions of abnormal returns on *Impact_Score*, breach size, negative sentiment, disclosure length, and firm-level controls across the 100 stochastic runs. Breach size is measured as the natural logarithm of one plus the number of records affected. Negative sentiment is measured as the proportion of words in the breach description classified as negative using the Loughran–McDonald financial dictionary. Disclosure length is measured as the natural logarithm of one plus the number of words in the breach description. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact_Score	-0.748	-0.381	-0.300	-0.296	-0.223	0.071	
	0.000	*0.096*	*0.195*	*0.244*	*0.311*	*0.941*	*0.011*
Breach Size	-0.093	-0.070	-0.064	-0.061	-0.051	-0.030	
	0.029	*0.118*	*0.162*	*0.188*	*0.249*	*0.491*	*0.069*
Negative Sentiment	0.938	5.332	5.799	5.946	6.871	9.021	
	0.075	*0.206*	*0.273*	*0.293*	*0.342*	*0.877*	*0.214*
Disclosure Length	0.111	0.220	0.267	0.272	0.326	0.428	
	0.048	*0.150*	*0.243*	*0.260*	*0.346*	*0.618*	*0.146*
Tobin's Q	-0.117	-0.044	-0.009	-0.007	0.024	0.182	
	0.196	*0.741*	*0.830*	*0.819*	*0.929*	*0.999*	*0.989*
Dividend	-0.655	-0.321	-0.196	-0.185	-0.068	0.325	
	0.255	*0.571*	*0.706*	*0.706*	*0.819*	*0.990*	*0.934*
R&D Intensity	-13.814	-9.477	-8.008	-8.119	-6.688	-4.335	
	0.065	*0.156*	*0.234*	*0.248*	*0.323*	*0.515*	*0.151*
Intangibles	4.420	6.488	7.658	7.606	8.702	12.025	
	0.010	*0.052*	*0.095*	*0.113*	*0.164*	*0.326*	*0.012*
Firm Size	0.243	0.299	0.330	0.330	0.358	0.433	
	0.000	*0.003*	*0.007*	*0.012*	*0.016*	*0.053*	*0.000*
Leverage	-0.841	-0.636	-0.561	-0.550	-0.457	-0.279	
	0.137	*0.328*	*0.385*	*0.398*	*0.479*	*0.661*	*0.472*
Constant	-5.438	-4.691	-4.420	-4.403	-4.047	-3.206	
	0.002	*0.007*	*0.012*	*0.015*	*0.021*	*0.087*	*0.000*
Observations	155	159	161	161	163	166	
R^2	0.073	0.085	0.095	0.098	0.106	0.191	

IA.19. Impact Score and Observable Breach Characteristics

Table XXI: This table reports the distributions of regression coefficients across 100 stochastic ChatGPT runs. The dependent variable is the standardised Impact Score. Panels A–C examine the individual associations with the number of affected records, negative sentiment, and description length (the natural logarithm of one plus the number of words), respectively. Panel D includes all three characteristics jointly. The corresponding heteroskedasticity-robust p-values are reported in italics. The final column reports Fisher–Brown combined p-values across the 100 runs, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Panel A: Breach Size							
Log Records Affected	-0.013	-0.003	0.002	0.003	0.007	0.029	
	0.135	*0.562*	*0.729*	*0.690*	*0.868*	*0.998*	*0.709*
Constant	0.038	0.237	0.257	0.262	0.300	0.354	
	0.000	*0.002*	*0.010*	*0.023*	*0.024*	*0.479*	*0.003*
R^2	0.000	0.000	0.001	0.002	0.003	0.016	
Panel B: Negative Sentiment							
Negative Sentiment	-2.462	-0.340	0.320	0.305	0.990	3.727	
	0.126	*0.595*	*0.793*	*0.738*	*0.902*	*0.994*	*0.756*
Constant	-0.065	0.056	0.091	0.085	0.115	0.203	
	0.155	*0.419*	*0.546*	*0.559*	*0.682*	*0.959*	*0.551*
R^2	0.000	0.000	0.000	0.001	0.001	0.010	
Panel C: Disclosure Length							
Log Description Length	-0.466	-0.394	-0.360	-0.360	-0.333	-0.087	
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.001*	*0.000*
Constant	0.457	1.541	1.643	1.647	1.778	2.092	
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
R^2	0.004	0.111	0.132	0.130	0.150	0.203	
Panel D: Joint							
Negative Sentiment	-7.249	-5.139	-4.318	-4.288	-3.557	-1.065	
	0.001	*0.024*	*0.059*	*0.094*	*0.111*	*0.589*	*0.030*
Log Records Affected	0.004	0.014	0.020	0.020	0.025	0.040	
	0.010	*0.119*	*0.193*	*0.256*	*0.377*	*0.807*	*0.164*
Log Description Length	-0.506	-0.402	-0.381	-0.368	-0.336	-0.099	
	0.000	*0.000*	*0.000*	*0.001*	*0.001*	*0.020*	*0.000*
Constant	0.607	1.757	1.918	1.887	2.040	2.493	
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.007*	*0.000*
R^2	0.015	0.129	0.161	0.157	0.183	0.253	

IA.20. Impact Score and News Coverage

Table XXII: The table reports summary statistics for the coefficient estimates from regressions of announcement-window cumulative abnormal returns on *Impact_Score*, news coverage, and firm-level controls across the 100 stochastic runs. News coverage is an indicator equal to one if the event has at least one recorded news appearance. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.768	-0.429	-0.385	-0.384	-0.331	-0.209	
	0.000	*0.019*	*0.040*	*0.056*	*0.076*	*0.245*	*0.000*
News Coverage	-0.409	-0.226	-0.129	-0.120	-0.008	0.136	
	0.424	*0.661*	*0.794*	*0.792*	*0.930*	*0.999*	*0.981*
Tobin's Q	-0.154	-0.109	-0.078	-0.075	-0.048	0.065	
	0.358	*0.516*	*0.655*	*0.663*	*0.764*	*0.999*	*0.897*
Dividend	-0.393	-0.120	-0.006	-0.005	0.085	0.412	
	0.362	*0.706*	*0.815*	*0.792*	*0.915*	*1.000*	*0.981*
R&D Intensity	-1.496	0.776	1.505	1.371	2.126	4.216	
	0.569	*0.777*	*0.844*	*0.832*	*0.906*	*1.000*	*0.993*
Intangibles	1.666	3.305	4.101	4.140	4.817	6.878	
	0.123	*0.262*	*0.348*	*0.361*	*0.450*	*0.694*	*0.363*
Firm Size	0.160	0.214	0.239	0.237	0.260	0.295	
	0.004	*0.012*	*0.024*	*0.031*	*0.042*	*0.122*	*0.000*
Leverage	-0.919	-0.722	-0.633	-0.635	-0.545	-0.382	
	0.168	*0.289*	*0.346*	*0.353*	*0.414*	*0.557*	*0.380*
Constant	-3.192	-2.524	-2.313	-2.271	-1.965	-1.308	
	0.010	*0.039*	*0.057*	*0.078*	*0.112*	*0.301*	*0.004*
Observations	193	200	201	201	203	209	
R^2	0.060	0.071	0.080	0.082	0.090	0.153	

Table XXIII: This table reports descriptive statistics for the average standardised Impact Score across 100 stochastic ChatGPT runs, grouped by whether the breach received news coverage.

News Coverage	N	Mean	SD	Median
No Coverage	212	-0.074	1.060	0.356
At Least One Appearance	22	0.024	0.646	0.202
Total	234			

IA.21. Correlations with Existing Cyber-Risk Measures

Table XXIV: The table reports summary statistics for the Spearman correlation coefficients between *Impact_Score* and each alternative cybersecurity risk measure across the 100 stochastic runs. Correlations use events for which both measures are available (about 45% of the sample for Florackis et al. (2023) and 77% for Jamilov et al. (2026)). CR Exposure, CR Presentation, CR Q&A, CR Executive, and CR Participants are the Jamilov et al. (2026) cyber-risk measures based on the whole earnings call, the presentation, the Q&A, executives' remarks, and other participants' remarks, respectively. The corresponding summary statistics for the p-values testing the null hypothesis of zero correlation are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Summary Statistics and Fisher–Brown Test Results

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Florackis	-0.35	-0.28	-0.24	-0.24	-0.2	-0.12	
	0	*0*	*0.01*	*0.03*	*0.04*	*0.19*	*0.000*
CR Exposure	-0.2	-0.11	-0.09	-0.09	-0.07	0	
	0.01	*0.12*	*0.21*	*0.26*	*0.36*	*1*	*0.067*
CR Presentation	-0.16	-0.11	-0.09	-0.08	-0.06	0.01	
	0.02	*0.15*	*0.22*	*0.29*	*0.39*	*0.95*	*0.119*
CR Q&A	-0.18	-0.09	-0.06	-0.07	-0.04	0.04	
	0.01	*0.24*	*0.38*	*0.41*	*0.56*	*0.95*	*0.349*
CR Executive	-0.2	-0.11	-0.1	-0.09	-0.07	0	
	0.01	*0.12*	*0.19*	*0.25*	*0.33*	*0.96*	*0.061*
CR Participants	-0.19	-0.11	-0.08	-0.09	-0.06	0.01	
	0.01	*0.12*	*0.26*	*0.3*	*0.41*	*0.99*	*0.106*

Table XXV: The table reports summary statistics for the Pearson correlation coefficients between *Impact_Score* and each alternative cybersecurity risk measure across the 100 stochastic runs. The corresponding summary statistics for the p-values testing the null hypothesis of zero correlation are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs.

Summary Statistics and Fisher–Brown Test Results

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Florackis	-0.3	-0.24	-0.21	-0.2	-0.17	-0.09	
	0	*0.01*	*0.03*	*0.05*	*0.07*	*0.35*	*0.000*
CR Exposure	-0.2	-0.1	-0.07	-0.08	-0.05	0	
	0.01	*0.17*	*0.31*	*0.34*	*0.46*	*1*	*0.165*
CR Presentation	-0.14	-0.07	-0.04	-0.04	-0.02	0.04	
	0.05	*0.34*	*0.6*	*0.57*	*0.78*	*1*	*0.736*
CR Q&A	-0.21	-0.12	-0.09	-0.09	-0.06	0	
	0	*0.1*	*0.22*	*0.27*	*0.39*	*0.96*	*0.064*
CR Executive	-0.19	-0.1	-0.07	-0.07	-0.05	0	
	0.01	*0.18*	*0.34*	*0.36*	*0.48*	*0.98*	*0.210*
CR Participants	-0.22	-0.12	-0.09	-0.1	-0.07	0	
	0	*0.1*	*0.22*	*0.24*	*0.31*	*0.96*	*0.042*

IA.22. Breach Severity and Firm Vulnerability Characteristics

Table XXVI: The table reports summary statistics for the coefficient estimates from linear regressions of the standardised *Impact_Score* on firm-level vulnerability characteristics across the 100 stochastic runs. The explanatory variables include firm size, R&D intensity, intangibles, leverage, and an indicator for the first recorded breach. The corresponding heteroskedasticity-robust p-values are displayed in italics. The final column reports the Fisher–Brown combined p-value across the 100 stochastic runs. The number of observations and R^2 are also reported.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Firm size	-0.011	0.071	0.080	0.084	0.097	0.138	
	0.001	*0.011*	*0.031*	*0.051*	*0.064*	*0.768*	*0.000*
R&D intensity	-5.766	-4.157	-3.623	-3.547	-2.933	0.590	
	0.002	*0.032*	*0.061*	*0.098*	*0.117*	*0.675*	*0.001*
Intangibles	-1.135	2.280	2.759	2.698	3.184	4.634	
	0.000	*0.020*	*0.046*	*0.089*	*0.115*	*0.824*	*0.000*
Leverage	0.052	0.170	0.251	0.250	0.312	0.525	
	0.002	*0.075*	*0.175*	*0.227*	*0.380*	*0.790*	*0.022*
First breach	0.140	0.362	0.448	0.447	0.531	0.739	
	0.002	*0.019*	*0.041*	*0.063*	*0.080*	*0.480*	*0.000*
Observations	193.000	199.750	201.000	201.290	203.000	209.000	
R^2	0.010	0.048	0.061	0.061	0.074	0.103	

IA.23. Industry and Time Patterns in Breach Severity

Table XXVII: Panel A reports the median and mean of sector-level average standardised Impact_Scores across the 100 stochastic runs, together with the median number of observations (N). Panel B reports summary statistics for the F-statistics from one-way ANOVA tests of equality of mean Impact Scores across the five sectors over the 100 runs. The corresponding summary statistics for the p-values are displayed in italics. The final column reports the p-value from the Fisher–Brown test of the null hypothesis that the p-values follow a uniform distribution.

Panel A: Average Impact_Score by Sector

Sector	Median	Mean	N
BFS	0.015	0.016	51
BSO	-0.039	-0.043	59
BSR	0.223	0.221	28
MED	0.086	0.095	40
UNKN	-0.137	-0.142	24

Panel B: One-way ANOVA

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
F-statistic	0.132	0.564	0.781	0.786	0.992	1.681	
	0.156	*0.413*	*0.538*	*0.559*	*0.689*	*0.971*	*0.788*

Table XXVIII: Panel A reports the median and mean of annual average standardised Impact Scores across the 100 stochastic runs, together with the median number of observations (N). Panel B presents the corresponding statistics for four-year periods, with the final period covering 2017–2021; because N values are medians across runs, period values need not equal the sum of annual values. Panel C reports summary statistics for the estimated coefficients from regressions of the standardised Impact Score on a linear calendar-year trend across the 100 runs. The corresponding summary statistics for the p-values testing the trend coefficient against zero, calculated using robust standard errors, are displayed in italics. The final column reports the p-value from the Fisher–Brown test of the null hypothesis that the p-values follow a uniform distribution.

Panel A: Annual Average Impact Score

Year	N	Median	Mean
2005	8	0.634	0.622
2006	23	0.424	0.423
2007	23	0.452	0.441
2008	7	0.623	0.613
2009	2	0.616	0.627
2010	13	-0.000	0.010
2011	10	-0.083	-0.108
2012	15	0.351	0.363
2013	12	0.266	0.260
2014	17	0.186	0.193
2015	7	-0.144	-0.124
2016	6	-0.034	-0.024
2017	7	-0.225	-0.238
2018	5	-1.215	-1.240
2019	32	-0.715	-0.717
2020	9	-0.246	-0.258
2021	5	-0.325	-0.336

Panel B: Average Impact Score by Period

Period	N	Median	Mean
2005–2008	61	0.485	0.477
2009–2012	41	0.150	0.145
2013–2016	42	0.125	0.129
2017–2021	58	-0.607	-0.606

Panel C: Linear Time-Trend Regression

	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Year trend	-0.098	-0.083	-0.080	-0.080	-0.077	-0.025	
	0.000	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
Observations	201						
R^2	0.164						

IA.24. Post-Announcement Stock Price Drift

Table XXIX: This table reports the distributions of coefficient estimates from regressions of post-announcement cumulative abnormal returns, CAR(2,20), on the Impact Score and firm-level controls. CAR(2,20) measures cumulative abnormal returns from trading day 2 to trading day 20 relative to the breach announcement, estimated using the market model. The dependent variable is expressed in percentage, and the Impact Score is standardised within each run. The regressions are estimated separately for 100 stochastic Impact Scores. The table reports the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum of the coefficient estimates across runs. Heteroskedasticity-robust p-values are reported in italics beneath the corresponding estimates. The final column reports Fisher–Brown combined p-values, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	0.261	0.996	1.240	1.219	1.424	2.040	
	0.000	*0.098*	*0.143*	*0.173*	*0.196*	*0.692*	*0.034*
Tobin's Q	0.661	0.882	0.932	0.943	1.002	1.326	
	0.059	*0.150*	*0.173*	*0.183*	*0.211*	*0.347*	*0.081*
Dividend	-0.727	-0.099	0.127	0.113	0.343	1.017	
	0.509	*0.786*	*0.880*	*0.855*	*0.938*	*0.995*	*0.996*
R&D Intensity	-51.986	-33.673	-30.597	-31.301	-27.937	-23.459	
	0.221	*0.492*	*0.527*	*0.512*	*0.549*	*0.628*	*0.711*
Intangibles	1.099	5.296	7.384	7.473	9.589	18.826	
	0.336	*0.658*	*0.731*	*0.725*	*0.795*	*0.957*	*0.953*
Firm Size	0.079	0.244	0.317	0.314	0.374	0.666	
	0.106	*0.401*	*0.473*	*0.489*	*0.590*	*0.862*	*0.633*
Leverage	-3.039	-2.354	-2.180	-2.185	-1.994	-1.354	
	0.115	*0.213*	*0.243*	*0.255*	*0.287*	*0.474*	*0.187*
Constant	-7.720	-4.352	-3.690	-3.604	-2.793	-1.026	
	0.098	*0.402*	*0.478*	*0.496*	*0.596*	*0.844*	*0.645*
Observations	191	198	199	199	201	207	
R^2	0.054	0.064	0.071	0.073	0.078	0.128	

Table XXX: This table reports the distributions of coefficient estimates from regressions of post-announcement cumulative abnormal returns, CAR(2,20) using Fama-French 3-factor model, on the Impact Score and firm-level controls. CAR(2,20) measures cumulative abnormal returns from trading day 2 to trading day 20 relative to the breach announcement, estimated using the Fama-French 3-factor model. The dependent variable is expressed in percentage, and the Impact Score is standardised within each run. The regressions are estimated separately for 100 stochastic Impact Scores. The table reports the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum of the coefficient estimates across runs. Heteroskedasticity-robust p-values are reported in italics beneath the corresponding estimates. The final column reports Fisher–Brown combined p-values, accounting for dependence across runs.

Variable	Min	Q1	Median	Mean	Q3	Max	Fisher–Brown p-value
Impact Score	-0.27	0.44	0.63	0.65	0.86	1.43	
	0	*0.27*	*0.4*	*0.43*	*0.54*	*0.98*	*0.334*
Tobin's Q	-0.17	0.02	0.1	0.09	0.16	0.39	
	0.53	*0.8*	*0.87*	*0.86*	*0.94*	*1*	*0.996*
Dividend	-3.3	-2.55	-2.27	-2.31	-2.05	-1.39	
	0.04	*0.1*	*0.15*	*0.15*	*0.19*	*0.35*	*0.043*
R&D Intensity	-39.65	-19.14	-16.15	-16.83	-13.67	-7.48	
	0.23	*0.64*	*0.68*	*0.68*	*0.74*	*0.85*	*0.918*
Intangibles	1.53	7.71	10.02	10.05	12.07	23.75	
	0.16	*0.53*	*0.6*	*0.6*	*0.68*	*0.93*	*0.828*
Firm Size	0.26	0.49	0.61	0.6	0.68	0.97	
	0.01	*0.12*	*0.16*	*0.19*	*0.26*	*0.55*	*0.063*
Leverage	-1.57	-0.52	-0.27	-0.32	-0.05	0.43	
	0.47	*0.8*	*0.88*	*0.86*	*0.95*	*1*	*0.996*
Constant	-8	-4.54	-3.94	-3.83	-2.67	-0.35	
	0.06	*0.38*	*0.45*	*0.48*	*0.61*	*0.95*	*0.577*
Observations	191	197.75	199	199.33	201	207	
R^2	0.014	0.021	0.025	0.026	0.029	0.067	