The Social Risks of Generative AI*

Marco Ceccarelli[†] Rex Wang Renjie[‡]

February 2026

Abstract

This paper shows that the equity market prices the novel social risks associated with generative AI. We exploit the release of ChatGPT as an information shock that updated investor beliefs about AI-related tail risks. Using pre-event ESG scores as proxies for firms' social risk management, we show that, controlling for AI productivity exposure, low-ESG firms underperform high-ESG firms by 4 percentage points over the two weeks following the release. The effect is concentrated in the Social pillar, particularly data privacy and security. Increases in option-implied downside risk indicate that changes in discount rates are the channel. A low-minus-high ESG portfolio of AI-exposed firms earns significant alphas during 2023–2024, in line with investors demanding compensation for bearing AI-related downside risks. Our findings are consistent with a tail-risk model in which social risks of AI are priced through discount rates.

JEL Classification: G11, G12, G14, G32, G34
Keywords: Artificial Intelligence, Generative AI, ChatGPT, Social Risks, Tail Risks

---

*We thank Albert Menkveld, Emilia Garcia-Appendini, Po-Hsuan Hsu, Sasan Mansouri, Sascha Steffen, and seminar and conference participants at the Applied Machine Learning Days (AMLD, Lausanne), Chinese University of Hong Kong, Shenzhen (CUHK-SZ), School of Management, Fudan University, National Taiwan University (NTU), National Cheng Kung University (NCKU), National Chengchi University (NCCU), National Taipei University (NTPU), National Tsing Hua University (NTHU), National Yang Ming Chiao Tung University (NYCU), ShanghaiTech University, University of Groningen, and Vrije Universiteit (VU) Amsterdam, for insightful conversations and comments.

[†]Vrije Universiteit (VU) Amsterdam. Email: `m.ceccarelli@vu.nl`.
[‡]Vrije Universiteit (VU) Amsterdam and Tinbergen Institute. Email: `renjie-rex.wang@vu.nl`.

1 Introduction

The recent uptake in generative artificial intelligence (AI), especially the rapid advancement in large language models (LLMs), represents a major technological shock to firms. Existing work has mainly focused on the *upside* of AI adoption, documenting effects on productivity, firm growth, labor reallocation, and systematic risk. By contrast, much less is known about how capital markets price the *downside risks* associated with the deployment of AI technologies.[1] These risks arise from a growing set of AI-related social and operational failures, including large-scale data breaches, algorithmic discrimination, AI-generated misinformation, and product liability claims involving autonomous systems. Such failures have become increasingly prominent: recent examples include regulatory fines for algorithmic discrimination in automated lending and hiring, litigation over AI-driven insurance claim denials, and large-scale data breaches linked to AI-powered systems, each generating substantial legal, regulatory, and reputational costs for the firms involved (e.g., Guenzel et al., 2025).

In this paper, we study whether and how equity markets price AI-related social risks. To guide our empirical analysis, we develop a simple, parsimonious tail-risk model in which firms differ in their adoption of AI technologies and in their social risk management quality. Greater AI adoption increases firms' productivity while also raising their exposure to discrete adverse events, while social risk management determines both the probability and severity of such events. The model yields clear, testable predictions about

[1] Notably, Acemoglu (2025) discusses the potential macro-economic implications of welfare-reducing AI innovations like advances in IT warfare and manipulation through social media and deepfakes. Babina et al. (2025) show that firms' investments in AI lead to an increase in firm-level systematic risk, consistent with these firms becoming more "growth-like." Eisfeldt and Schubert (2024) provides a review of the growing finance literature studying AI.

how firm values should respond to information shocks that shift investors' assessment of AI-related tail risk. We test the model's predictions empirically in several steps.

We start with an event-study framework centered on the release of ChatGPT on November 30, 2022, which is a salient information shock that increased investors' awareness of both the productivity potential and the social risks of generative AI (Eisfeldt et al., 2025). This event provides a sharp setting to assess whether equity markets reprice firms differentially when AI-related tail risks become salient. To capture cross-sectional variation in social-risk management, we use ESG ratings measured prior to the event. These scores summarize firms' exposure to, and management of, operational and regulatory risks central to AI deployment and are predetermined with respect to the ChatGPT shock.[2]

Figure 1: CARs Around the Release of ChatGPT, by ESG Score Quintile

This figure shows the average announcement CAR$[-1,+10]$ around the launch of ChatGPT on November 30, 2022, sorted by quintiles of MSCI IVA ESG scores available by October 31, 2022. The red bars represent the 95% confidence intervals for each quintile.

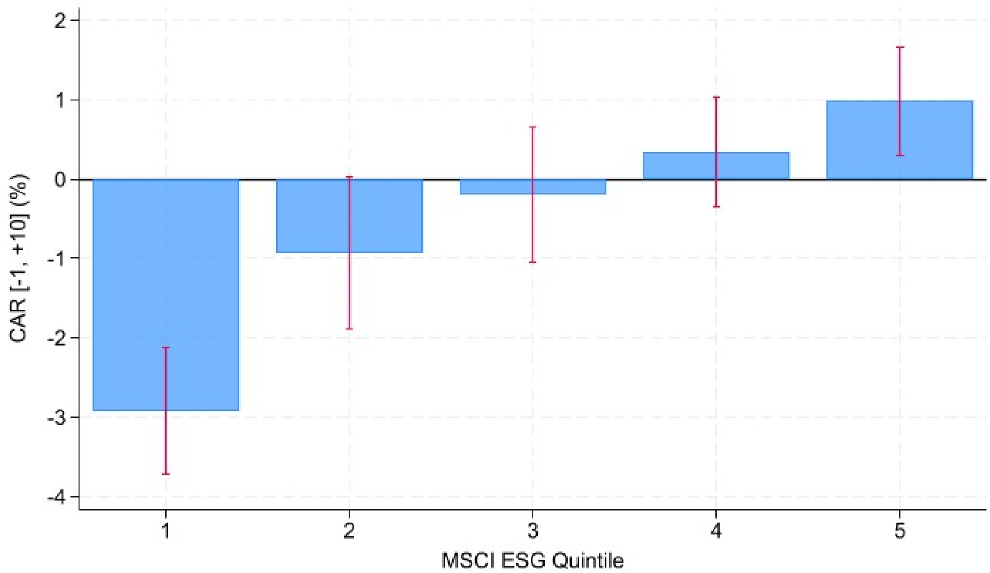


Figure 1 illustrates the resulting cross-sectional return patterns. Stocks in the bottom

[2]Using a pre-existing, third-party measure also mitigates concerns about researcher discretion and data mining. For a related discussion, see, e.g., Engle et al. 2020.

ESG quintile experience cumulative abnormal returns (CARs) of -2.9 percentage points (pp) in the two weeks following the release, while stocks in the top quintile exhibit significantly positive CARs of about 1.0pp over the same window.

To test this relationship more formally, we merge CRSP's set of publicly listed U.S. firms with Compustat fundamentals and ESG scores from several rating providers. Our primary measure of social risk management is the MSCI IVA ESG score, which is widely used by institutional investors and shown to be economically informative (Berg, Heeb, and Kölbel, 2024). We confirm the robustness of our findings with ESG ratings from Standard & Poor's (S&P) and Sustainalytics. To ensure that our results are driven by AI adoption, rather than AI development, we exclude technology firms.

To disentangle AI-related downside risk from expected AI-driven productivity gains, we explicitly control for firms' AI exposure using several complementary proxies. These include changes in the share of AI-skilled workers (Babina et al., 2024), as well as task-based workforce exposure to AI and firm-level data assets (Eisfeldt et al., 2025). Together, these measures capture both realized AI investments and ex ante AI readiness. This empirical design allows us to isolate repricing attributable to AI-related social risk, rather than differential growth expectations, and to directly test the model-implied interactions between AI adoption and social-risk management.

Our regression analysis confirms a strong, positive relationship between ESG scores and CAR[-1, +10] after controlling for various firm characteristics and industry fixed effects. This result is both statistically and economically significant: a one-standard-deviation increase in ESG scores relates to over 1pp higher CARs. Notably, the effect is concentrated among firms with the weakest ESG performance. Firms in the bottom ESG quintile experience, on average, 2.3pp more negative returns than those in the middle quintile, whereas differences across higher quintiles are economically small and statistically insignificant. This asymmetry aligns with the convex jump probability function in our model: tail-risk exposure is convex in poor governance, meaning that the marginal valuation effect of social-risk management is largest for poorly managed firms.

The relationship between ESG scores and CARs is also substantially stronger for firms with greater AI exposure, suggesting that social-risk management and AI adoption are complementary. Once this interaction is included, the standalone ESG coefficient becomes insignificant, indicating that ESG matters primarily through its interaction with AI exposure rather than as a generic firm-quality characteristic.

Not all dimensions of a firm's ESG rating are equally relevant. When we study the Environmental, Social, and Governance pillars separately, the complementarity with AI exposure is concentrated in the Social pillar. Within this pillar, sub-scores such as *'Privacy data security and management'* exhibit the strongest association with returns. These findings support the interpretation that investors price in AI-related tail risks.

Our theoretical framework implies that the repricing we document could be driven by changes in firms' discount rates, specifically by tail risk premia, as well as changes in cash flow expectations. The relative importance of these two channels is an empirical question, which we test next. For the tail-risk channel, we study changes in two option-based measures around the release of ChatGPT. First, following Bakshi, Kapadia, and Madan (2003), we examine realized skewness, which captures shifts in the asymmetry of the return distribution toward the left tail. Second, following Kelly, Pástor, and Veronesi (2016), we analyze the implied volatility smirk slope, $SlopeD$, which measures the relative expensiveness of deep out-of-the-money put options and thus isolates the market's pricing of extreme downside risk. Consistent with our model, firms with worse social risk management experience a larger decline in realized skewness and a larger increase in *SlopeD* around the event. Effectively, once AI-related risks become salient, investors perceive these firms as more exposed to left-tail outcomes. Importantly, both effects are driven entirely by the Social pillar, reinforcing the interpretation that investors are repricing AI-related social risks rather than generic firm traits or overall quality.

We now turn to the cash flow channel and study changes in analyst earnings forecasts at the one-year, five-year, and long-term horizons. If ESG scores were to primarily capture differences in AI-driven productivity potential, we would expect analysts to revise earnings expectations differentially for high- versus low-ESG firms around the event. We do not see this occurring at any horizon, suggesting that the observed repricing is unlikely to reflect changes in expected cash flows. This further strengthens our interpretation that the valuation effects operate through discount rates, specifically tail-risk premia.

In the absence of shocks to the probability or severity of tail risk events, the realized returns of firms that are exposed to these shocks should be higher, as investors will require compensation for bearing these risks. We test this conjecture in a portfolio-based asset pricing exercise. Using only information available prior to the release of ChatGPT, we form two long-short portfolios that go long bottom-quintile ESG firms and short top-quintile ESG firms (i.e., a 'low-minus-high' ESG portfolio), separately among firms with high and low AI exposure. These portfolios will capture firms that have *both* a high exposure to and a low management of AI tail risks.

We find that, between January 2023 and December 2024, the high-AI-exposure portfolio generates cumulative returns of around 55% and statistically significant daily alphas across various risk factor models. The same ESG-based sort among low-AI-exposure firms yields no significant alpha. This pattern reinforces our main insight: the return spread is specific to the interaction of weak social-risk management with AI adoption. This is consistent with investors demanding ongoing compensation for holding stocks with higher exposure to AI-related social tail risk.

Finally, as an out-of-sample test, we exploit the unexpected ousting of OpenAI's CEO, Sam Altman, on November 17, 2023. Unlike the ChatGPT launch, this event did not convey new information about AI productivity but instead heightened concerns about governance and safety in AI deployment. Contemporary reporting attributed the decision to disagreements over the handling of AI security and safety risks (Reuters, 2023a; WSJ, 2025). Consistent with our model, low-ESG firms again underperformed high-ESG firms around this second shock, with effects concentrated among firms with high AI exposure.

Our paper adds to the growing literature documenting the impact of (generative) AI as a major technology shock to firm value and corporate policies (see, e.g., Eisfeldt and Schubert, 2024, for a comprehensive review). Earlier theoretical work examines how technological advances affect firm investments and valuations (e.g., Kogan and Papanikolaou, 2014). Babina et al. (2024) measure firm-level AI investments using employee resumes and find that AI-investing firms experience higher growth in sales, employment, and market valuations. Babina, Fedyk, He, and Hodson (2023) document that AI adoption is associated with significant reorganization of firms' workforces. Eisfeldt et al. (2025) build on the work of Eloundou, Manning, Mishkin, and Rock (2024) to measure firms' workforce exposure to Generative AI and document that high-exposure firms outperformed low-exposure firms significantly following the release of ChatGPT. Notably, Babina et al. (2025) shows that the systemic risk of firms with larger AI investments increases, consistent with these firms becoming more "growth-like". Our paper adds by studying the *downside* risks of AI adoption. We show that investments in corporate social responsibility play an important role in reducing firms' exposure to AI-related tail risks.

We also contribute to the broader literature on the externalities of AI adoption. On the positive side, studies have shown how AI can complement human skills (Cao, Jiang, Wang, and Yang, 2024), reduce gender inequality (Bao, Huang, and Lin, 2024), and help forecast firm outcomes like capital investments (Jha, Qian, Weber, and Yang, 2024). However, AI also brings potential negative externalities (Sunstein, 2024; Guenzel et al., 2025). We add to this literature by showing that investors treat the risks implied by AI adoption as value-relevant and price them in through a tail-risk channel.

Our paper also contributes to the literature that proxies for exposure to risk factors through ESG performance (see, e.g., Gillan, Koch, and Starks (2021) for a review). We contribute by drawing attention to an increasingly important—but relatively underexplored—type of risks, namely those related to the widespread adoption of AI, and show that ESG ratings convey useful information to manage (and potentially hedge) exposure to this new class of risks. The tail-risk framework provides a sharper characterization compared to prior work, suggesting that ESG reduces overall volatility. We show it is specifically *tail risk*—the risk of discrete adverse events—that investments in ESG performance help manage.

2 A Tail-Risk Model of AI Adoption

This section develops a model of how AI related social risks affect firm value. Three observations motivate our modeling choice. First, AI related harms tend to arrive as discrete adverse events rather than as gradual changes in cash flow volatility. Examples include data breaches, lawsuits over algorithmic discrimination, AI generated misinformation, and product liability cases involving autonomous systems. When such events occur, they often trigger large and abrupt stock price responses, suggesting that markets view them as jump like shocks. Second, these risks are inherently one sided. There is no natural analogue to a positive data breach or a beneficial AI failure. As a result, variance alone is not sufficient to characterize exposure. Instead, AI related social risks primarily operate through tail risk and negative skewness. Third, firm level governance shapes both the likelihood and the severity of AI related incidents. Strong oversight reduces the probability of adverse events, while effective internal controls and crisis management mitigate losses conditional on an incident. A jump risk framework captures these features in a parsimonious way, allowing for discrete adverse events, asymmetric downside risk, and heterogeneity in risk management quality, while offering tractable pricing predictions.

2.1 Firms and Cash Flows

Consider a continuum of firms indexed by $i \in [0, 1]$. Each firm is characterized by two attributes: AI adoption intensity $z_i \in [0, 1]$ and social-risk management quality $s_i \in [0, 1]$. The latter captures capabilities in data privacy, product safety, and human capital management. In the absence of adverse events, firm i generates expected cash flow $\bar{\pi}_i = \bar{\pi}_0(1 + \varphi z_i)$, where $\varphi > 0$ governs the productivity gains from AI adoption.

2.2 Tail-Risk Structure

AI adoption exposes firms to discrete adverse events. Realized cash flow is:

$$\pi_i = \bar{\pi}_i + J_i, \tag{1}$$

where J_i is a jump component: $J_i = -\xi_i$ with probability p_i, and 0 otherwise.

Assumption 1 (Jump Probability)**.**

$$p_i = p_0 + \eta \cdot z_i \cdot (1 - s_i)^2, \quad \eta > 0. \tag{2}$$

This specification captures three features: (i) jump probability is increasing in AI adoption z_i—firms deploying more AI face greater exposure to AI-related incidents; (ii) jump probability is convex in poor governance $(1 - s_i)^2$; (iii) the interaction between z_i and $(1 - s_i)^2$ implies that social-risk management matters more for AI-intensive firms. We adopt a convex specification to capture the intuition that marginal improvements in governance have larger effects for poorly-managed firms: moving from very poor to moderate governance eliminates the most egregious vulnerabilities, while moving from good to excellent governance yields smaller incremental gains. The quadratic form is parsimonious. The parameter η governs the sensitivity of jump probability to AI adoption for poorly-governed firms. This could be driven, for example, by policymakers' regulatory and enforcement actions.

While we have a linear multiplier in z_i for parsimony, the model's qualitative predictions are invariant to more general specifications. For example, if the probability is modeled as $p_i = p_0 + \eta \cdot z_i^\gamma \cdot (1 - s_i)^2$ with $\gamma \geq 1$, reflecting potential network effects or complexity-driven risks in AI deployment, the directional predictions for the cross-partial $\partial^2 p/\partial z \partial s$ are preserved and the predicted return sensitivity below is only strengthened.

Assumption 2 (Jump Magnitude)**.**

$$\mathbb{E}[\xi_i \mid \textit{jump}] = \xi_0 + \nu(1 - s_i), \quad \nu > 0. \tag{3}$$

Firms with weaker social-risk management suffer larger losses when adverse events occur, reflecting both direct costs (fines, remediation) and indirect costs (reputational damage, legal liability). The parameter ν governs the sensitivity of loss magnitude to social risk management quality. For parsimony, we model the expected loss magnitude as depending on s_i but not on z_i. A natural extension would include a term κz_i with $\kappa > 0$ in Equation (3), so that firms deploying AI at greater scale also face larger losses conditional on an incident. Under this generalization, both the expected tail loss $p_i \mathbb{E}[\xi_i]$ and the tail risk premium λ^τ become more sensitive to z_i for low-s_i firms, which strengthens all comparative statics derived below without altering their sign. We adopt the simpler specification because the key cross-sectional variation in our empirical setting comes from s_i—how well firms prevent and respond to incidents—rather than from scale of deployment z_i.

Expected cash flow is therefore: $\mathbb{E}[\pi_i] = \bar{\pi}_i - p_i \cdot \mathbb{E}[\xi_i]$, where the second term captures the expected loss from AI-related tail events. The cash flow distribution for these firms features a longer left tail—most of the time the firm earns $\bar{\pi}_i$, but with probability p_i it suffers a large downside loss.

2.3 Equilibrium Pricing

In the cross-section, firms differ only in their jump-risk exposure; the baseline volatility σ_0 is common. The jump component contributes to both the variance and skewness of firm i's cash flows:

$$\mathrm{Var}(J_i) = p_i \cdot (\mathbb{E}[\xi_i])^2 \cdot (1 - p_i) \approx p_i \cdot (\mathbb{E}[\xi_i])^2, \tag{4}$$

$$\mathrm{Skew}(J_i) \approx -p_i \cdot (\mathbb{E}[\xi_i])^3, \tag{5}$$

where the approximations use $p_i \ll 1$ (rare tail-risk events). Because jumps are one-sided, higher jump variance mechanically implies more negative skewness. Both p_i and $\mathbb{E}[\xi_i]$ are decreasing in s_i, so $p_i \cdot (\mathbb{E}[\xi_i])^2$ and $p_i \cdot (\mathbb{E}[\xi_i])^3$ induce the same cross-sectional ordering. The distinction between the second and third powers becomes empirically relevant in Prediction 4 below, where option-implied skewness scales with $(\mathbb{E}[\xi_i])^3$.

Following the mean-variance-skewness preference framework (e.g., Kraus and Litzenberger, 1976; Harvey and Siddique, 2000), a representative investor maximizes:

$$\max_{w_i} \ \mathbb{E}[R_p] - \frac{\gamma}{2}\mathrm{Var}(R_p) + \frac{\delta}{6}\mathrm{Skew}(R_p), \quad \gamma > 0, \ \delta > 0, \tag{6}$$

where w_i denotes portfolio weights, γ is variance aversion, and δ captures preference for positive skewness (equivalently, aversion to negative skewness). Consider firm i's contribution to the investor's portfolio. The firm-specific (idiosyncratic) component of return variance and skewness is driven entirely by the jump:

$$\frac{\partial \mathrm{Var}(R_p)}{\partial w_i} \propto p_i \cdot (\mathbb{E}[\xi_i])^2, \qquad \frac{\partial \mathrm{Skew}(R_p)}{\partial w_i} \propto -p_i \cdot (\mathbb{E}[\xi_i])^3. \tag{7}$$

In equilibrium, the investor requires compensation for both contributions. The first-order condition implies that, in addition to the standard systematic risk premium, firm i's required return includes a tail-risk premium:

$$\lambda^\tau(z_i, s_i) = \underbrace{\gamma \cdot p_i \cdot (\mathbb{E}[\xi_i])^2}_{\text{variance aversion}} + \underbrace{\delta \cdot p_i \cdot (\mathbb{E}[\xi_i])^3}_{\text{skewness aversion}}. \tag{8}$$

For parsimony, we collect both terms into a single reduced-form expression. Because jumps are one-sided, the variance and skewness terms are co-monotone in s_i: firms with higher $p_i \cdot (\mathbb{E}[\xi_i])^2$ also have more negative skewness. We can therefore write:

$$\lambda^\tau(z_i, s_i) = \mu \cdot p(z_i, s_i) \cdot (\mathbb{E}[\xi_i])^2, \quad \mu > 0, \tag{9}$$

where μ is a reduced-form price-of-risk parameter that absorbs both γ and the skewness aversion scaled by $\mathbb{E}[\xi_i]$. This simplification preserves the cross-sectional ranking of tail-risk premia because the variance and skewness channels reinforce each other: firms with higher jump variance also have more negative skewness, so a single index suffices.

The required return is:

$$r_i = r_f + \beta_i \mathbb{E}[R_m - r_f] + \lambda^\tau(z_i, s_i), \tag{10}$$

where the first two terms reflect compensation for systematic risk and λ^τ is the additional premium for firm-specific tail risk. We adopt the CAPM for the systematic component for simplicity, but the analysis extends to any asset pricing model that admits a firm-specific tail risk premium.

Firm value is then:

$$V_i = \frac{\bar{\pi}_i - p(z_i, s_i)\mathbb{E}[\xi_i]}{r + \lambda^\tau(z_i, s_i)}, \tag{11}$$

where $r \equiv r_f + \beta_i \mathbb{E}[R_m - r_f]$ collects the systematic risk terms. Low-s_i firms face a double penalty: lower expected cash flows (higher expected tail losses in the numerator) and higher discount rates (larger tail risk premium in the denominator).

2.4 Empirical Predictions

We now apply the model to derive predictions for our empirical analysis, exploiting the launch of ChatGPT on November 30, 2022. This event was unexpected by financial markets, and the extraordinary uptake of ChatGPT (e.g., reaching 100 million users within two months) made both AI's capabilities and its associated risks salient to investors. Before ChatGPT, generative AI was largely a specialist topic; afterward, it became a mainstream concern.

We model the release of ChatGPT as an event with two components. First, a common productivity update: the shock reveals AI is more productive than previously believed, so that $\phi_0 \to \phi_1 > \phi_0$. Second, a tail-risk update: before ChatGPT, investors' assessments of the risk parameters (η, ν) were attenuated, either because AI-related social risks were genuinely difficult to quantify absent a concrete demonstration of the technology's capabilities, or because these risks received insufficient attention in portfolio decisions (Bordalo, Gennaioli, and Shleifer, 2013). After the shock, investors updated towards higher effective values $\eta_1 > \eta_0$ and $\nu_1 > \nu_0$, implying a discrete increase in perceived tail-risk exposure and hence in the tail-risk premium $\lambda_1^\tau > \lambda_0^\tau$.

This formulation nests two interpretations. Under a Bayesian view, ChatGPT provided a public signal about the true deployment potential of generative AI, causing investors to raise both the productivity parameter ϕ and the risk-mapping parameters (η, ν). Under a salience-based view, the risk parameters were known but underweighted in pricing until a vivid public event made them concrete. Our empirical predictions are identical under either interpretation; we require only that the tail risk premium increases, i.e., that $\lambda_1^\tau > \lambda_0^\tau$, regardless of whether this reflects learning or reweighting.

Assuming the firm value in (11), the change in firm value around the shock is:

$$\Delta V_i = \underbrace{\frac{(\varphi_1 - \varphi_0)\bar{\pi}_0 z_i}{r}}_{\text{Productivity Gains}} - \underbrace{\frac{p(z_i, s_i)\mathbb{E}[\xi_i]}{r}}_{\text{Expected Tail Loss}} - \underbrace{\frac{\lambda_1^\tau(z_i, s_i)\bar{\pi}_i^{\text{new}}}{r^2}}_{\text{Tail Risk Discount}}. \tag{12}$$

Note that the first term, the productivity component of the shock, is independent of s_i, while the latter two terms, both the expected tail loss and the tail-risk discount, are decreasing in s_i, so the negative valuation impact of newly salient tail risks is larger for poorly-managed firms.

Prediction 1 (Value). *The change in firm value around the AI information shock is strictly increasing in social-risk management quality:* $\partial \Delta V_i / \partial s_i > 0$.

Moreover, the jump probability $p(z_i, s_i) = p_0 + \eta z_i (1 - s_i)^2$ implies that AI adoption amplifies the penalty for poor social-risk management. For firms with low z_i, the tail-risk (both probability and severity) is muted; for high-z_i firms, the difference between high- and low-s_i is amplified.

Prediction 2 (Interaction with AI Exposure)**.** *The value response to social-risk management quality is stronger for firms with greater AI adoption:* $\partial^2 \Delta V_i / \partial s_i \partial z_i > 0$.

These two predictions concern the overall valuation response to the AI information shock but remain silent about the relative importance of cash-flow versus discount-rate channels. In the model, both channels are potentially operative and important: valuation changes may reflect (i) revisions to expected cash flows, through higher productivity gains or higher expected losses from AI-related adverse events; or (ii) changes in discount rates arising from the repricing of tail risk. To distinguish between these two channels, we further derive the following testable predictions.

Because the difference between the first two terms in (12) is increasing in s_i, with the second term (expected tail-loss component) decreasing in s_i, expected cash flows are increasing in social-risk management quality. If the repricing operates through expected cash flows, firms with stronger (weaker) social-risk management should experience upward (downward) revisions in expected earnings around the AI information shock.

Prediction 3 (Cash-Flow Channel)**.** *Revisions in earnings forecasts around the AI information shock are increasing in social-risk management quality:* $\partial \Delta \mathbb{E}[CF_i] / \partial s_i > 0$.

The model also delivers an implication for the discount-rate channel. Since both p_i and $\mathbb{E}[\xi_i]$ are decreasing in s_i, and skewness scales with $-p_i \cdot (\mathbb{E}[\xi_i])^3$, the increase in perceived left-tail risk when AI risks become salient is larger for firms with weaker social-risk management. Accordingly, if the repricing operates through a tail-risk premium, we should observe differential changes in measures of perceived downside risk across firms.[3]

Prediction 4 (Changes in Skewness)**.** *The change in cash flow skewness around the AI shock is less negative for firms with superior social-risk management:* $\partial \Delta Skew_i / \partial s_i > 0$.

We test these predictions empirically by examining valuation responses to AI-related information shocks and distinguishing between cash-flow and discount-rate channels. The next section describes the data and empirical measurement used in the analysis.

3 Data and Measurement

This section describes the data and the construction of the key variables used to test the model's predictions.

3.1 Firm Valuations

We start our data collection with the CRSP universe of U.S. publicly listed firms and download daily stock prices for the period between November 2021 and December 2023. We add the returns of the Fama-French 3-factor portfolios, momentum portfolio returns, and the risk-free rate from Kenneth French's website.

To estimate each firm's market beta, we regress one year of daily excess returns on a constant and the daily market factor. For this, we use the period starting one month before the release of ChatGPT (the 250 trading days between November 3, 2021, and October 18, 2022, stopping 31 trading days before November 30, 2022). To ensure the

[3]Note that while the model also implies that volatility increases more for low-s_i firms, changes in implied volatility are difficult to attribute to a specific channel, as they may reflect general uncertainty or heterogeneous cash-flow dispersion. Changes in negative skewness more directly isolate the repricing of left-tail risk, which is the distinctive feature of the jump structure in our model.

precision of our estimates, we compute abnormal returns only for stocks for which we can observe a minimum of 70 daily observations in the estimation period. We compute CAPM alphas as the daily excess return on the stock minus the stock's beta times the market excess return. Abnormal returns based on other risk factor are computed analogously.

With this sample in hand, we obtain several firm-level variables based on the latest available financial information before the event date, using quarterly fundamentals from Compustat. Specifically, we compute: *Firm size* as the natural logarithm of total assets; *Market-to-book* as the market value of equity plus book value of liabilities divided by total assets; *Leverage* as the book value of debt relative to total assets; *Tangibility* as the firm's PP&E assets relative to total assets; *ROA* as the firm's return-on-assets; and *Beta* as the firm's market beta estimated using the CAPM model over trading days [-280, -31] relative to November 30, 2022.

3.2 ESG Ratings as a Proxy for Social-Risk Management

In the model, the key firm-level state variable is AI-related social-risk management quality, s_i, which is not directly observable to the econometrician. We therefore use ESG ratings as an empirical proxy. Formally, the ESG score S_i provides a noisy signal of the latent quality:

$$S_i = s_i + \epsilon_i, \quad \mathbb{E}[\epsilon_i] = 0. \tag{13}$$

ESG ratings are well suited to proxy for s_i for three reasons. First, they directly assess dimensions central to AI-related social risks. In particular, the Social pillar evaluates data privacy, product safety, human capital management, and customer welfare, which are precisely the domains in which AI adoption generates exposure to tail risks such as data misuse, algorithmic bias, AI-induced harms, and workforce disruption. Second, ESG ratings capture both risk exposure and risk management. For example, MSCI separately assesses firms' exposure to material risks and the effectiveness of their risk controls, mirroring the model in which both the probability p_i and severity ξ_i of AI-related tail events depend on s_i. Third, ESG ratings aggregate dispersed and otherwise hard-to-observe information about firm practices, synthesizing evidence from regulatory filings, incident databases, and third-party audits (Berg et al., 2022).

An alternative approach would construct proxies for s_i using textual or machine-learning methods applied to firms' disclosures. We argue that ESG ratings are preferable in this setting. They are designed to capture what firms do rather than what they say, are less sensitive to strategic disclosure or event-driven narrative adjustment, and mitigate concerns about researcher discretion because they are pre-existing third-party measures. Moreover, text-based measures tend to capture firms' exposure to or attention toward AI, not the quality of social-risk management *conditional* on AI adoption.

We begin with aggregate ESG scores to avoid cherry-picking, but also since the latent management quality s_i reflects a firm's overall capability to manage emerging risks, which does not necessarily align with a single pillar (e.g., environmental impacts from AI infrastructure may load onto Environmental pillar, while AI oversight relates to Governance pillar). We then decompose the scores into pillars to identify which dimensions drive the results. We obtain ESG ratings of firms from three different sources. Our preferred ESG rating provider is MSCI IVA (from which we obtain ESG scores for the year 2022), given their popularity among investors (Berg et al., 2024). Given the relatively large disagreement among individual ESG rating providers, we complement this with ESG ratings from Sustainalytics and S&P Global.

To validate that ESG ratings capture meaningful variation in firms' social-risk management quality, we conduct an out-of-sample predictive test. In Appendix Table A1, we examine whether ESG ratings available as of October 31, 2022 (i.e., before the ChatGPT shock) predict the incidence of cyber attacks and data breaches over 2023–2024, using cybersecurity incident data from Audit Analytics.[4] We find that higher-rated firms are significantly less likely to experience such events, consistent with ESG scores providing an informative signal about latent management quality s_i. This result establishes a direct link between the proxy and the type of discrete adverse events that generate jump risk in the model.

3.3 Proxies for AI Adoption

It is important to control for AI driven productivity gains in order to isolate them from the effects attributed to ESG ratings. A key concern is that ESG ratings may simply proxy for overall firm quality and thus indirectly capture productivity improvements associated with AI adoption. To address this concern, and to examine how AI adoption interacts with social risk exposure, we need empirical proxies for firm-level AI adoption intensity (z_i, in the model). We use three complementary measures:

1. *Change in share of AI workers* (Babina et al., 2024) measures the change from 2010 to 2021 in the fraction of a firm's workforce with AI skills, based on employee resumes. This captures realized investments in AI human capital prior to the event.

[4] It is crucial that the incidents occur after the measurement date of the ESG ratings. Otherwise, the scores might respond to past incidents rather than capturing forward-looking risk-management capacity.

2. *Workforce AI exposure* (Eisfeldt et al., 2025) quantifies the extent to which a firm's workforce performs tasks that are susceptible to automation by generative AI. This captures potential exposure to AI driven productivity changes and disruption.

3. *Data assets* (Eisfeldt et al., 2025) captures the firm's stock of data that can be leveraged by AI technologies. This captures how firms with larger data assets are better positioned to benefit from, but also more exposed to, AI adoption.[5]

We combine the three measures to construct an indicator for high AI adoption intensity and exposure, $GenAI$, which equals one if a firm is above the sample median on all three dimensions. This definition classifies about 17% of firms in our sample as having substantial AI readiness.

3.4 Earnings Forecasts and Options-Implied Tail-Risk

To examine whether ESG scores are associated with changes in expected cash flows, we use analyst earnings forecast data from I/B/E/S. We focus on forecast revisions over the three months following the release of ChatGPT, from December 2022 through February 2023, a window during which information about the productivity implications of generative AI became salient. We study revisions at three forecasting horizons: one year ahead, five years ahead, and long term growth forecasts. Following the standard approach in the literature, we measure analyst revisions using standardized unexpected forecasts (SUF), defined as the change in the consensus earnings forecast scaled by the cross sectional standard deviation of forecasts. This normalization facilitates comparison across firms and horizons and isolates economically meaningful revisions in expected cash flows.

[5] We thank the authors for making their data publicly available.

We measure option-implied downside tail risk using two complementary quantities based on equity options data from the Ivy DB OptionMetrics Surface File. First, we construct model-free implied skewness following Bakshi, Kapadia, and Madan (2003), which summarizes the asymmetry of the risk-neutral return distribution. Skewness is defined as the third central moment of the risk-neutral distribution, normalized by the variance raised to the power of 3/2. A decrease in skewness indicate a shift of risk-neutral probability mass from the right to the left tail, suggesting a higher relative price of downside outcomes. But note that this skewness measures captures changes in tail asymmetry but could be affected by both downside and upside tails.

Second, we measure downside tail risk using the implied volatility slope, SlopeD following Kelly et al. (2016). SlopeD is estimated as the slope coefficient from regressing implied volatilities of out-of-the-money put options (with deltas between -0.5 and -0.1) on their corresponding deltas. An increase in SlopeD indicates that deeper out-of-the-money puts are relatively more expensive, reflecting a higher marginal cost of protection against extreme downside tail events. Because SlopeD measures relative expensiveness across strikes, it is orthogonal to the overall level of implied volatility and therefore comparable across assets with different baseline risk, isolating the pricing of downside tail risk within the left tail.

A potential concern is that our findings are driven by cash-flow effects specific to technology firms rather than by the pricing of AI-related social risk. Technology firms tend to have higher ESG ratings and are more likely to develop frontier AI technologies, making it difficult to disentangle downside-risk pricing from expected AI-driven growth.

To address this concern, we exclude technology firms from the sample, thereby focusing on firms that primarily adopt and deploy AI rather than develop it.[6] We winsorize all continuous variables except for ESG scores at the 1% and 99% levels.[7] The summary statistics of our main variables are presented in Table 1.

[Table 1 about here.]

4 Results

We begin by examining the relationship between firms' ESG ratings and cumulative announcement returns around the launch of ChatGPT, and then test how this relationship varies with firms' exposure to AI. We next investigate which ESG dimensions drive the observed effects. To distinguish between expected cash-flow and discount-rate channels, we study changes in analyst earnings forecasts and option-implied measures of downside tail risk. Finally, we use the unexpected removal of Sam Altman as OpenAI's CEO as an out-of-sample test of our main hypotheses.

4.1 ESG Ratings and Announcement Returns

To test the relationship between abnormal returns around the AI shock and firms' ESG scores, we estimate the following linear OLS regression:

$$CAR\left[-1,+10\right]_i = \beta ESG_i + \alpha \mathbf{\Gamma_i} + \mu_i + \epsilon_i \tag{14}$$

[6]Following Acemoglu et al. (2022), we define the technology sector as NAICS 51 ("Information") and NAICS 54 ("Professional, Scientific, and Technical Services").

[7]We are not winsorizing ESG scores since these are bound between 0 and 10 by the ESG rating agency.

$CAR\,[-1,+10]_i$ is the cumulative abnormal return for firm i from one trading day before to ten trading days after the launch of ChatGPT. ESG_i is the ESG rating of firm i, as provided by MSCI IVA, available one month before the event. $\mathbf{\Gamma_i}$ is a vector of firm-level controls, μ_i is an industry fixed effect, and ϵ_i is the error term clustered at the NAICS 3-digit industry level. The control variables include firm size, market-to-book ratio, leverage ratio, asset tangibility, and return-on-assets (ROA). We additionally control for a firm's market beta to account for changes in systematic risk associated with AI adoption: as shown by Babina et al. (2025), firms with greater AI investment tend to experience increases in their market betas as they become more growth-oriented. This control addresses the concern that ESG characteristics may be correlated with AI-induced shifts in systematic risk rather than directly with investors' assessment of AI-related risks.

Table 2 shows the results. Overall, the results show a positive and significant relationship between ESG scores and CARs, which is also economically sizable. For example, the results in column 1 imply that a one-standard-deviation higher ESG score is associated with 1.11pp ($= 1.168 \times 0.95$) higher CARs. This strongly supports Prediction 1 from the model: the stock market reaction to ChatGPT is significantly more positive for firms with higher ESG ratings.

To assess whether the relationship between ESG performance and CARs is linear, we replace the continuous ESG score with indicator variables for its five quintiles, using the third quintile as the benchmark. Column (2) shows that the effect is concentrated among firms with poor ESG performance. Firms in the lowest ESG quintile experience, on average, a 2.29pp more negative market reaction relative to firms in the middle quintile. By contrast, differences across higher ESG quintiles are economically smaller and statistically insignificant.

[Table 2 about here.]

This pattern has two implications. First, it closely matches the mechanism in our model where tail-risk exposure is convex in poor social-risk management, thus firms with the weakest ESG practices face a disproportionately higher probability and severity of AI-related adverse events. Consequently, the marginal valuation effect of social-risk management is largest in the lower tail, generating a pricing response that is concentrated among poorly managed firms rather than spread symmetrically across the ESG distribution. Second, the absence of a positive premium for top-quintile ESG firms mitigates the concern that ESG scores simply proxy for overall firm quality or superior productivity gains from AI adoption. Instead, the evidence points to an asymmetric downside-risk channel in which investors penalize firms with the greatest exposure to AI-related tail risk, rather than rewarding high-ESG firms for expected cash-flow improvements.

Note that beyond industry affiliation, there may be substantial within-industry heterogeneity in firms' exposure to AI that is not captured by our industry fixed effects. If low-ESG firms within a given industry are also slower or less effective adopters of AI, the estimated relationship could reflect differential AI-driven productivity expectations rather than the pricing of downside risk. To address this concern, we explicitly control for firm-level AI adoption intensity. Column (3) includes all three of our AI measures: Δ*Share AI Workers* from Babina et al. (2024), as well as *Workforce AI Exposure* and *Data Assets* from Eisfeldt et al. (2025); while column (4) uses our composite *GenAI* indicator that aggregates these dimensions. Across specifications, the coefficient on the

ESG score remains robust. [8] Moreover, Appendix Table A3 shows no significant relationship between ESG scores and any of our AI adoption measures, indicating that relatively more sustainable firms are neither better nor worse than relatively less sustainable firms in terms of their pace of AI adoption. Taken together, these results suggest that the repricing we document is not driven by within-industry differences in AI-related productivity potential, but instead reflects investors' assessment of firms' exposure to AI-related downside social risks across the broader economy.

4.1.1 Interaction with AI Exposure

In columns (5) and (6), we interact ESG scores with *GenAI*. The coefficient on $ESG \times GenAI$ is positive and highly significant (1.98, $t = 2.84$), confirming the model's prediction that social-risk management and AI exposure are complementary. The relation between ESG scores and CARs is substantially stronger for firms with high AI exposure, consistent with the idea that social-risk management becomes economically more relevant precisely when firms adopt AI more intensively and investors are more likely to perceive and price AI-related downside risks.

Using ESG quintile indicators in column (6), we find that the interaction effect is particularly pronounced for firms in the lowest ESG quintile: firms in the lowest ESG-quintile with high AI exposure ($GenAI$=1) experience an 4.79pp ($t = -3.49$) more negative announcement returns relative to other firms. At the same time, the standalone coefficients on ESG (and on ESG Q1) become insignificant once the interaction is included. This

[8] The number of observations decreases from columns (1)–(2) to columns (3)–(6) due to missing data for the AI exposure variables. In Appendix Table A2, columns (7) and (8) re-estimate the specifications in columns (1) and (2) on the restricted sample for which AI exposure data are available; the coefficients are qualitatively similar.

pattern indicates that ESG matters primarily through its interaction with AI exposure, rather than as a generic firm-quality or sustainability characteristic.

These results imply that the market penalty for poor social-risk management is amplified when firms are more exposed to AI, exactly in line with Prediction 2 of the model. More broadly, they suggest that the effects we document are specific to AI-related social risks rather than to other ESG risks unrelated to AI.

4.1.2 Robustness

In this section, we briefly summarize the battery of robustness tests that confirm our main findings. To conserve space, we only report the main coefficients of interest in Table 3.

First, we replicate our main results in column (4) of Table 2 using event windows that start one trading day before the event and end, respectively, one, three, five, and fifteen days after the event. Moreover, to ensure that our effects are not transitory, Figure 2 plots the daily ARs over one month after the event date. We do not observe any reversals.

Second, we estimate abnormal returns relative to the four-factor Fama-French-Carhart model and obtain results that are almost identical in magnitude to those of our baseline specification.

Third, we use two alternative ESG rating providers, S&P Global and Sustainalytics, and apply the noise-correction procedure proposed by Berg et al. (2021) as well as taking the average of the three ESG ratings.[9] Appendix Figure A1 showcases the results based on quintiles of ESG scores for the two alternative rating providers.

Finally, we follow Cohn et al. (2024) to account for the cross-sectional correlations in

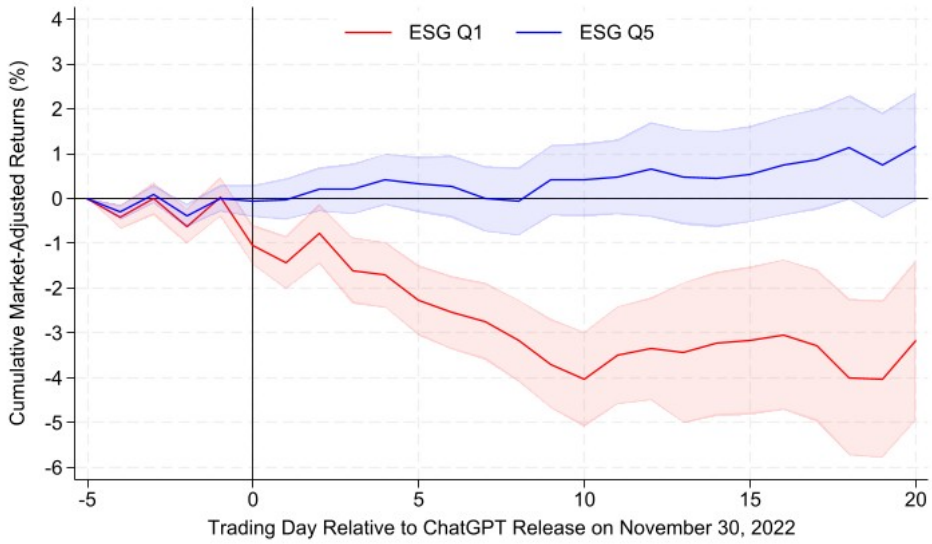
[9] The coefficient on Sustainalytics is negative because Sustainalytics reports *risk* ratings, i.e., higher values imply a worse ESG performance.

event returns and address concerns about false positives.[10]

[Table 3 about here.]

Figure 2: Cumulative Abnormal Returns Over 20 Trading Days

This figure plots the abnormal returns from -5 to 20 trading days around the launch of ChatGPT on November 30, 2022, averaged across stocks with ESG scores in the top quintile (blue) and bottom quintile (red). We use the MSCI IVA scores available by October 31, 2022. The ARs of each stock are calculated using the CAPM model estimated over trading days [-280, -31]. The shaded area corresponds to the 95% confidence intervals.



4.2 Which ESG Dimensions Matter?

This section explores which dimensions of ESG drive the relationship, following our "funnel" approach from aggregate ESG to pillar scores to sub-scores.

[10]The coefficients for this test are reported in absolute values and not in percentages, which means that the main coefficient of interest is about 1.5pp, in line with the magnitudes documented in the previous section.

4.2.1 E/S/G Pillar Decomposition

We start by examining the effect of the individual Environmental, Social, and Governance pillar scores on the announcement returns. We run the following linear OLS regression, where Env_i, Soc_i, and Gov_i are the respective pillar scores:

$$CAR\ [-1,+10]_i = \beta_1 Env_i + \beta_2 Soc_i + \beta_3 Gov_i + \beta_4 GenAI_i + \alpha\mathbf{\Gamma_i} + \mu_i + \epsilon_i \tag{15}$$

[Table 4 about here.]

The results are reported in Table 4. We start in column (1) by looking at the relationship between CARs and firms' Social pillar score. As expected, we find a positive and statistically significant effect: a one standard deviation increase in the Social pillar scores relates to a 0.53 percent higher CAR (=1.48×0.361).

Column (2) tests Prediction 2 by interacting the pillar score with the *GenAI* indicator. The interaction effect between *GenAI* and *Soc* is positive and significant (1.02, $t = 2.16$), while the baseline coefficient on the Social pillar is close to zero (-0.03, $t = -0.12$). In other words, the positive effect of the Social pillar score is concentrated in firms with high AI adoption or substantial AI investment, which are more exposed to the AI-related tail risks.

Columns (3) and (4) examine the Environmental and Governance pillar, respectively. We find insignificant results for both pillar scores, suggesting that the relevant information for firms' AI-related social-risk management quality is captured by the Social pillar scores. Finally, column (5) includes all three pillar scores and their interactions with *GenAI* jointly. The $Social \times GenAI$ interaction remains positive and significant (0.949,

$t = 2.02$). This distinction is consistent with the model's mechanism: AI adoption intensifies exposure to social risks such as data privacy violations and product safety failures, risks captured by the Social pillar, whereas the environmental costs of AI (e.g., energy consumption) may depend more on external factors such as the availability of renewable energy (Feher et al., 2025) than on the intensive margin of firm-level AI adoption.

4.2.2 Sub-Score Analysis

In the next step, we break down the ESG scores into the various criteria provided by MSCI IVA. Table 5 shows the results from linear regressions of the CARs on the individual components of the ESG scores. We report only the components available for at least a third of our sample. To conserve space, we only tabulate the main coefficient of interest for our preferred specification, which controls for firm characteristics, $GenAI$, and industry fixed effects.

We find that data privacy components load significantly and with the expected signs. For example, a one standard deviation decrease in *Privacy data security* relates to CARs that are 2.2pp more negative ($= 0.76 \times 2.92$). This aligns with our interpretation that investors reward high-ESG firms because of their lower exposure and better ability to manage AI-related tail risks. We also find significant effects for better human capital development and corruption practices. The former likely reflects firms' ability to attract AI talent, while the latter likely reflects firms' exposure to cybersecurity risks.

[Table 5 about here.]

4.3 Channels: Expected Cash Flows versus Discount Rate

The predictions tested so far concern the overall valuation response to the AI information shock, but they remain silent about the relative importance of cash-flow versus discount-rate channels. In the model, both channels are potentially operative. Changes in firm value may reflect revisions to expected cash flows, either through (i) higher expected productivity gains from AI adoption; (ii) higher expected losses associated with AI-related adverse events; or (iii) changes in discount rates arising from the repricing of tail risk. Distinguishing between these channels is essential for understanding whether markets interpret AI primarily as a source of growth prospects or as a source of heightened downside risk. We now turn to empirical tests that separately identify these channels.

4.3.1 The Cash Flow Channel

The repricing we document could be driven by changes in expected cash flows. As shown in our model, firms with weaker social-risk management face higher expected losses from AI-related adverse events, reducing expected earnings (the numerator effect in equation (11)). If investors update these expectations around ChatGPT's release, we would expect differential analyst forecast revisions. Table 6 reports regressions of standardized unexpected forecasts (SUF) on ESG scores at various horizons. The results show no significant relationship between ESG scores and analyst forecast revisions at any horizon—one-year, five-year, or long-term forecasts. This absence of differential forecast revisions suggests that the repricing we document does not operate primarily through changes in expected cash flows.

[Table 6 about here.]

We find no significant relationship between ESG scores and analyst forecast revisions at any horizon: one-year, five-year, or long-term earnings growth. This evidence does not support the interpretation that ESG scores proxy for differences in expected AI-driven productivity gains. The absence of forecast revisions suggests that the repricing we document is not driven by differential cash-flow expectations.

4.3.2 The Discount Rate Channel

We now turn to the discount rate channel. In our model, investors demand a tail-risk premium for bearing exposure to discrete adverse events. This premium reflects compensation for negatively skewed payoffs, a higher-moment effect distinct from expected changes in cash flows. Our model predicts that firms with weaker social-risk management will experience a larger increase in perceived left-tail risk when AI risks become salient.

To test this prediction, we use two complementary option-based measures of skewness and downside tail risk constructed from equity options data. First, we measure changes in model-free implied skewness ($\Delta Skew$) following Bakshi et al. (2003), which captures the asymmetry of the risk-neutral return distribution. A decline in implied skewness reflects a shift of risk-neutral probability mass toward the left tail, indicating a higher relative price of downside outcomes, though this measure may also reflect movements in the right tail.

As shown in columns (1)-(3) of Table 7, we find a significant positive relationship between firms' ESG scores and changes in realized skewness from November to December 2022. Firms with lower ESG scores experience a larger decline in skewness around the event, indicating a more pronounced shift of return probability mass toward the left tail.

The coefficient on ESG in column (1) is 0.006 ($t = 2.34$), implying that a one–standard deviation lower ESG score is associated with an approximately 8% larger decline in skewness ($= 0.006 \times 0.95/0.073$), relative to the pre-event average skewness of 0.073. Column (2) shows that this effect is mostly driven by the Social pillar (0.004, $t = 3.11$).

[Table 7 about here.]

We further measure downside tail risk using the implied volatility smirk slope, *SlopeD*, following Kelly, Pástor, and Veronesi (2016). *SlopeD* is estimated from the cross-section of out-of-the-money put implied volatilities (deltas between -0.5 and -0.1); an increase in *SlopeD* indicates that deeper out-of-the-money puts become relatively more expensive, reflecting a higher marginal cost of protection against extreme downside events. Because it captures relative pricing across strikes, *SlopeD* is orthogonal to the overall level of implied volatility and isolates variation in downside tail risk.

Turning to the *SlopeD* results in columns (4)-(6) of Table 7, we find a significant negative relationship between ESG scores and changes in the price of downside protection. Firms with lower ESG scores experience a larger increase in *SlopeD* around the event (column (4): -0.014, $t = -1.83$), indicating a larger rise in perceived downside tail risk. Column (5) shows that the effect is again mostly driven by the Social pillar (-0.006, $t = -1.81$). These findings are consistent with Prediction 4 of the model: investors perceive firms with weaker social-risk management as more exposed to AI-related tail risks and price this exposure into option markets.

The interaction term between the social pillar and *GenAI* in columns 3 and 6 is not statistically significant. Two considerations explain why the complementarity with AI exposure is weaker here than in the return regressions. The first is econometric.

The option-matched sample is smaller, reducing power for detecting interaction effects. Options-based measures are also noisier than cumulative returns: realized skewness is estimated from a short window of daily returns, and $SlopeD$ depends on the availability and liquidity of out-of-the-money puts. Both factors may attenuate interaction coefficients toward zero.

The second is conceptual. Option-based tail-risk measures primarily capture how investors reprice downside states once AI-related social risks become salient. This repricing can occur broadly across firms with weak social-risk management and does not need to vary strongly with the intensity of current AI adoption. By contrast, the effect of this repricing on firm value depends on each firm's *exposure* to the repriced states. In our model, AI adoption amplifies this exposure, so a given shift in the pricing of tail risk translates into a larger valuation impact for high-AI firms with poor social-risk management. As a result, the ESG–AI interaction appears most clearly in returns where common tail-risk repricing is scaled by firm-specific AI exposure, rather than in option-implied measures, which primarily capture the repricing of downside risk itself.

Taken together, the evidence in this subsection points to a discount-rate channel driven by the repricing of AI-related tail risk. We find no indication that ESG scores are associated with differential revisions in expected earnings around the ChatGPT release, inconsistent with a cash-flow channel. Instead, firms with weaker social-risk management experience larger increases in option-implied downside risk, with effects concentrated in the Social pillar and consistent with our model's focus on discrete adverse events and tail-risk premia.

4.4 Out-of-Sample Validation

4.4.1 Portfolio Evidence: AI-Related Tail-Risk Premia

Our model implies that investors require compensation for bearing AI-related tail risk, so that firms with high exposure to such risk should earn higher expected returns, in the absence of news AI information shocks. The event-study evidence documents repricing around discrete shocks, but cannot speak to whether this repricing indeed generates a persistent risk premium. To probe this, we run a portfolio test using only information available prior to the release of ChatGPT.

Specifically, we form two long-short portfolios. The first sorts firms with high AI exposure ($GenAI = 1$) into ESG quintiles and goes long the bottom quintile (highest tail-risk exposure) and short the top quintile (lowest tail-risk exposure). The second applies the identical sort to firms with low AI exposure ($GenAI = 0$). Both portfolios are equally weighted and formed on January 1, 2023, using pre-event ESG scores and AI exposure measures, and held without rebalancing through December 31, 2024.[11]

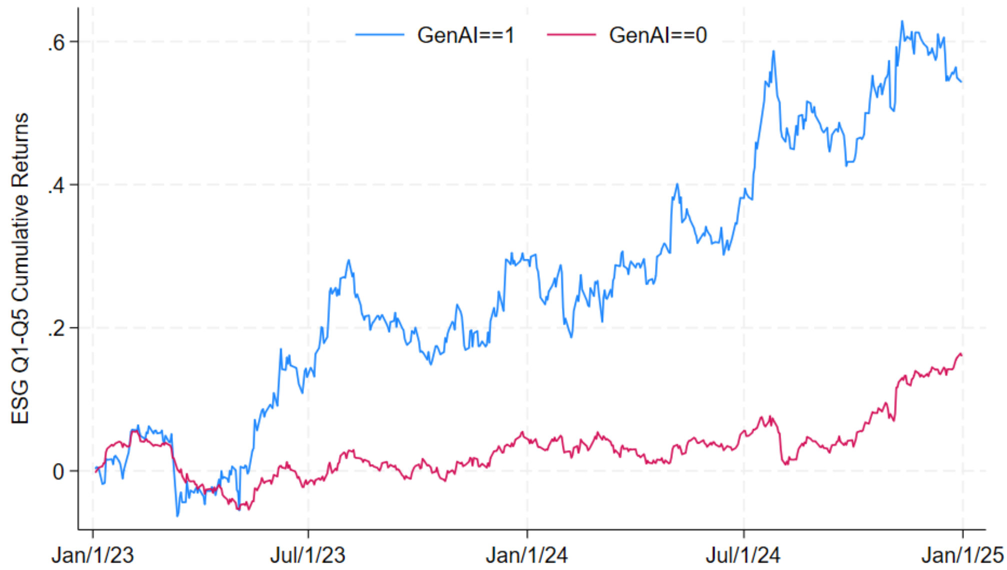
Figure 3 plots the cumulative log returns of both portfolios. Among high-AI-exposure firms, the long-short portfolio accumulates returns of approximately 55% over the two-year holding period, with a broadly monotonic upward trajectory.[12] To put this in perspective, the S&P 500 returned approximately 48% over the same period. In contrast, the same ESG-based sort within the low-AI-exposure group generates cumulative returns of only about 16%, with most of the gains realized only after the 2024 presidential election

[11] We apply equally-weighting since the sample of high AI exposure firms is small, and value-weighting might capture the returns of individual stocks. Nevertheless, our results remain broadly consistent when using value-weighting.

[12] We exclude technology firms from this analysis, so the portfolio returns do not simply reflect the broader AI-driven rally in technology stocks.

Figure 3: Cumulative Returns of ESG Q1–Q5 Long-Short Portfolios

This figure plots daily cumulative log returns of two long-short portfolios from January 2023 through December 2024. Each portfolio goes long stocks in the bottom ESG quintile (Q1) and short stocks in the top ESG quintile (Q5), formed using MSCI IVA scores available by October 31, 2022. The blue line restricts the sort to firms with high AI exposure ($GenAI = 1$); the red line restricts to firms with low AI exposure ($GenAI = 0$), with $GenAI$ indicating if a firm is above the sample median on all three AI exposure measures: ΔShare AI workers, Workforce AI exposure, and Data assets. Portfolios are equal-weighted and held without rebalancing.



announcement.

Table 8 formalizes this comparison. We estimate the portfolio alphas using different risk factor models: Fama and French (1993) three-factor model, Carhart (1997) four-factor model, Fama and French (2015) five-factor model, and the five-factor plus momentum model. The high-AI-exposure portfolio earns a statistically significant daily alpha of 8-9 basis points across all specifications, while the low-AI-exposure portfolio generates near-zero and insignificant alpha under any specification.

[Table 8 about here.]

These results have three implications. First, the tail-risk repricing documented in our event study is not transitory: it generates a persistent return spread consistent with investors demanding ongoing compensation for bearing AI-related social risk. Second, the stark contrast between the two portfolios rules out the interpretation that our findings simply capture an ESG risk premium unrelated to AI. If ESG scores proxied for a generic firm-quality characteristic that commands a risk premium regardless of AI exposure, both portfolios should generate significant alphas. The fact that only the high-AI-exposure portfolio does so indicates that the premium is specific to the interaction of weak social-risk management with AI adoption, exactly as implied by our model. Third, the absence of alpha in the low-AI-exposure portfolio mitigates the concern that ESG scores proxy for unobserved dimensions of firm quality correlated with future returns, since any such confound would need to operate exclusively among AI-intensive firms.

We note two caveats regarding the magnitude of the estimated alpha. First, the portfolio is formed from a relatively narrow subset of firms (*GenAI* firms constitute approximately 17% of our sample, totaling 211 stocks, and the long-short strategy selects from the extreme ESG quintiles within this group) and is held without rebalancing, so the implied alpha does not account for transaction costs and capacity constraints. The economic significance lies in the cross-sectional pattern, i.e., the comparison within high-AI-exposure firms, rather than the absolute magnitude. Second, while the t-statistics indicate statistical significance at the 5% level, they remain below 3, so we do not claim to have identified a novel priced risk factor (Harvey, Liu, and Zhu, 2016). Rather, the portfolio test serves as an out-of-sample validation that the AI-related tail-risk channel documented in our event study carries forward into realized returns.[13]

[13]For context, Eisfeldt et al. (2025) report a daily alpha of 0.45% for a portfolio sorted on workforce exposure to generative AI, which captures the productivity upside of AI adoption; our estimated alpha of approximately 0.08% is considerably smaller, consistent with the interpretation that it reflects a tail-risk premium for the *downside* of AI adoption rather than expected productivity gains. The productivity gains will be differenced out in our high-minus-low ESG portfolio with high AI exposure.

4.4.2 The Sam Altman Event

Finally, as another out-of-sample test, we exploit the unexpected and short-lived ousting of Sam Altman, CEO of OpenAI, on November 17, 2023. Unlike our main event, the launch of ChatGPT, this episode did not convey new information about AI capabilities or productivity, but instead reflected heightened concerns about governance, safety, and oversight in the development and deployment of advanced AI systems. We therefore interpret this episode as an exogenous shock to perceived AI social risk. The decision surprised OpenAI employees, investors, and the broader public, and was widely attributed to disagreements between Altman and the board over the pace of AI development and the handling of AI security and safety issues.[14] Contemporary reporting points to alleged misrepresentations to the board regarding AI security practices and the approval of field trials and enhancements to GPT-4 (Reuters, 2023a,b). Although Altman was reinstated as CEO within five days following internal opposition and the threat of mass resignations, the episode underscores persistent tensions between rapid AI deployment and social-risk management. We employ shorter event windows for this shock to reflect its brief duration.

This event can be seen as a negative shock to the saliency of AI-related risks. If our model is correct and high-ESG firms are better equipped to manage these risks, we would expect them to perform better around the event. We test this by examining the CARs in the three-day event window surrounding the ousting.

Table 9 presents the results. Column (1) shows that a one-standard-deviation increase in ESG scores is associated with 0.47pp (= 0.489 × 0.95) higher CARs around the event and statistically significant at the 5% level. When we decompose into quintiles (column

[14] WSJ (2025) provides a detailed account; see also Hagey (2025).

(2)), firms in the bottom two ESG quintiles experience significantly lower returns than the middle quintile: ESG Q1 firms underperform by 0.81pp ($t = -1.83$), and ESG Q2 firms underperform by 0.57pp ($t = -1.72$). These effects remain largely unaffected after controlling for firms' AI exposure in column 3.

Column (4) tests the interaction prediction using the *GenAI* indicator. Consistent with our main findings, the interaction $ESG \times GenAI$ is positive and significant (0.84, $t = 2.31$), indicating that the relationship between ESG scores and returns around the Altman event is substantially stronger for firms with high AI exposure. While overall returns were negative across all firms, those in the bottom quintile in terms of ESG ratings with high AI exposure performed significantly worse than their higher-scoring peers. Overall, these results are consistent with investors viewing firms with weak social-risk management as more exposed to AI-related risks.

[Table 9 about here.]

5 Conclusion

Guided by an AI-related tail-risk model, we document how the equity market prices these risks in. We exploit a major AI-related information shock, the release of ChatGPT on November 30, 2022. Over the two weeks following the event, firms with low ESG scores experience significantly lower cumulative abnormal returns than firms with high ESG scores. The return differential is economically large and driven primarily by differences in the Social pillar, particularly firms' data privacy and product safety practices. These results are robust to controlling for multiple proxies of AI-driven productivity exposure, like firms' AI-related labor inputs, workforce AI exposure, and data assets. The effects are most pronounced among firms with greater ex-ante AI exposure.

To understand the underlying mechanisms, we study implied risk measures from stock options and revisions in analyst forecasts. Around the ChatGPT release, low-ESG firms exhibit significantly larger increases in option-implied downside risk, as reflected in both realized skewness and the implied volatility smirk slope. In contrast, we find no corresponding differential revisions in analysts' earnings forecasts at either short or long horizons.

Out of sample tests confirm the validity of our findings. First, we construct portfolios going long low ESG stocks and short high ESG stocks, separately for firms with and without high ex-ante AI exposure. If the implications of our model are correct, such portfolios will generate abnormal returns since investors require compensation for bearing AI-related downside risk. Indeed, we find that the portfolio of high AI-exposure firms starkly outperforms the low-AI-exposure portfolio over the period from 2023 to 2024. Second, we run an out-of-sample test using the temporary ousting of OpenAI's CEO, Sam Altman, an event that heightened the salience of AI safety concerns without signaling productivity improvements. Our findings are again consistent with the pricing in of AI-driven tail risks: low-ESG firms again underperform, with effects concentrated among firms with high AI exposure.

Taken together, our results align closely with a parsimonious tail-risk model in which AI adoption exposes firms to discrete adverse events such as data breaches, algorithmic discrimination, and regulatory penalties, whose probability and severity depend on the quality of firms' social-risk management. ESG ratings—particularly the Social pillar—

contain valuable information that proxies for this latent firm characteristic. When exogenous shocks update beliefs about AI-related social risks, investors reprice firms by updating discount rates to reflect changes in perceived tail risk. However, firms with stronger social-risk management are perceived as better positioned to absorb these tail risks.

References

Acemoglu, Daron, 2025, The simple macroeconomics of AI, *Economic Policy* 40, 13–58.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo, 2022, Artificial intelligence and jobs: Evidence from online vacancies, *Journal of Labor Economics* 40, S293–S340.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson, 2024, Artificial intelligence, firm growth, and product innovation, *Journal of Financial Economics* 151, 103745.

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson, 2023, Firm investments in artificial intelligence technologies and changes in workforce composition, *Working Paper* .

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson, 2025, Artificial intelligence and firms' systematic risk, *Working Paper* .

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan, 2003, Stock return characteristics, skew laws, and the differential pricing of individual equity options, *The Review of Financial Studies* 16, 101–143.

Bao, Leo, Difang Huang, and Chen Lin, 2024, Can artificial intelligence improve gender equality? evidence from a natural experiment, *Management Science, forthcoming* .

Berg, Florian, Florian Heeb, and Julian Kölbel, 2024, The economic impact of ESG ratings, *Working Paper* .

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, 2022, Aggregate confusion: The divergence of ESG ratings, *Review of Finance* 26, 1315–1344.

Berg, Florian, Julian Kölbel, Anna Pavlova, and Roberto Rigobon, 2021, ESG confusion and stock returns: Tackling the problem of noise, *Working Paper* .

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013, Salience and asset prices, *American Economic Review* 103, 623–28.

Cao, Sean, Wei Jiang, Junbo Wang, and Baozhong Yang, 2024, From man vs. machine to man + machine: The art and AI of stock analyses, *Journal of Financial Economics* 160, 103910.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Cohn, Jonathan B., Travis L. Johnson, Zack Liu, and Malcolm Wardlaw, 2024, Past is prologue: Inference from the cross section of returns around an event, Working paper.

Eisfeldt, Andrea L., and Gregor Schubert, 2024, AI and finance, *Working Paper* .

Eisfeldt, Andrea L., Gregor Schubert, Miao Ben Zhang, and Bledi Taska, 2025, The labor impact of generative AI on firm values, Working paper.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, 2024, GPTs are GPTs: Labor market impact potential of LLMs, *Science* 384, 1306–1308.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel, 2020, Hedging climate change news, *The Review of Financial Studies* 33, 1184–1216.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Feher, Adam, Emilia Garcia-Appendini, and Roxana Mihet, 2025, AI investment and climate policy, *Working Paper* .

Gillan, Stuart L, Andrew Koch, and Laura T Starks, 2021, Firms and social responsibility: A review of ESG and CSR research in corporate finance, *Journal of Corporate Finance* 66, 101889.

Guenzel, Marius, Shimon Kogan, Marina Niessner, and Kelly Shue, 2025, AI personality extraction from faces: Labor market implications, *Working Paper* .

Hagey, Keach, 2025, *The Optimist: Sam Altman, OpenAI, and the Race to Invent the Future* (W.W. Norton Company).

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, … and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *The Journal of Finance* 55, 1263–1295.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, 2024, ChatGPT and corporate policies, *Working Paper* .

Kelly, BRYAN, ĽUBOŠ Pástor, and PIETRO Veronesi, 2016, The price of political uncertainty: Theory and evidence from the option market, *The Journal of Finance* 71, 2417–2480.

Kogan, Leonid, and Dimitris Papanikolaou, 2014, Growth opportunities, technology shocks, and asset prices, *Journal of Finance* 69, 675–718.

Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *The Journal of Finance* 31, 1085–1100.

Reuters, 2023a, OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, Accessed: 2024-11-24.

Reuters, 2023b, Sam Altman to return as OpenAI CEO after his tumultuous ouster, Accessed: 2024-11-23.

Sunstein, Cass R, 2024, AI, reducing internalities and externalities, *Working Paper* .

WSJ, 2025, The secrets and misdirection behind Sam Altman's firing from OpenAI, Accessed: 2025-05-21.

Table 1: Data Sample and Summary Statistics

This table reports the summary statistics of our main variables. *CAR[-1, +10]* is the abnormal return of each stock, calculated using the CAPM model estimated over [-280, -31] trading days, and then cumulated over [-1, +10] trading days around the release of ChatGPT on November 30, 2022. *ESG* is the most recent MSCI IVA ESG score of a firm as of October 31, 2022, while *Env*, *Soc*, and *Gov* refer to the Environmental, Social, and Governance pillar scores, respectively. *S&P Global ESG* and *Sustainalytics ESG* are standardized ESG scores from two other rating agencies, S&P Global and Sustainalytics. *Firm size* is the natural logarithm of total assets; *Market-to-book* is calculated as (PRCC_F*CSHPRI+ DLC+DLTT+PSTKL - TXDITC)/AT; *Leverage* is calculated as (DLTT + DLC)/AT; *Tangibility* is the firm's tangible assets relative to total assets ratio, calculated as PPENT/AT; *ROA* is return-on-asset, calculated as OIADP/AT. All firm characteristics are obtained from Compustat for the fiscal quarter prior to November 2022. Δ*Share AI workers* is the change in the share of a firm's AI-related employees from 2010 to 2021 following Babina et al. (2024). *Workforce AI exposure* measures workforce exposure to generative AI, *Data assets* captures the firm's stock of data that could be deployed with AI technologies, both are from Eisfeldt et al. (2025). *GenAI* is an indicator equal to one if a firm is above the sample median on all three AI exposure measures. Δ*Skew* and Δ*SlopeD* measure changes in realized skewness and the implied volatility slope around the event, respectively. *Altman's removal CAR [-1, +2]* measures firms' cumulative abnormal returns around the announcement of Sam Altman's unexpected removal from OpenAI on November 17, 2023. All continuous variables, except ESG scores, are winsorized at the 1% and 99% levels.

				Percentile				
	N	Mean	St. Dev.	10th	25th	50th	75th	90th
CAR [-1, +10] (%)	1,830	-0.62	10.03	-11.19	-5.53	-0.75	3.82	9.35
ESG	1,830	4.92	0.95	3.70	4.30	4.90	5.60	6.20
Env	1,830	4.61	2.37	1.60	2.80	4.50	6.40	7.60
Soc	1,830	4.47	1.48	2.70	3.50	4.30	5.40	6.40
Gov	1,830	5.84	1.08	4.40	5.20	5.90	6.60	7.10
S&P Global ESG (standardized)	1,792	-0.03	0.90	-1.00	-0.67	-0.18	0.36	1.29
Sustainalytics ESG (standardized)	1,888	-0.05	0.99	-1.30	-0.81	-0.12	0.61	1.18
Firm size	1,830	8.15	1.72	5.96	6.88	8.04	9.37	10.53
Market-to-book	1,830	1.40	1.80	0.15	0.42	0.82	1.70	3.09
Leverage	1,830	0.32	0.27	0.03	0.11	0.29	0.45	0.63
Tangibility	1,830	0.24	0.24	0.01	0.05	0.15	0.36	0.64
ROA	1,830	0.00	0.06	-0.07	0.00	0.01	0.03	0.05
Beta	1,825	1.02	0.36	0.55	0.76	1.00	1.26	1.48
ΔShare AI workers (%)	1,339	0.07	0.36	0.00	0.00	0.00	0.05	0.17
Task AI exposure	1,356	0.35	0.06	0.27	0.32	0.36	0.40	0.43
Data assets	1,339	0.01	0.01	0.00	0.01	0.01	0.01	0.02
GenAI	1,168	0.17	0.37	0.00	0.00	0.00	0.00	1.00
ΔSkew	1,331	-0.01	0.13	-0.13	-0.05	-0.01	0.03	0.10
Pre-event skew	1,397	0.07	0.11	-0.02	0.04	0.06	0.10	0.17
ΔSlopeD	1,730	0.00	0.30	-0.16	-0.05	-0.01	0.03	0.16
Pre-event SlopeD	1,730	0.41	0.37	0.16	0.21	0.30	0.50	0.84
Altman's removal CAR [-1, +2] (%)	1,884	-1.21	5.46	-6.26	-3.43	-1.18	0.89	3.50

Table 2: ChatGPT Announcement Returns and ESG Scores

This table reports coefficients from OLS regressions of firms' cumulative abnormal returns during the [-1, +10] trading days around the release of ChatGPT on November 30, 2022, and firms' most recent MSCI IVA ESG score as of October 31, 2022. The abnormal returns of each stock are calculated using the CAPM model estimated over trading days [-280, -31] and then cumulated over [-1, +10]. All columns include NAICS 3-digit industry fixed effects. ESG Q1-Q5 are indicator variables for the different quintiles of the ESG score. *GenAI* is an indicator variable equal to one if a firm is above the sample median on all three AI exposure measures: ΔShare AI workers, Workforce AI exposure, and Data assets. Other firm-level controls, including firm size, market-to-book, leverage, tangibility, ROA, and market beta, are reported in Appendix Table A2 and suppressed here for brevity. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	CAR [-1, +10] in %					
	(1)	(2)	(3)	(4)	(5)	(6)
ESG	1.168***		0.891***	0.802***	0.528	
	(3.618)		(2.986)	(2.786)	(1.611)	
ESG Q1		-2.289**				-0.451
		(-2.468)				(-0.571)
ESG Q2		-0.809				
		(-1.112)				
ESG Q4		0.391				
		(0.479)				
ESG Q5		0.695				
		(0.967)				
ESG × GenAI					1.984***	
					(2.844)	
ESG Q1 × GenAI						-4.789***
						(-3.492)
ΔShare AI workers			2.027**			
			(2.077)			
Workforce AI exposure			-2.753			
			(-0.314)			
Data assets			28.631			
			(0.344)			
GenAI				1.856***	-8.491**	2.491***
				(4.141)	(-2.363)	(4.291)
Observations	1,825	1,825	1,168	1,168	1,168	1,168
Adjusted R-squared	0.100	0.099	0.150	0.149	0.154	0.149
Controls	Yes	Yes	Yes	Yes	Yes	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 3: Robustness

This table summarizes the robustness tests for our main results. In panel A, we re-estimate our baseline specification in column (4) of Table 2 with different event windows and models for the cumulative abnormal returns. We also consider ESG scores from two alternative data providers, S&P Global and Sustainalytics, and follow the noise-correction procedure developed by Berg et al. (2021) to instrument ESG ratings with ratings from other providers. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. In panel B, we implement the recent event-study methodology developed by Cohn et al. (2024), which accounts for cross-sectional correlations in event returns and addresses concerns about false positives. We report the coefficient estimates and the corresponding p-values.

Panel A: Robustness

	ESG	
	Coeff.	(t-stat)
Alternative event windows for CAR		
CAR[-1, +1]	0.232	(1.227)
CAR[-1, +3]	0.383	(1.686)
CAR[-1, +5]	0.519	(2.094)
CAR[-1, +15]	0.754	(2.712)
Alternative models for abnormal returns		
Market-adjusted CAR[-1, +10]	0.499	(1.845)
FFC4-factor CAR[-1, +10]	0.679	(2.493)
Altnerative ESG data		
S&P Global (standardized)	0.640	(2.489)
S&P Global (instrumented by MSCI)	2.726	(2.698)
Sustainalytics (standardized)	-0.386	(-1.089)
Sustainalytics (instrumented by MSCI)	-3.063	(-2.317)
Averaging the three standardized scores	1.140	(2.422)

Panel B: Alternative event study method as in Cohn et al. (2024)

	Coeff.	p-value	
		CDF	Param.
ESG	1.403	0.010	0.008
GenAI	1.190	0.125	0.109
Firm size	-0.439	0.415	0.408
Market-to-book	0.058	0.815	0.826
Leverage	2.907	0.185	0.194
Tangibility	-1.715	0.385	0.488
ROA	-4.302	0.855	0.788
Beta	-0.469	1.000	0.998
Observations		1,168	

Table 4: ChatGPT Announcement Returns and E/S/G Pillars

This table reports coefficients of OLS regressions of firms' cumulative abnormal returns during the [-1, +10] trading days around the release of ChatGPT on November 30, 2022, and firms' most recent MSCI IVA Environmental, Social, and Governance pillar scores as of October 31, 2022. *GenAI* is an indicator variable equal to one if a firm is above the sample median on all three AI exposure measures. Other firm-level controls, including firm size, market-to-book, leverage, tangibility, ROA, and market beta, are suppressed here for brevity. All regressions include firm controls and NAICS 3-digit industry fixed effects. *t*-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	CAR [-1, +10] in %				
	(1)	(2)	(3)	(4)	(5)
Soc	0.361**	-0.025			-0.082
	(2.173)	(-0.118)			(-0.381)
Social × GenAI		1.018**			0.949**
		(2.158)			(2.015)
Env			0.284		0.308
			(1.455)		(1.502)
Env × GenAI			0.299		0.187
			(1.028)		(0.649)
Gov				0.093	0.087
				(0.375)	(0.344)
Gov × GenAI				0.057	0.245
				(0.161)	(0.623)
GenAI		-2.776	0.097	1.703	-5.259
		(-1.356)	(0.059)	(0.796)	(-1.374)
Observations	1,825	1,168	1,168	1,168	1,168
Adjusted R-squared	0.094	0.148	0.151	0.144	0.152
Controls	Yes	Yes	Yes	Yes	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes

Table 5: ChatGPT Announcement Returns and Specific ESG Criteria

This table reports coefficients of OLS regressions of firms' cumulative abnormal returns during the [-1, +10] trading days around the release of ChatGPT on November 30, 2022, and firms' most recent MSCI IVA scores for different sub-scores ('criteria') as of October 31, 2022. We re-estimate our baseline specification, substituting the average ESG scores with individual ESG sub-scores. Given the significance of missing data for many individual criteria, we only report the results for sub-scores available for at least one-third of our baseline sample. For brevity, we only present the main coefficients of interest. All specifications control for firm characteristics and NAICS 3-digit industry fixed effects. *t*-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

Dep. variable	Coef.	(*t*-stat.)	Obs.
Carbon emissions	0.112	(0.586)	1,168
Energy efficiency exposure	0.047	(0.343)	1,168
Water stress	0.176	(1.092)	1,120
Human capital development	**0.288****	**(2.221)**	**1,153**
Human capital development exposure	-0.235	(-1.126)	1,153
Human capital development management	**0.269***	**(1.778)**	**1,153**
Privacy data security	**0.761*****	**(4.626)**	**576**
Privacy data security exposure	**-1.169*****	**(-2.840)**	**576**
Privacy data security management	**0.426****	**(2.250)**	**572**
Corporate governance	0.099	(0.571)	1,168
Corruption institutional	**-0.170***	**(-1.782)**	**1,153**
Corporate behavior	-0.077	(-0.423)	1,168
Corporate behavior ethics	-0.067	(-0.366)	1,168
Tax transparency	-0.329	(-1.525)	1,165
Corporate behavior tax transparency	-0.104	(-0.720)	1,167

Table 6: ESG Scores and Analyst Forecast Revisions

This table reports coefficients from OLS regressions of standardized unexpected forecasts (SUF) around the release of ChatGPT on firms' most recent MSCI IVA ESG scores. SUF is calculated as the change in consensus forecast divided by the standard deviation of forecasts. We examine one-year in columns (1)-(3), five-year in column (4), and long-term earnings forecasts in column (5). All regressions control for firm characteristics (firm size, market-to-book, leverage, tangibility, ROA, and market beta, are suppressed here for brevity) and NAICS 3-digit industry fixed effects. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	Analyst EPS Forecast Revisions				
		1Y		5Y	Long-Term
	(1)	(2)	(3)	(4)	(5)
ESG	-0.008				
	(-0.220)				
Soc		-0.002	-0.010	0.023	-0.017
		(-0.147)	(-0.498)	(0.763)	(-0.695)
Soc × GenAI			-0.026	-0.046	-0.012
			(-0.549)	(-1.126)	(-0.510)
GenAI			0.109	0.189	0.108
			(0.580)	(0.823)	(0.848)
Observations	1,697	1,697	1,115	648	699
Adjusted R-squared	0.040	0.040	0.039	0.107	0.028
Controls	Yes	Yes	Yes	Yes	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes

Table 7: ESG Scores and Changes in Options-Implied Tail Risk

This table reports coefficients from OLS regressions of changes in firms' skewness measures around the release of ChatGPT, measured between November and December 2022, on firms' most recent MSCI IVA ESG scores, as of October 31, 2022. Columns 1–3 use changes in realized skewness ($\Delta Skew$) as the dependent variable. Columns 4–6 use changes in the IV slope ($\Delta SlopeD$), defined as the difference between out-of-the-money put IV and at-the-money IV, capturing the market's pricing of left-tail risk. $GenAI$ is an indicator variable equal to one if a firm is above the sample median on all three AI exposure measures. All regressions control for the pre-event skewness or IV slope, other firm characteristics (firm size, market-to-book, leverage, tangibility, ROA, and market beta, are suppressed here for brevity), and NAICS 3-digit industry fixed effects. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	ΔSkew			ΔSlopeD		
	(1)	(2)	(3)	(4)	(5)	(6)
ESG	0.006**			-0.014*		
	(2.336)			(-1.829)		
Soc		0.003***	0.003**		-0.007**	-0.005*
		(3.193)	(2.446)		(-2.017)	(-1.818)
Soc × GenAI			0.004			0.003
			(0.769)			(0.304)
GenAI			-0.029			-0.005
			(-0.999)			(-0.128)
Observations	1,316	1,316	956	1,536	1,536	1,040
Adjusted R-squared	0.489	0.489	0.522	0.019	0.018	0.009
Controls	Yes	Yes	Yes	Yes	Yes	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: Risk-Adjusted Portfolio Alphas

This table reports daily alphas (in %) from time-series regressions of long-short portfolio returns on standard risk factors. Each portfolio goes long the bottom ESG quintile and short the top ESG quintile, formed using MSCI IVA scores available by October 31, 2022. The first row restricts the sort to firms with high AI exposure ($GenAI = 1$); the second to firms with low AI exposure ($GenAI = 0$). FF3 denotes the Fama and French (1993) three-factor model; FFC4 adds momentum (Carhart, 1997); FF5 is the Fama and French (2015) five-factor model; FF5+M adds momentum to the five-factor model. Portfolios are equal-weighted and held from January 2023 through December 2024 without rebalancing. t-statistics based on Newey–West standard errors (five lags) are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	Daily Portfolio Alpha (%)			
	FF3	FFC4	FF5	FF5+M
GenAI = 1: ESG Q1 - Q5	0.087**	0.086**	0.082**	0.081**
	(2.296)	(2.278)	(2.130)	(2.118)
GenAI = 0: ESG Q1 - Q5	0.009	0.009	0.007	0.007
	(0.474)	(0.443)	(0.398)	(0.374)

Table 9: AI Concerns and ESG Scores: Sam Altman Removal

This table reports coefficients from OLS regressions of cumulative abnormal returns during the [-1, +2] trading days around the announcement of Sam Altman's unexpected removal from OpenAI on November 17, 2023, on firms' most recent MSCI IVA ESG score as of December 31, 2022. Returns are estimated using the CAPM model over trading days [-280, -31]. All columns include firm controls (firm size, market-to-book, leverage, tangibility, ROA, and market beta, are suppressed here for brevity) and NAICS 3-digit industry fixed effects. ESG Q1-Q5 are indicator variables for the different quintiles of the ESG score. *GenAI* is an indicator variable equal to one if a firm is above the sample median on all three AI exposure measures. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

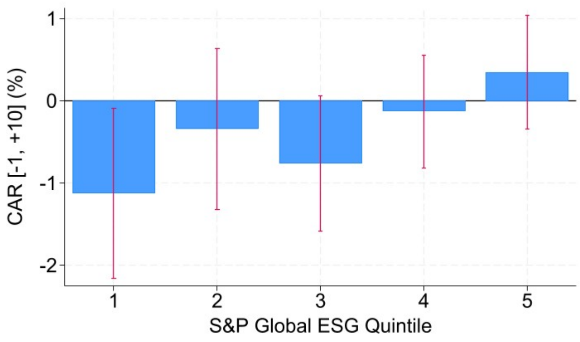
	Sam Altman's Removal CAR [-1, +2] in %			
	(1)	(2)	(3)	(4)
ESG	0.489**		0.458**	0.329*
	(2.359)		(2.383)	(1.810)
ESG Q1		-0.810*		
		(-1.831)		
ESG Q2		-0.567*		
		(-1.722)		
ESG Q4		-0.230		
		(-0.789)		
ESG Q5		0.395		
		(1.219)		
ESG × GenAI				0.839**
				(2.312)
GenAI			-0.407	-4.779**
			(-0.925)	(-2.253)
Observations	1,755	1,755	1,138	1,138
Adjusted R-squared	0.058	0.056	0.093	0.096
Controls	Yes	Yes	Yes	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes

Internet Appendix for

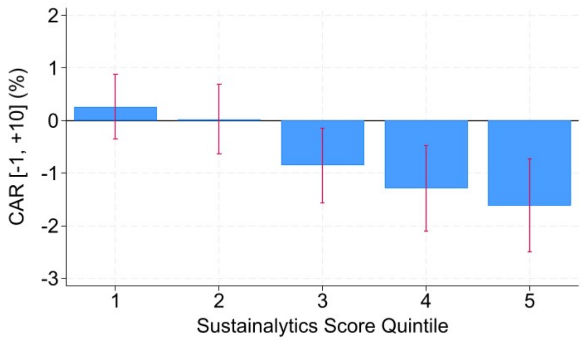
"The Social Risks of Generative AI"

Figure A1: Alternative ESG Score and ChatGPT Release CAR

This figure shows the average announcement CAR[-1, +10] around the launch of ChatGPT on November 30, 2022, sorted by quintiles of ESG scores available by October 31, 2022. We use the standardized ESG scores from S&P Global in panel (a) and from Sustainalytics in panel (b). We plot the 95% confidence intervals for each quintile.



(a)



(b)

Table A1: ESG Scores and Future Incident Likelihood

This table examines whether ESG ratings predict future cybersecurity incidents. The dependent variable in columns (1)–(3) is an indicator equal to one if a firm experienced at least one cyber attack or data breach during 2023–2024, as recorded in the Audit Analytics Cybersecurity database. Column (4) uses the total number of cybersecurity incidents over the same period as the dependent variable. The main explanatory variable is the firm's most recent MSCI IVA ESG score as of October 31, 2022. Columns (1) and (2) report OLS linear probability model estimates; column (3) reports logit marginal effects; column (4) reports Poisson regression estimates. *Firm size* is the natural logarithm of total assets; *Market-to-book* is calculated as (PRCC_F*CSHPRI+ DLC+DLTT+PSTKL - TXDITC)/AT; *Leverage* is calculated as (DLTT + DLC)/AT; *Tangibility* is the firm's tangible assets relative to total assets ratio, calculated as PPENT/AT; *ROA* is return-on-asset, calculated as OIADP/AT; *Beta* is the firm's market beta estimated using the CAPM model over trading days [-280, -31] relative to November 30, 2022. All specifications include NAICS 3-digit industry fixed effects. The reduction in observations in column (3) is due to the logit model dropping industries with no within-group variation in the outcome. t-statistics (columns 1–2) and z-statistics (columns 3–4), based on robust standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	Cyberattack			Incidents
Model	OLS		Logit	Poisson
	(1)	(2)	(3)	(4)
ESG	-0.009*	-0.014**	-0.016*	-0.020**
	(-1.866)	(-2.198)	(-1.950)	(-2.101)
Data assets		0.446	0.245	-0.660
		(0.284)	(0.133)	(-0.339)
Firm size	0.027***	0.030***	0.036***	0.045***
	(4.917)	(4.490)	(6.272)	(5.884)
Market-to-book	0.002	0.001	0.001	0.006
	(1.021)	(0.444)	(0.201)	(1.496)
Leverage	0.038**	0.061**	0.085***	0.095***
	(2.031)	(2.078)	(2.941)	(3.069)
Tangibility	-0.051	-0.052	-0.083	-0.076
	(-1.479)	(-1.087)	(-1.403)	(-1.297)
ROA	-0.209**	-0.188*	-0.298**	-0.349*
	(-2.220)	(-1.843)	(-2.323)	(-1.921)
Beta	0.003	0.011	0.018	0.017
	(0.159)	(0.502)	(0.626)	(0.503)
Observations	1,825	1,338	993	1,338
(Pseudo) Adjusted R-squared	0.076	0.090	0.179	0.657
Industry FE	Yes	Yes	Yes	Yes

Table A2: ChatGPT Announcement Returns and ESG Scores

Columns (1)-(6) of this table reports full coefficient estimates from the regressions reported in Table 2, including controls omitted for brevity in the main text. *Firm size* is the natural logarithm of total assets; *Market-to-book* is calculated as (PRCC_F*CSHPRI+ DLC+DLTT+PSTKL - TXDITC)/AT; *Leverage* is calculated as (DLTT + DLC)/AT; *Tangibility* is the firm's tangible assets relative to total assets ratio, calculated as PPENT/AT; *ROA* is return-on-asset, calculated as OIADP/AT; *Beta* is the firm's market beta estimated using the CAPM model over trading days [-280, -31] relative to November 30, 2022. All other firm characteristics are obtained from Compustat for the fiscal quarter prior to November 2022. Columns (7) and (8) re-estimate the specifications in columns (1) and (2), respectively, restricting the sample to firms for which AI exposure data are available.

	CAR [-1, +10] in %							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ESG	1.168***		0.891***	0.802***	0.528		0.896***	
	(3.618)		(2.986)	(2.786)	(1.611)		(3.009)	
ESG Q1		-2.289**				-0.451		-1.657**
		(-2.468)				(-0.571)		(-2.326)
ESG Q2		-0.809						-1.295**
		(-1.112)						(-2.236)
ESG Q4		0.391						-0.765
		(0.479)						(-1.097)
ESG Q5		0.695						0.682
		(0.967)						(1.124)
ESG × GenAI					1.984***			
					(2.844)			
ESG Q1 × GenAI						-4.789***		
						(-3.492)		
ΔShare AI workers			2.027**					
			(2.077)					
Workforce AI exposure			-2.753					
			(-0.314)					
Data assets			28.631					
			(0.344)					
GenAI				1.856***	-8.491**	2.491***		
				(4.141)	(-2.363)	(4.291)		
Firm size	-0.362**	-0.335*	-0.039	-0.022	-0.033	0.090	0.019	0.066
	(-2.077)	(-1.910)	(-0.166)	(-0.094)	(-0.138)	(0.399)	(0.087)	(0.296)
Market-to-book	-0.420***	-0.409***	-0.462***	-0.403***	-0.430***	-0.375***	-0.393***	-0.393***
	(-5.034)	(-4.754)	(-4.189)	(-3.448)	(-3.594)	(-3.010)	(-3.445)	(-3.326)
Leverage	2.259**	2.190**	2.908***	2.810**	2.881**	2.763**	2.820**	2.791**
	(2.193)	(2.134)	(2.743)	(2.435)	(2.485)	(2.415)	(2.379)	(2.365)
Tangibility	-1.612	-1.595	-0.000	0.398	0.416	0.461	-0.209	-0.243
	(-0.910)	(-0.912)	(-0.000)	(0.175)	(0.179)	(0.196)	(-0.092)	(-0.109)
ROA	-13.342***	-13.301***	-10.780	-11.477	-12.378	-11.694	-11.616	-11.862
	(-3.544)	(-3.530)	(-1.397)	(-1.440)	(-1.538)	(-1.388)	(-1.487)	(-1.509)
Beta	-0.974	-1.048	-1.630	-1.515	-1.492	-1.604	-1.342	-1.352
	(-1.072)	(-1.170)	(-1.556)	(-1.489)	(-1.499)	(-1.526)	(-1.377)	(-1.358)
Observations	1,825	1,825	1,168	1,168	1,168	1,168	1,168	1,168
Adjusted R-squared	0.100	0.099	0.150	0.149	0.154	0.149	0.154	0.149
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table A3: ESG Scores and AI Exposure Measures

This table reports the relation between firms' ESG scores and their AI exposure and investments. We use firms' most recent MSCI IVA ESG score as of October 31, 2022. Δ*Share AI workers* is the change in the share of a firm's AI-related employees from 2010 to 2021 following Babina et al. (2024). *Workforce AI exposure* measures workforce exposure to generative AI, *Data assets* captures the firm's stock of data that could be deployed with AI technologies, both are from Eisfeldt et al. (2025). All specifications include NAICS 3-digit industry fixed effects. Column (5) further controls for other firm characteristics from Compustat, measured in the fiscal quarter preceding November 2022. The controls (omitted for brevity) include firm size, market-to-book, leverage, asset tangibility, ROA, and market beta. t-statistics, based on standard errors clustered by NAICS 3-digit industry, are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

	MSCI ESG				
	(1)	(2)	(3)	(4)	(5)
ΔShare AI workers	0.076			0.016	-0.043
	(1.409)			(0.246)	(-0.747)
Workforce AI exposure		0.106		-1.108	0.005
		(0.139)		(-1.110)	(0.006)
Data assets			11.906	17.796	7.601
			(1.011)	(1.117)	(0.538)
Observations	1,339	1,356	1,339	1,168	1,168
Adjusted R-squared	0.261	0.292	0.296	0.269	0.372
Controls	No	No	No	No	Yes
NAICS3-Industry FE	Yes	Yes	Yes	Yes	Yes