



UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS
DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 17

Does the Language of Artificial Intelligence Move Markets?

*A Semantic Analysis of S&P 500 SEC Filings and the Cross-Section of Stock
Returns, 2010–2026*

Manel Kammoun Djerrycharli Tandjambianda Simon-Pierre Boucher

Département des sciences administratives
Université du Québec en Outaouais

manel.kammoun@uqo.ca · djerrycharli.tandjambianda@uqo.ca · simon-pierre.boucher@uqo.ca

Gatineau – Pavillon Alexandre-Taché
283, boulevard Alexandre-Taché
Gatineau, Québec, Canada J9A 1L8

June 2026

Version 1.0

Abstract

We construct a *semantic* measure of artificial-intelligence (AI) engagement for every S&P 500 firm by embedding the text of its annual SEC filings (Form 10-K) with a sentence transformer and scoring each passage by cosine similarity to professionally curated AI “anchor” concepts—internal AI adoption, AI infrastructure exposure, and AI product-disruption risk. Linking these scores to monthly equity prices from Yahoo Finance and to the Fama–French factor data, we assemble a panel of 7869 firm-years (500 firms, fiscal years 2010–2026) and ask whether the intensity of AI language predicts the cross-section of subsequent returns. Our central result is **conditional and economically large**. Over the full sample the relation between AI-disclosure intensity and future returns is positive but modest and only marginally significant; a risk-adjusted long–short portfolio earns an insignificant $\sim 2\%$ /year. After the public launch of ChatGPT (30 November 2022), however, the relation *more than sextuples*: the spread between the top and bottom quintile of AI disclosure rises from +2.4 percentage points per year before 2022 to +14.6 points per year afterward. The break survives section-weighting choices, alternative signal definitions and horizons, dimension reduction, and—crucially—the exclusion of the “Magnificent 7” mega-caps. A battery of identification tests sharpens the interpretation: (i) a *placebo* across five non-AI semantic factors shows the post-2022 jump is specific to AI-related language (ESG and climate disclosure, by contrast, earn *negative* post-2022 spreads); (ii) a market-model *event study* finds that firms with higher pre-event AI disclosure earned +10.1% higher cumulative abnormal returns over the year after ChatGPT ($t=1.97$); and (iii) the effect is absorbed by firm fixed effects and roughly halved by sector fixed effects, indicating a *cross-sectional, regime-dependent re-rating*—the market repriced *which firms and sectors* are AI-exposed once AI became salient, rather than rewarding year-to-year changes in any single firm’s wording. The findings provide direct, text-based evidence for an attention- and narrative-driven repricing of AI exposure, and illustrate the value of embedding-based disclosure measurement for asset pricing.

Keywords: textual analysis, sentence embeddings, artificial intelligence, disclosure, cross-section of returns, investor attention, event study, generative AI

JEL Classification: G12, G14, G17, M41, O33

1. Introduction

The arrival of generative artificial intelligence has been the defining corporate-narrative shift of the 2020s. In the span of barely three years following the public launch of ChatGPT in late 2022, the market capitalization of AI-linked firms swelled by trillions of dollars, “AI” vaulted past every other theme in frequency on earnings calls, and the language of corporate disclosure filled with references to large language models, generative tools, intelligent automation, and the compute infrastructure that powers them. A technology that had been the preserve of research labs and back-office analytics suddenly became the organizing story of public equity markets. This paper asks a question that is simple to state but surprisingly demanding to answer well: did the *language* of AI in firms’ mandatory disclosures carry information about their subsequent stock returns—and if so, when did that information begin to be priced, and through what economic channel?

Answering the question convincingly requires three ingredients that have only recently become jointly feasible. The first is a large, machine-readable corpus of corporate disclosure. The second is a measurement technology that captures the *meaning* of text rather than the mere frequency of keywords—one that recognizes that a firm describing how it “deploys machine-learning models to optimize its supply chain” is talking about AI even when the three-letter acronym never appears. The third is a clean, look-ahead-free mapping from disclosure to tradeable returns. We bring all three to bear. Our laboratory is the universe of S&P 500 constituents over fiscal years 2010–2026. Our raw material is the full text of their annual reports (Form 10-K), parsed section by section. Our measurement engine is the EDGAR Research Database, which embeds every text chunk with the all-MiniLM-L6-v2 sentence transformer and scores it by cosine similarity to a set of carefully written “anchor” passages that define eight economic concepts. Three of these anchors isolate distinct facets of AI engagement, each corresponding to a different way the technology bears on firm value: *internal AI adoption* (using AI to run the business more productively), *AI infrastructure exposure* (capital and revenue tied to the compute build-out), and *AI product-disruption risk* (the danger that AI commoditizes or displaces the firm’s own offerings). The result is a continuous, comparable, firm-year measure of how intensively—and in what sense—each firm engages with AI in its own words, available for the entire index and spanning a decade and a half.

We pair these semantic scores with monthly adjusted prices for all 500 tickers from Yahoo Finance and with the Fama–French five factors plus momentum. To avoid look-ahead bias we adopt the standard portfolio-formation convention. Because a 10-K covering fiscal year Y is filed and public by early in year $Y+1$, we form signals at the end of June of $Y+1$ and measure the forward twelve-month return from July of $Y+1$ to June of $Y+2$. The roughly three-month cushion between the typical filing date and portfolio formation guarantees that the signal is in the public information set before the holding period begins, even for late filers. This yields 6,852 firm-years with a strictly

tradeable forward return, spanning formation years 2010 through 2024.

Our headline finding is that the pricing of AI language is *not* a stable factor premium but a regime-dependent re-rating that switched on with the generative-AI boom. For most of the 2010s, sorting S&P 500 firms on the intensity of their AI disclosure produced only a faint return pattern: the equal-weighted spread between the highest and lowest quintiles averaged about +2.4 percentage points per year—economically modest and statistically fragile—and a risk-adjusted long–short portfolio earned an alpha that was both small and statistically insignificant. The information was there in the filings, but the market was not paying for it. After the launch of ChatGPT in November 2022, the very same sort produced a spread of +14.6 percentage points per year, a more than sixfold jump. This is not a subtle tilt buried in a long-run average; it is a sharp, datable change in how the cross-section prices a fixed characteristic. The re-rating is visible from three independent angles: it appears in every post-2022 formation cohort taken on its own, in a cumulative long–short wealth curve that drifts sideways for a decade and then inflects sharply upward at the end of 2022, and in a year-by-year Fama–MacBeth slope that hovers near zero through 2021 before turning large and positive.

A natural worry is that this pattern simply reflects the well-known rally of a handful of technology mega-caps—that we have rediscovered the Magnificent 7 dressed in textual clothing. We confront this concern head-on and reject it. The post-2022 spread is +14.1 percentage points ($t=1.9$) when the “Magnificent 7” are excluded and +14.5 points when the ten largest firms by market capitalization are dropped: removing the very names that dominate the headline rally barely moves the estimate, so the effect is broad-based across the index rather than an artifact of Nvidia, Microsoft, or their peers. It is equally robust to how the signal is built and measured—to using the simple mean, the section-weighted mean, or the top-5% tail of similarity scores; to six-, twelve-, and twenty-four-month holding horizons; and to replacing the three correlated AI signals with their first principal component, a single index that alone explains 87% of their common variation.

Establishing that the spread is large and well-timed is only half the task; we then deploy three complementary tests to pin down what the result is—and, equally important, what it is not. The first is a *placebo* exercise that re-runs the entire research design on the five non-AI semantic factors in the database (ESG, climate, cybersecurity, automation/labor, and macro risk). The logic is simple: if the post-2022 jump merely reflected a generic “more disclosure earns more” phenomenon, or a mechanical artifact of our embedding-and-sorting pipeline, every factor should display the same break. They do not. The post-2022 spread is large and positive for the AI factors, essentially zero for macro-risk language, and—most tellingly—*negative* for ESG and climate disclosure, meaning high-ESG firms *underperformed* after 2022. The post-ChatGPT re-rating is specific to AI-related content, not a property of disclosure intensity in general. The second test is a market-model *event study* around the 30 November 2022 ChatGPT launch. Firms in the top tercile of pre-event AI

disclosure—measured from their FY2021 10-K, the last report filed before anyone could have anticipated the event—earned cumulative abnormal returns +10.1 percentage points higher than the bottom tercile over the following year ($t=1.97$) and +14.6 points higher over eighteen months. Because abnormal returns are purged of each firm’s market-beta exposure, this is a clean cross-sectional response to an unanticipated narrative shock rather than a reflection of high-AI firms simply being higher-beta in a recovering market. The third test uses two-way fixed effects and difference-in-differences to localize the mechanism within the cross-section. The premium is fully absorbed by firm fixed effects and roughly halved by sector fixed effects, which tells us that the durable, priced signal is the persistent *level* of a firm’s (and its sector’s) AI exposure relative to peers, not the year-to-year wiggle in its wording.

Taken together, the four pieces of evidence—the datable break, the AI-specific placebo, the event-study timing, and the fixed-effects anatomy—converge on a single interpretation: an attention- and narrative-driven repricing. The textual evidence of firms’ AI engagement accumulated quietly in their filings for a decade and was barely priced. A salient catalyst—the consumer breakthrough of generative AI—then drew investor attention to that pre-existing information, and capital flowed toward the firms and sectors whose disclosures had already revealed the greatest AI exposure. What looks like new information was in large part old information newly attended to. This is the asset-pricing signature of limited attention and narrative economics, observed here not through an indirect sentiment proxy but directly in the semantic content of mandatory disclosure.

Contribution. We make four contributions. *First*, methodologically, we demonstrate the value of embedding-based semantic measurement of disclosure for asset pricing: a cosine-similarity-to-anchor design captures economically meaningful, multi-faceted exposure (adoption vs. infrastructure vs. risk) that keyword counts cannot, and it does so at scale across the entire S&P 500 and a decade and a half of filings. *Second*, substantively, we provide text-based, look-ahead-free evidence that the cross-sectional pricing of AI exposure is a post-2022 regime phenomenon rather than a long-standing premium. *Third*, we discipline the finding with a placebo across non-AI factors and an event study, distinguishing a genuine AI-specific repricing from a mechanical or market-wide artifact. *Fourth*, we document the *anatomy* of the effect—its concentration in smaller firms, its operation substantially through a sector channel, and its disappearance under firm fixed effects—which sharpens the economic interpretation and warns against naive trading strategies.

The remainder of the paper is organized as follows. Section 2 situates the study in five literatures and states our hypotheses. Section 3 describes the data and the construction of the semantic measures. Section 4 lays out the empirical methodology. Section 5 presents the main results, and Section 6 the robustness and identification tests. Section 7 discusses mechanisms and limitations, and Section 8 concludes.

2. Related Literature and Hypotheses

Our study sits at the intersection of five strands of research: textual analysis in finance, the pricing of corporate disclosure, technology and innovation in asset pricing, investor attention and narrative economics, and the nascent literature on the economics of artificial intelligence. Each contributes a building block—a measurement tool, an object of study, a candidate risk factor, a behavioral mechanism, and a comparison case—and our contribution is to combine them: we measure a technology theme from firms’ own words, price it in the cross-section, and show that its pricing turns on precisely when a behavioral mechanism predicts it should. We take the strands in turn and close with the hypotheses they jointly motivate.

2.1. Textual analysis and natural language processing in finance

A now-large literature extracts economic signals from financial text. The foundational work of [Tetlock \(2007\)](#) showed that the tone of a *Wall Street Journal* column forecasts aggregate returns, and [Tetlock et al. \(2008\)](#) extended this to firm-level news. [Loughran and McDonald \(2011\)](#) demonstrated that finance-specific sentiment dictionaries materially outperform general-purpose word lists in 10-K analysis, launching a wave of dictionary-based studies and the widely used Loughran–McDonald word lists; [Loughran and McDonald \(2016\)](#) survey this field. A second generation moved from counting words to modeling them: [Gentzkow et al. \(2019\)](#) review text-as-data methods in economics, while [Kelly et al. \(2021\)](#) and [Bybee et al. \(2024\)](#) use machine-learning representations of news to construct interpretable factors and a “state of the economy” from text. Our paper belongs to the most recent, embedding-based generation of this work. The limitations of keyword counting are acute for a theme like AI: a naive count cannot tell “we do not currently use artificial intelligence” apart from “we are an artificial-intelligence leader,” it misses the paraphrase of a firm that describes machine learning without ever writing “AI,” and it treats boilerplate and substance identically. We instead measure the *semantic similarity* of disclosure to anchor concepts using a sentence transformer ([Reimers and Gurevych, 2019](#)), which maps passages into a vector space where meaning, not surface form, governs proximity. This follows a growing practice of using contextual embeddings to quantify disclosure content ([Cong et al., 2022](#)), and is closest in spirit to similarity-to-concept designs that score documents against curated reference passages.

2.2. The pricing of corporate disclosure

A parallel literature asks whether the *content* and *changes* of mandatory disclosure predict returns. [Cohen et al. \(2020\)](#) show that firms that change the language of their 10-K relative to the prior year (“lazy prices” when they do not) experience predictable future returns and operating outcomes.

Hoberg and Phillips (2010) and Hoberg and Phillips (2016) use 10-K product descriptions to build text-based industry classifications and measures of product-market fluidity that carry asset-pricing content. Lopez-Lira (2023) extracts risk exposures from 10-K risk-factor sections. The general lesson is that the qualitative portions of filings contain value-relevant information that markets process incompletely and with delay. We contribute a focused, concept-level measure—AI engagement—and show that its pricing is strongly state-dependent, a dimension the disclosure-pricing literature has largely left implicit.

2.3. Technology, innovation, and the cross-section of returns

Asset pricing has long sought to connect technological change to risk and return. Hirshleifer et al. (2013) show that innovative efficiency predicts returns; Kogan et al. (2017) build market-based measures of innovation from patents and link them to growth and reallocation; Kogan and Papanikolaou (2014) show that investment-specific technology shocks shape the cross-section. A more recent literature studies AI exposure directly: Babina et al. (2024) measure firm-level AI investment from job-posting and resume data and document faster growth among AI-adopting firms; Eisfeldt et al. (2023) construct a firm-level measure of *generative*-AI exposure from the task content of occupations and show that a generative-AI factor was priced precisely in the weeks around ChatGPT’s release. Acemoglu and Restrepo (2020) and Webb (2020) study the labor-market exposure of firms to AI and automation. Our contribution relative to this work is to measure AI exposure from the firm’s *own disclosure*—a self-reported, narrative signal available for the entire index at annual frequency—and to characterize the full time profile of its pricing across a decade and a half, not only the event window. Where Eisfeldt et al. (2023) use occupation-task exposure, we use the language firms choose, providing a complementary and highly granular measure.

2.4. Investor attention, salience, and narrative economics

Why might value-relevant disclosure be priced only after a catalyst? The behavioral literature offers a clear mechanism: limited investor attention. Barber and Odean (2008) show that retail investors are drawn to attention-grabbing stocks; Da et al. (2011) construct a direct attention proxy from Google search and link it to prices; DellaVigna and Pollet (2009) document under-reaction to information released when attention is low; and Hirshleifer et al. (2009) show that earnings news is underweighted when investors are distracted. Shiller (2017) argues that contagious narratives drive major asset-price movements, and Baker and Wurgler (2006) show that sentiment differentially affects hard-to-value, speculative stocks. Our results provide a sharp illustration: the information about firms’ AI engagement existed in filings for years but was repriced abruptly when a salient consumer event—ChatGPT—focused investor attention on AI. The asymmetry we document (AI

rewarded, ESG penalized after 2022) is itself consistent with a shift in the prevailing market narrative.

2.5. ESG and climate disclosure as a comparison case

Finally, our placebo design connects to the literature on the pricing of ESG and climate disclosure. [Pástor et al. \(2021\)](#) model how sustainability preferences depress the expected returns of green assets in equilibrium, and [Pástor et al. \(2022\)](#) show that green stocks' strong *realized* returns through 2021 reflected unexpected increases in climate concern rather than high expected returns. [Bolton and Kacperczyk \(2021\)](#) document a carbon premium. Consistent with the post-2021 reversal in this literature, we find that high-ESG and high-climate disclosure earned *negative* spreads after 2022, exactly when AI disclosure began to earn large positive ones—underscoring that the post-2022 re-rating reflects a thematic rotation toward AI rather than a generic disclosure effect.

2.6. Hypotheses

Our analysis is organized around four hypotheses.

H1 (Unconditional pricing).

If AI exposure carries a stable risk premium or persistent mispricing, then sorting on AI-disclosure intensity yields a positive and significant risk-adjusted return spread over the full 2010–2026 sample.

H2 (Regime dependence).

If instead AI exposure is repriced only once AI becomes salient, the return spread is small before and large after the late-2022 generative-AI catalyst, with a detectable structural break.

H3 (AI-specificity).

If the post-2022 spread reflects a genuine AI repricing rather than a mechanical or market-wide artifact, it appears for AI-related semantic factors but not for unrelated factors (e.g. ESG, climate, macro risk).

H4 (Cross-sectional, not within-firm).

If the market repriced *which* firms are AI-exposed rather than rewarding incremental wording, the effect is concentrated in the cross-section and attenuates under firm and sector fixed effects.

The data, as we show, decisively favor H2–H4 over H1.

3. Data and Variable Construction

3.1. The semantic factor panel

The backbone of our study is the EDGAR Research Database, an infrastructure that downloads, parses, and semantically scores SEC filings for S&P 500 firms. For each filing the pipeline proceeds in four steps. It (i) extracts the canonical sections of the 10-K—Risk Factors, Management’s Discussion and Analysis (MD&A), legal proceedings, the financial statements, and so on—so that the location as well as the content of each passage is known; (ii) splits each section into overlapping text chunks of a few sentences, the overlap ensuring that a thought spanning a chunk boundary is not lost; (iii) embeds every chunk into a 384-dimensional vector with the all-MiniLM-L6-v2 sentence transformer, a compact and widely validated model that maps semantically similar passages to nearby points in the embedding space; and (iv) scores each chunk by its cosine similarity to a set of *anchor* passages that define eight economic concepts. The anchors are short, deliberately written paragraphs that articulate each concept in natural language. The AI-adoption anchor, for instance, describes a firm “internally using artificial intelligence, machine learning, generative AI, large language models, automation, predictive analytics . . . to improve productivity, operations, efficiency, decision-making.” Cosine similarity to such an anchor measures how closely a passage’s *meaning* matches the concept, on a scale from -1 to 1 , regardless of the exact words used—so that a sentence on “deploying neural networks to forecast demand” scores highly on AI adoption even though it shares no keyword with the anchor. This is precisely the paraphrase- and context-sensitivity that keyword counts lack.

Chunk-level scores are aggregated to the filing level using section weights that emphasize the parts of the document where the concept is most likely to be discussed substantively (e.g. MD&A and Risk Factors), and then to the firm-fiscal-year level. For each concept the database reports the mean, maximum, standard deviation, top-5% tail, and section-weighted mean (*wmean*) of chunk similarities, along with first and second differences, percentage changes, and three-year rolling statistics. The eight concepts comprise three AI facets—*AI adoption*, *AI infrastructure*, *AI disruption risk*—and five others used as placebos: *ESG commitment*, *climate risk*, *cybersecurity*, *automation/labor*, and *macro risk*.

The resulting panel covers all 500 current S&P 500 members across fiscal years 2010–2026, a total of 7,869 firm-years. We take the section-weighted mean (*wmean*) as our headline signal because it up-weights the forward-looking, strategy-rich sections; we verify robustness to the simple mean and the top-5% tail in Section 6. Our three primary signals are therefore *ai_adoption_wmean* (**AI Adopt**), *ai_infrastructure_wmean* (**AI Infra**), and *ai_disruption_risk_wmean* (**AI Risk**); we also form a standardized opportunity composite, $AI\ Compos = \frac{1}{2}[z(AI\ Adopt) + z(AI\ Infra)]$.

3.2. Price and factor data

We download monthly adjusted-close prices for all 500 tickers from Yahoo Finance over 2009–2026 (handling share-class tickers such as BRK.B and BF.B), the SPDR S&P 500 ETF (SPY) as the market benchmark, and the Fama–French five factors (Fama and French, 2015), the momentum factor, and the risk-free rate from the Kenneth R. French data library. Monthly simple returns are computed from adjusted prices, which incorporate dividends and splits. A market-capitalization proxy is formed as shares outstanding (from Yahoo Finance) times the historical price; we use it for value-weighting and as a size control, acknowledging the shares-held-constant approximation.

3.3. Timing and the construction of tradeable returns

The single most important design choice in a disclosure–return study is timing, because a signal that inadvertently peeks at information unavailable in real time can manufacture predictability that no investor could have traded on. A 10-K covering fiscal year Y becomes public only when filed, typically in February or March of year $Y+1$. To guarantee that our signal is in the public information set before any return is measured, we follow the Fama–French portfolio convention: signals from fiscal year Y are used to form portfolios at the end of June of $Y+1$, and we measure the forward twelve-month return from July of $Y+1$ through June of $Y+2$. The roughly three-month gap between the typical filing date and June 30 builds in a margin that rules out look-ahead bias even for late filers, while the July-to-June holding window is the standard one in the cross-sectional literature and so keeps our estimates comparable to it. After requiring a valid forward return with adequate monthly coverage, we obtain 6,852 firm-years spanning formation years 2010–2024, with return realizations extending through mid-2026.

3.4. Controls and standardization

For each formation year we compute cross-sectional z -scores of every signal, so that all coefficients are expressed per cross-sectional standard deviation and the powerful secular rise in AI language (documented below) is differenced out—in any given year we compare firms to their contemporaneous peers. We control for the three workhorse return determinants: *size* (log market-cap proxy), *momentum* (the firm’s own return over the twelve months ending at formation, skipping the last month), and *idiosyncratic volatility* (the annualized standard deviation of monthly returns over the trailing year). In regressions, returns and continuous controls are winsorized at the 1st and 99th percentiles to limit the influence of outliers; in portfolio sorts we use raw returns. Appendix A gives precise definitions.

3.5. Summary statistics

Table 1 reports summary statistics for the estimation sample. The average forward twelve-month return is 18.6% (reflecting the strong 2011–2026 bull market for U.S. large caps), with a wide cross-sectional dispersion (standard deviation 41.5%). The AI signals are small in level—cosine similarities of disclosure to narrow anchors—but exhibit meaningful cross-sectional variation. Table 2 reports correlations. Two features stand out. First, the three AI signals are highly collinear (pairwise ρ between 0.75 and 0.86): firms that write more about adopting AI also write more about AI infrastructure and AI risk. They are best understood as three lenses on a single latent construct—*AI salience*—a point we formalize with principal-components analysis in Section 6. Second, the raw (unconditional) correlation between each AI signal and the forward return is positive but small (≈ 0.04), foreshadowing that the unconditional relation is weak and that the action lies in its time variation.

Table 1. Summary Statistics

Variable	Mean	SD	p5	Median	p95
AI adoption (wmean)	0.040	0.028	-0.000	0.038	0.089
AI infrastructure (wmean)	-0.023	0.033	-0.073	-0.024	0.033
AI disruption risk (wmean)	0.026	0.030	-0.018	0.024	0.080
Forward 12m return	0.186	0.415	-0.279	0.142	0.735
Forward return, mkt-adj.	0.036	0.395	-0.401	-0.005	0.537
Momentum (12m)	0.178	0.337	-0.257	0.141	0.728
Volatility (ann.)	0.269	0.144	0.126	0.240	0.492
Log market cap	23.725	1.178	21.942	23.636	25.741

Notes: This table reports distributional statistics for the firm-year estimation panel, which comprises 6,852 firm-years (497 firms) over formation years 2010–2024 with a tradeable forward return. The three AI signals—*AI adoption*, *AI infrastructure*, and *AI disruption risk*—are section-weighted mean cosine similarities (wmean) of a firm’s 10-K text to the corresponding anchor concept. *Forward 12m return* is the buy-and-hold return from July of fiscal year $Y+1$ through June of $Y+2$; the market-adjusted variant nets out the contemporaneous return on the S&P 500 ETF (SPY). *Momentum* is the firm’s trailing 12-month own return, *volatility* the annualized standard deviation of its monthly returns, and *log market cap* the natural logarithm of the market-capitalization proxy. For each variable the table reports the mean, standard deviation (SD), and the 5th, 50th (median), and 95th percentiles.

Table 2. Correlations among Signals, Controls, and Returns

	(1)	(2)	(3)	(4)	(5)	(6)
(1) AI adoption	1.00	0.80	0.86	0.07	0.01	0.04
(2) AI infrastructure	0.80	1.00	0.74	0.11	0.02	0.04
(3) AI disruption risk	0.86	0.74	1.00	0.07	0.02	0.04
(4) Size	0.07	0.11	0.07	1.00	0.08	-0.10
(5) Momentum	0.01	0.02	0.02	0.08	1.00	0.00
(6) Forward return	0.04	0.04	0.04	-0.10	0.00	1.00

Notes: This table reports Pearson correlation coefficients among the three within-year standardized AI signals, the size and momentum controls, and the forward 12-month return, computed over the estimation panel. Columns (1)–(6) correspond to the row variables in the same order. The three AI signals are mutually highly correlated (pairwise ρ between 0.74 and 0.86), consistent with their being three lenses on a single latent “AI salience” construct, whereas each AI signal’s unconditional correlation with the forward return is small (≈ 0.04), foreshadowing that the action lies in the time variation of the relation rather than its unconditional level.

3.6. The secular rise of AI language

Before turning to returns, Figure 1 documents the raw phenomenon that motivates our within-year design. The cross-sectional mean similarity to all three AI concepts trended steadily upward through the 2010s and inflected sharply after 2022; by 2025–2026 AI language saturates S&P 500 disclosure. This is precisely why our tests rely on within-year cross-sectional ranking: the interesting question is not whether the index as a whole talks more about AI over time (it emphatically does), but whether, *in a given year*, firms that talk about AI more than their peers subsequently earn different returns. Figure 2 shows that AI disclosure is also unevenly distributed across sectors, concentrated in Information Technology, Communication Services, and Industrials—a fact that will matter for interpreting the sector channel.

3.7. Which firms disclose the most AI? Descriptive evidence

Table 3 lists the ten most AI-intensive S&P 500 firms by average AI-adoption score over the full sample. The list is instructive in two ways. First, it has strong face validity: it features enterprise-technology and software firms (Hewlett Packard Enterprise, Intel, IBM, Tyler Technologies, Autodesk), an algorithm-centric platform (Uber), and data-intensive financials (BNY Mellon, Berkshire Hathaway), exactly the kinds of firms one would expect to write substantively about deploying AI. Second, the list is *not* confined to the technology sector: industrials (Cummins) and energy (EOG Resources) appear, confirming that intensive AI disclosure cuts across the GICS taxonomy and that our continuous, text-based measure captures variation a sector dummy would miss.

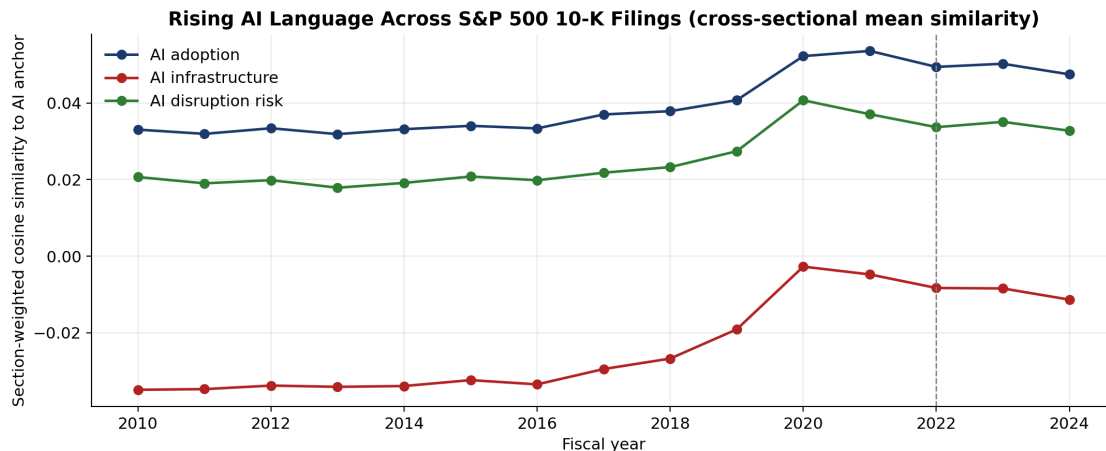


Figure 1. The secular rise of AI language in S&P 500 disclosure. The figure plots the cross-sectional mean of the section-weighted cosine similarity of 10-K text to each of the three AI anchors (adoption, infrastructure, and disruption risk), by fiscal year over 2010–2026. All three measures trend steadily upward through the 2010s and inflect sharply after the November 2022 launch of ChatGPT (dashed vertical line), by which point AI language saturates index disclosure. This secular trend motivates the within-year cross-sectional design used throughout the paper.

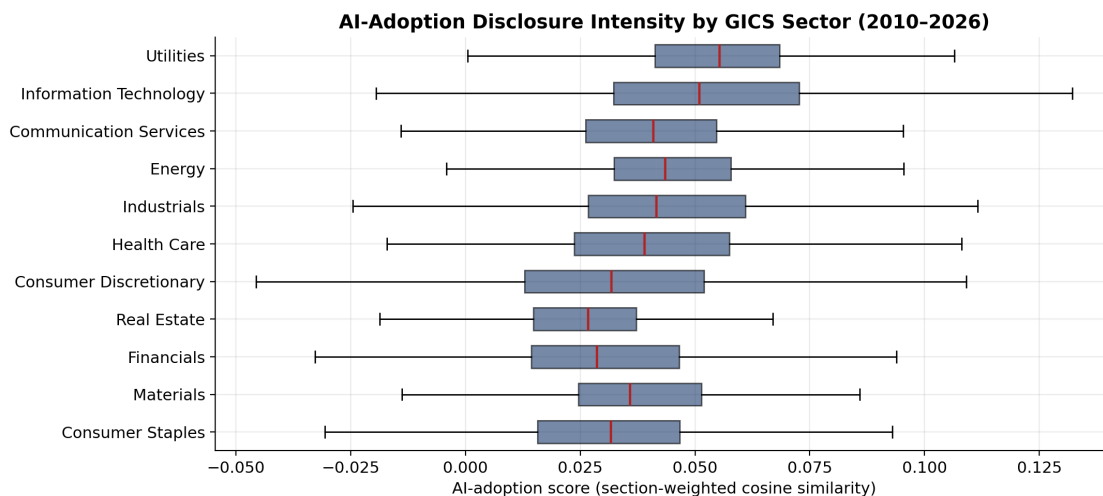


Figure 2. AI disclosure across GICS sectors. The figure shows box plots of the firm-level AI-adoption score within each GICS sector (boxes span the interquartile range, the red line marks the sector median, and outliers are suppressed for legibility). AI disclosure is concentrated in technology-intensive sectors—Information Technology, Communication Services, and Industrials—a concentration that motivates the sector-channel analysis and the sector-neutral robustness checks reported later.

A further descriptive fact is central to our interpretation: the AI signal is highly *persistent*. The year-over-year cross-sectional rank autocorrelation of the AI composite is 0.88, and even at a five-year horizon it remains 0.73. Firms that are AI-intensive disclosers stay AI-intensive; the measure captures a slow-moving structural characteristic of the firm rather than transient wording. This persistence has two implications we exploit later: it explains why firm fixed effects absorb most of the signal (there is little within-firm variation to identify), and it underpins the difference-in-differences design, in which we fix each firm’s exposure at its pre-event level precisely because that level is stable.

Table 3. The Most AI-Intensive S&P 500 Firms

Ticker	Company	Sector	Avg. AI-adoption
HPE	Hewlett Packard Enterprise	Information Technology	0.126
UBER	Uber	Industrials	0.123
INTC	Intel	Information Technology	0.121
IBM	IBM	Information Technology	0.113
TYL	Tyler Technologies	Information Technology	0.108
BNY	BNY Mellon	Financials	0.106
BRK.B	Berkshire Hathaway	Financials	0.104
ADSK	Autodesk	Information Technology	0.103
CMI	Cummins	Industrials	0.096
EOG	EOG Resources	Energy	0.095

Notes: This table lists the ten S&P 500 firms with the highest average AI-adoption disclosure score—the section-weighted mean cosine similarity of their 10-K text to the AI-internal-adoption anchor—computed over fiscal years 2010–2026. For each firm the table reports its ticker, company name, GICS sector, and average AI-adoption score. The ranking has strong face validity, featuring enterprise-technology and software firms alongside an algorithm-centric platform and data-intensive financials; importantly, it is not confined to the Information Technology sector, with industrials and energy firms also appearing—evidence that intensive AI disclosure cuts across the sector taxonomy that a categorical control would impose.

4. Methodology

We bring six complementary research designs to bear, ranging from simple portfolio sorts to a market-model event study. Each carries its own identifying assumptions and its own vulnerabilities: a sort is transparent but model-free and silent on risk; a factor regression adjusts for risk but imposes a linear factor structure; a panel with fixed effects controls for unobserved heterogeneity but discards the cross-sectional variation that may matter most; an event study buys clean timing at the cost of a single window. By design these weaknesses do not overlap, so a conclusion that survives all six

is far more credible than one resting on any one. Using the battery together lets us triangulate the economic interpretation rather than lean on a single estimate.

4.1. Portfolio sorts

Each formation year we sort firms into quintiles on a given (within-year standardized) AI signal and compute the equal-weighted (EW) and value-weighted (VW) forward return of each quintile, along with the high-minus-low (Q5–Q1) spread that a self-financing long–short investor would have earned. We report the time-series average of the annual spreads and test its significance with Newey–West (Newey and West, 1987) standard errors (three lags) to account for any serial correlation induced by the overlapping return windows. The appeal of the sort is that it imposes no model and makes no functional-form assumption—it simply asks what high-AI firms earned relative to low-AI firms—and the EW–VW comparison is itself informative: when the equal-weighted spread exceeds the value-weighted one, the effect is concentrated in smaller index members rather than driven by the largest names.

4.2. Factor-model spanning of a long–short portfolio

A raw spread could in principle be no more than compensation for bearing known risks. To ask whether the AI spread survives risk adjustment, we build an annually rebalanced, equal-weighted, self-financing long–short portfolio (top minus bottom quintile), record its monthly return, and regress it on (i) the CAPM market factor and (ii) the Fama–French five factors plus momentum (FF6). The intercept (alpha) measures the average return the portfolio earned beyond what its loadings on these established factors can explain, and the slopes reveal which risks the strategy is implicitly taking; we use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors throughout. A significant alpha would indicate that AI disclosure is priced over and above standard exposures, while an insignificant one would say the spread is spanned by known factors.

4.3. Two-way fixed-effects panel regressions

We estimate

$$r_{i,t \rightarrow t+12} = \beta \text{AI}_{i,t} + \gamma' X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}, \quad (1)$$

where r is the winsorized forward return, $\text{AI}_{i,t}$ the standardized signal, X the controls, and α_i, δ_t firm and year fixed effects. Standard errors are double-clustered by firm and year (Petersen, 2009; Cameron et al., 2011). Firm fixed effects absorb all time-invariant cross-firm differences, so β identifies whether *within-firm* changes in AI language predict returns; year fixed effects absorb common shocks. Contrasting this estimate with the cross-sectional ones isolates whether the signal lives across or within firms (H4).

4.4. Fama–MacBeth regressions

We also run annual cross-sectional regressions of the forward return on the AI signal and controls and average the slopes following [Fama and MacBeth \(1973\)](#), with Newey–West errors on the time series of coefficients. This estimator weights each year equally and yields an interpretable per-standard-deviation premium; plotting the annual slopes reveals the time profile of the premium directly.

4.5. Structural-break and difference-in-differences designs

To test regime dependence (H2) we split the sample at the 30 November 2022 ChatGPT launch and compare sub-period spreads, and we estimate pooled interaction models of the form $r = \beta_1 \text{AI} + \beta_2 (\text{AI} \times \text{Post}) + \gamma' X + \varepsilon$, where $\text{Post} = \mathbb{1}[\text{formation year} \geq 2022]$ and β_2 captures the change in the premium. We complement this with a difference-in-differences design in which treatment is the firm’s *pre-determined* AI exposure—fixed at its 2019–2021 average so it cannot respond to the event—interacted with Post, within firm and year fixed effects. An event-time version interacts pre-determined exposure with each year dummy (relative to 2021) to trace the dynamic response.

4.6. Market-model event study

Finally, we run a classic event study ([MacKinlay, 1997](#)) around the ChatGPT launch, which converts the annual-frequency portfolio evidence into a high-frequency, cleanly dated test. The launch is close to an ideal natural experiment for our purposes: it was a consumer-product event, unanticipated by equity markets, that sharply raised the salience of AI without itself changing any firm’s fundamentals overnight. For each stock we estimate a market model ($r_{i,d} = a_i + b_i r_d^m + e_{i,d}$) over the estimation window January–October 2022 and compute daily abnormal returns $\text{AR}_{i,d} = r_{i,d} - (\hat{a}_i + \hat{b}_i r_d^m)$ over the event window November 2022–June 2024. Cumulative abnormal returns (CAR) are compared across terciles of *pre-event* AI disclosure—taken from the FY2021 10-K, the last report filed before the launch and therefore immune to any reaction to it—and we regress firm-level CAR on pre-event exposure with size, beta, and sector controls. Because abnormal returns net out each firm’s market-beta exposure, the event study isolates the cross-sectional response to the AI-narrative shock from the contemporaneous (and initially negative) market drift of 2022–2023, addressing the natural objection that high-AI firms simply rode a beta-driven recovery.

5. Results

5.1. Portfolio sorts: a weak unconditional relation

Table 4 reports the quintile sorts over the full sample. The relation between AI disclosure and future returns is positive and broadly monotone but economically modest. The equal-weighted Q5–Q1 spread is +4.6 percentage points per year for AI adoption ($t=1.83$), +3.9 for AI infrastructure ($t=1.58$), +5.3 for AI disruption risk ($t=2.27$), and +4.9 for the composite ($t=1.88$). Value-weighting attenuates the adoption and composite spreads, indicating the effect is stronger among smaller index members. The left panel of Figure 3 displays the full-sample pattern: returns rise gently from the lowest to the highest AI quintile. On its own, this is suggestive but far from decisive evidence—and, as we now show, it averages over two very different regimes.

Table 4. Portfolio Sorts on AI Disclosure

Signal (weighting)	Q1	Q2	Q3	Q4	Q5	Q5–Q1	t
AI adoption (EW)	16.2	17.9	19.1	18.6	20.8	+4.6	1.83
value-weighted	12.8	15.6	19.1	16.9	15.4	+2.6	0.84
AI infrastructure (EW)	16.8	18.4	17.3	19.6	20.6	+3.9	1.58
value-weighted	11.4	18.2	15.0	20.0	16.9	+5.5	2.01
AI disruption risk (EW)	16.1	18.7	17.2	19.2	21.4	+5.3	2.27
value-weighted	13.7	17.6	16.7	17.1	17.6	+3.9	1.74
AI composite (EW)	16.1	19.0	17.3	19.2	21.0	+4.9	1.88
value-weighted	11.7	17.8	16.2	19.5	16.6	+4.9	1.61

Notes: This table reports mean forward 12-month returns (in percent) for portfolios sorted into quintiles on each within-year standardized AI-disclosure signal, over the full sample of formation years 2010–2024. For each signal we report both the equal-weighted (EW) and value-weighted portfolio returns; Q1 is the lowest-disclosure quintile and Q5 the highest. The column Q5–Q1 is the high-minus-low spread, and t is its Newey–West (three-lag) t -statistic computed on the time series of 15 annual spreads. Equal-weighted spreads that exceed their value-weighted counterparts indicate that the relation is stronger among smaller index members.

5.2. Does AI disclosure earn risk-adjusted alpha unconditionally? (H1)

Table 5 reports CAPM and FF6 alphas of the long–short portfolios. The AI-opportunity composite earns a raw spread of about +2.3% per year, of which the FF6 alpha is +1.9%—but statistically insignificant ($t=1.24$). The AI-adoption and AI-infrastructure long–shorts are similarly insignificant after risk adjustment. The one exception is the AI-disruption-risk spread, whose FF6 alpha of +2.8% per year is marginally significant ($t=2.10$), hinting at a small risk premium for firms that flag AI as a competitive threat. The factor loadings (not the focus here) show the AI long–short is roughly

market-neutral with modest negative profitability (RMW) and investment (CMA) tilts, consistent with a growth-oriented long leg. The verdict on H1 is clear: **unconditionally, AI disclosure is at best weakly priced.** This is not the end of the story but the beginning of it.

Table 5. Risk-Adjusted Performance of the AI Long–Short Portfolio

Signal	Mean LS (%/yr)	CAPM α (%/yr)	t	FF6 α (%/yr)	t
AI adoption	+2.1	+1.5	1.08	+2.0	1.46
AI infrastructure	+1.7	+0.9	0.61	+1.1	0.84
AI composite	+2.3	+1.4	0.85	+1.9	1.24
AI disruption risk	+3.0	+2.1	1.38	+2.8	2.10

Notes: This table reports the risk-adjusted performance of the equal-weighted high-minus-low (Q5–Q1) long–short portfolio formed on each AI signal and rebalanced annually, over the full sample (178 monthly return observations). *Mean LS* is the average raw long–short return; *CAPM α* and *FF6 α* are the intercepts from time-series regressions of the long–short return on the CAPM market factor and on the Fama–French five factors plus momentum, respectively. All means and alphas are annualized and expressed in percent per year. The t -statistics are based on Newey–West (three-lag) heteroskedasticity- and autocorrelation-consistent standard errors. Over the full sample the alphas are economically small and, with the marginal exception of AI disruption risk, statistically insignificant.

5.3. Panel and Fama–MacBeth estimates

Table 6 juxtaposes the two-way fixed-effects panel with the Fama–MacBeth estimator. The contrast is instructive. In the Fama–MacBeth (cross-sectional) regressions, the AI-adoption and AI-infrastructure premia are positive—+1.3 and +1.6 percentage points per standard deviation—with t -statistics near 1.6–1.7, indicating a weak but positive cross-sectional premium on average. In the two-way fixed-effects panel, however, every AI coefficient collapses to essentially zero. Because firm fixed effects remove all permanent cross-firm differences, this tells us the signal lives *across* firms, not *within* them: a given firm writing relatively more about AI from one year to the next does not predict its own subsequent return (anticipating H4). Throughout, size carries the familiar negative premium ($t \approx -8$) and volatility a positive one, consistent with the broader cross-sectional literature in this large-cap universe.

Table 6. Panel and Fama–MacBeth Return Regressions

	(1) Two-way FE	(2) Fama–MacBeth
AI adoption (z)	-0.0047 ($t=-0.24$)	-0.0000 ($t=-0.00$)
AI infrastructure (z)	+0.0134 ($t=0.61$)	+0.0158 ($t=1.54$)
AI disruption risk (z)	-0.0153 ($t=-1.00$)	+0.0001 ($t=0.01$)
Size (z)	-0.2242 ($t=-7.17$)	-0.0330 ($t=-8.21$)
Momentum (z)	-0.0024 ($t=-0.25$)	+0.0081 ($t=1.08$)
Volatility (z)	+0.0197 ($t=2.12$)	+0.0687 ($t=3.56$)
Firm FE	Yes	No
Year FE	Yes	(by construction)
N (obs / years)	6811	15

Notes: This table reports forward-return regressions estimated two ways. Column (1) is a two-way (firm $\times$ year) fixed-effects panel with standard errors double-clustered by firm and year; the firm fixed effects absorb all permanent cross-firm differences, so the AI coefficients identify whether *within-firm* changes in AI language predict returns. Column (2) is the Fama–MacBeth estimator—the average of annual cross-sectional slopes—with Newey–West t -statistics on the time series of coefficients. The dependent variable is the winsorized forward 12-month return, and all signals and controls are within-year cross-sectional z -scores, so coefficients are expressed per cross-sectional standard deviation. The contrast between the two columns is the object of interest: the cross-sectional (Fama–MacBeth) AI premia are positive while the within-firm (panel) coefficients collapse to zero, locating the signal across rather than within firms.

5.4. The structural break: AI pricing switched on after ChatGPT (H2)

The full-sample averages are not so much wrong as misleading: they blend two regimes that look nothing alike, and in doing so they understate the post-2022 effect as badly as they overstate the pre-2022 one. This regime change is the paper’s central result. Splitting the sample at the November 2022 ChatGPT launch, the center and right panels of Figure 3 reveal a striking transformation. Before 2022 the relationship between AI quintile and forward return is nearly flat—a firm in the top AI quintile earned little more than one in the bottom quintile. After 2022 the same schedule tilts steeply upward, rising monotonically from the lowest to the highest quintile. The high-minus-low composite spread is +2.4 percentage points per year pre-ChatGPT but +14.6 points per year post-ChatGPT—a more than sixfold increase in the price of an essentially unchanged characteristic.

Table 7 quantifies the break two ways. In sub-period sorts, the AI-adoption Q5–Q1 spread rises from +2.2% per year ($t=2.04$) pre-2022 to +14.3% ($t=2.15$) post-2022; the composite rises from +2.5% to +14.6%; AI infrastructure from +1.4% to +13.6%. In the pooled interaction regressions, the post-2022 incremental slope is positive for all three signals and statistically significant for AI infrastructure ($t=2.63$)—the facet most directly tied to the chip, cloud, data-center, and power firms

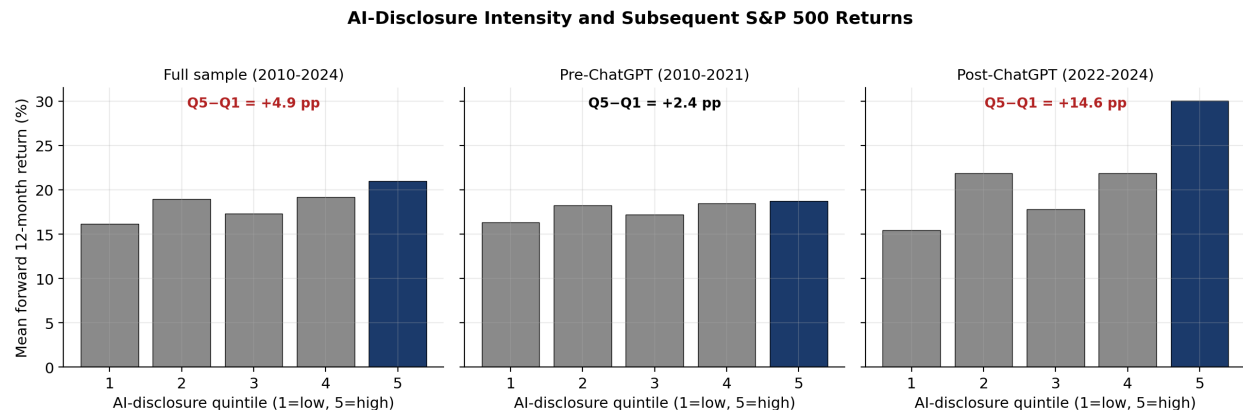


Figure 3. Forward returns by AI-disclosure quintile, before and after ChatGPT. Each panel plots the mean forward twelve-month return across quintiles of the AI-disclosure composite: the left panel pools the full sample, the center panel covers the pre-ChatGPT period (2010–2021), and the right panel the post-ChatGPT period (2022–2024). The quintile schedule is nearly flat before 2022 and steeply increasing afterward; the high-minus-low (Q5–Q1) spread rises from +2.4 percentage points per year before ChatGPT to +14.6 points after.

that led the 2023–2025 rally. Figure 4 shows the annual spreads: small and mixed-sign through 2021, uniformly large and positive for every post-2022 cohort.

Table 7. The Post-ChatGPT Structural Break

Signal	Q5–Q1 spread		Interaction
	Pre (2010–21)	Post (2022–24)	AI×Post
AI adoption	+2.2 ($t=2.04$)	+14.3 ($t=2.15$)	+0.0199 ($t=1.17$)
AI infrastructure	+1.4 ($t=0.91$)	+13.6 ($t=2.44$)	+0.0277 ($t=2.63$)
AI composite	+2.5 ($t=1.58$)	+14.6 ($t=2.07$)	+0.0270 ($t=1.78$)

Notes: This table quantifies the change in the pricing of AI disclosure around the November 2022 launch of ChatGPT. Columns 2–3 report the equal-weighted Q5–Q1 forward-return spread (in percent per year, with Newey–West t -statistics in parentheses) separately for the pre-period (formation years 2010–2021) and the post-period (2022–2024). Column 4 reports the coefficient on the interaction of the within-year standardized AI signal with the post-2022 indicator from a pooled OLS regression that includes size and momentum controls and year-clustered standard errors. The spread rises roughly sixfold across the break for every AI signal, and the interaction term is positive throughout—most strongly for AI infrastructure, the facet tied to the compute build-out.

The same regime change is visible in two further lenses. Figure 5 plots the cumulative value of one dollar invested in the equal-weighted, annually rebalanced, self-financing high-minus-low portfolio: it drifts sideways-to-up for a decade and then inflects upward after the dashed ChatGPT line. Figure 6 plots the year-by-year Fama–MacBeth slope on the AI composite (with controls): it hovers near zero for the entire 2010s and then turns sharply positive, peaking in the 2024 cohort.

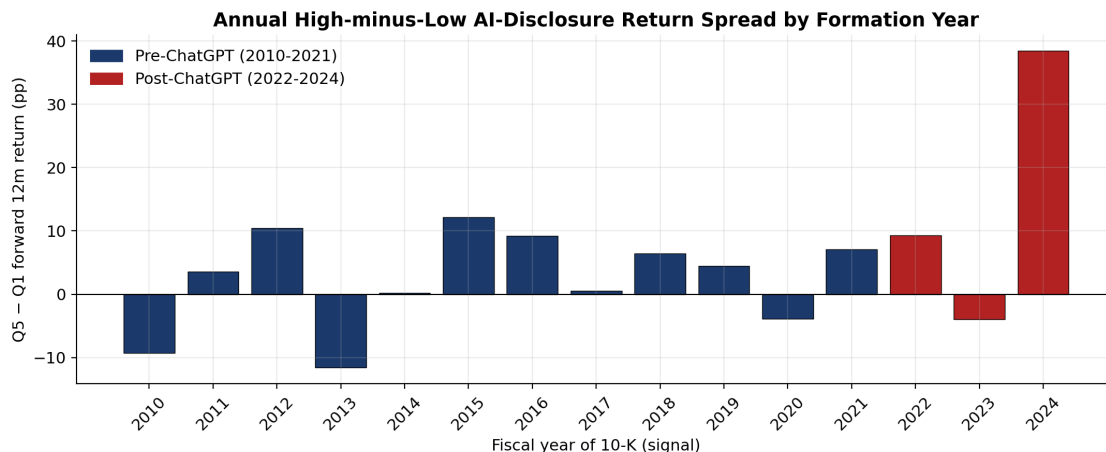


Figure 4. Annual AI-disclosure return spreads by cohort. The figure plots the annual high-minus-low (Q5–Q1) forward-return spread on the AI-disclosure composite, by 10-K fiscal (formation) year. Spreads are small and of mixed sign through 2021, whereas the post-ChatGPT cohorts (shown in red) are uniformly and substantially positive—showing that the regime change appears in every post-2022 cohort taken on its own, not only in the period average.

The latter figure also delivers an important nuance: the size-controlled premium is largest in the most recent cohort, consistent with the *AI infrastructure* super-cycle of 2024–2025 (the build-out of data centers and accelerated compute). We return to this concentration when discussing limitations.

Finally, Figure 7 confirms the pattern is not an artifact of a few influential observations: the binned relation between the AI composite and forward returns is essentially flat in the pre-ChatGPT sub-period and distinctly upward-sloping afterward, across the entire distribution of AI disclosure rather than only at the extreme quintiles.

5.5. The anatomy of the premium: size and sector

Where in the cross-section does the post-2022 premium concentrate? Table 8 and Figure 8 report an $AI \times size$ double sort. Before 2022 the AI gradient is flat within both size halves. After 2022 it is steep, and dramatically so among smaller firms: small, high-AI firms earned a mean forward return of 34.6%, versus 19.5% for small, low-AI firms—a 15-point gap—while the corresponding gap among large firms is 7 points. The premium is thus *not* merely a mega-cap phenomenon (a point we reinforce in Section 6); if anything it is strongest among the smaller, less-followed names, consistent with an attention-based mechanism in which information diffuses more slowly into smaller stocks.

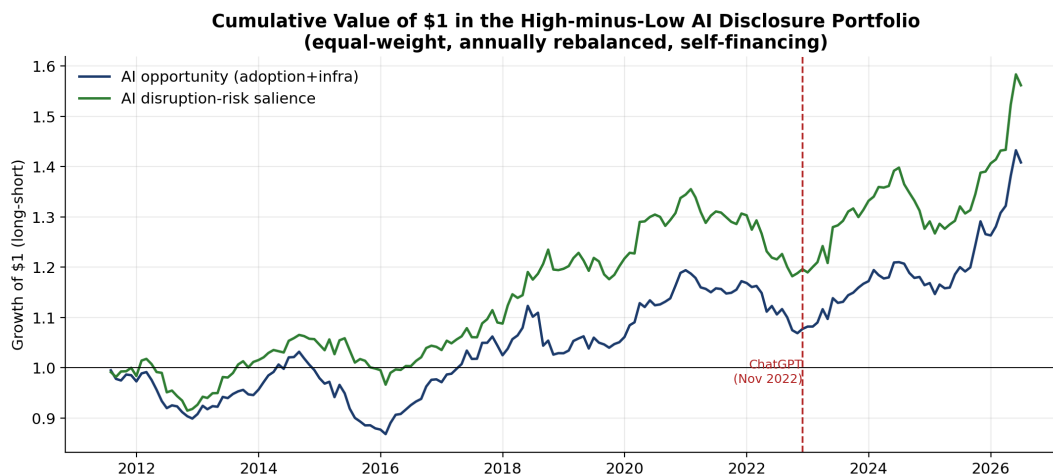


Figure 5. Cumulative performance of the AI long–short portfolio. The figure plots the cumulative value of \$1 invested in the equal-weighted, annually rebalanced, self-financing high-minus-low AI-disclosure portfolio, shown for both the AI-opportunity composite and AI disruption-risk signals. Both curves drift sideways-to-up for a decade and then inflect sharply upward after the dashed ChatGPT line, visualizing the regime change as a kink in accumulated wealth.

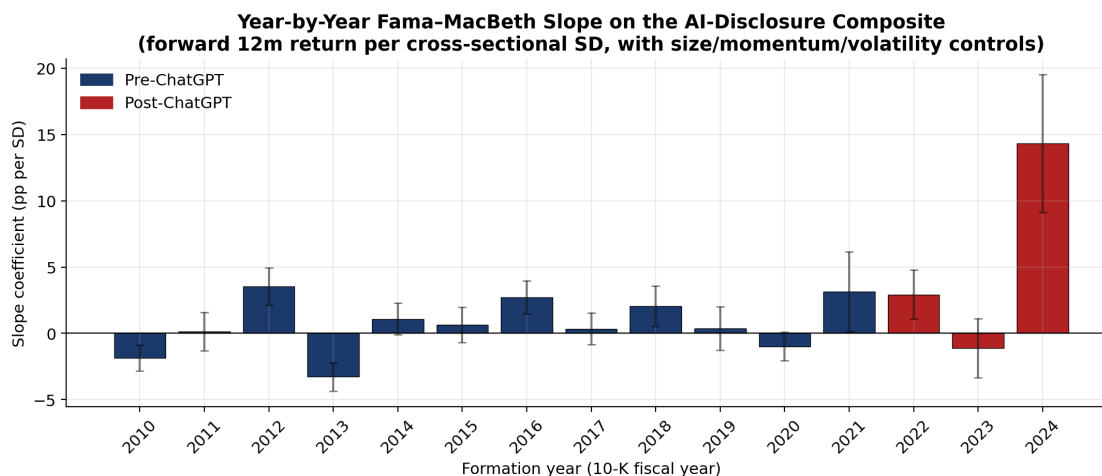


Figure 6. The AI premium over time: year-by-year Fama–MacBeth slopes. The figure plots the annual Fama–MacBeth slope on the AI-disclosure composite—the forward return earned per cross-sectional standard deviation of AI disclosure—controlling for size, momentum, and volatility, with error bars of ± 1 standard error. The slope hovers near zero throughout the 2010s and turns sharply positive after ChatGPT, peaking in the 2024 cohort, consistent with the 2024–2025 AI-infrastructure super-cycle.

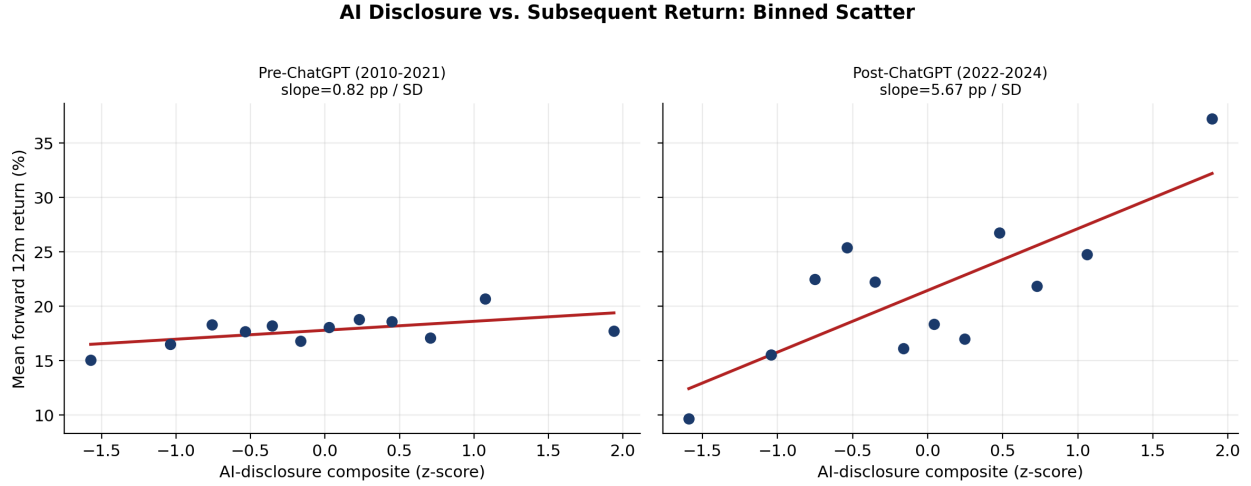


Figure 7. Binned relation between AI disclosure and returns, by sub-period. The figure shows a binned scatter of the forward twelve-month return against the AI-disclosure composite, plotted separately for the pre- and post-ChatGPT periods. Each point is a portfolio of firms within a bin of the AI signal and the line is an OLS fit. The conditional relation is essentially flat before ChatGPT and distinctly upward-sloping afterward, across the entire distribution of AI disclosure rather than only at the extreme quintiles—confirming that the effect is not driven by a few influential observations.

Table 8. AI Disclosure $\times$ Size Double Sort

	Low AI	Mid AI	High AI	Hi–Lo
<i>Panel A. Pre-ChatGPT (2010–2021)</i>				
Small	19.7	21.9	23.6	+3.9
Large	13.8	14.3	13.6	-0.1
<i>Panel B. Post-ChatGPT (2022–2024)</i>				
Small	19.5	19.2	34.5	+15.1
Large	15.8	18.3	22.9	+7.1

Notes: This table reports mean forward 12-month returns (in percent) from an independent double sort on AI disclosure and firm size. In each formation year firms are split into a small and a large half by the market-capitalization proxy, and within each half into AI terciles (Low, Mid, High); the column Hi–Lo is the high-minus-low AI return within each size group. Panel A covers the pre-ChatGPT period (formation years 2010–2021) and Panel B the post-ChatGPT period (2022–2024). The AI gradient is flat in both size halves before 2022 and steep afterward, and dramatically steeper among small firms—consistent with an attention-based mechanism in which information diffuses more slowly into smaller, less-followed stocks.

5.6. Economic significance

The statistical break is matched by an economic one, and it is worth dwelling on the magnitudes because they are unusual. The Sharpe ratio of the equal-weighted AI-composite long–short

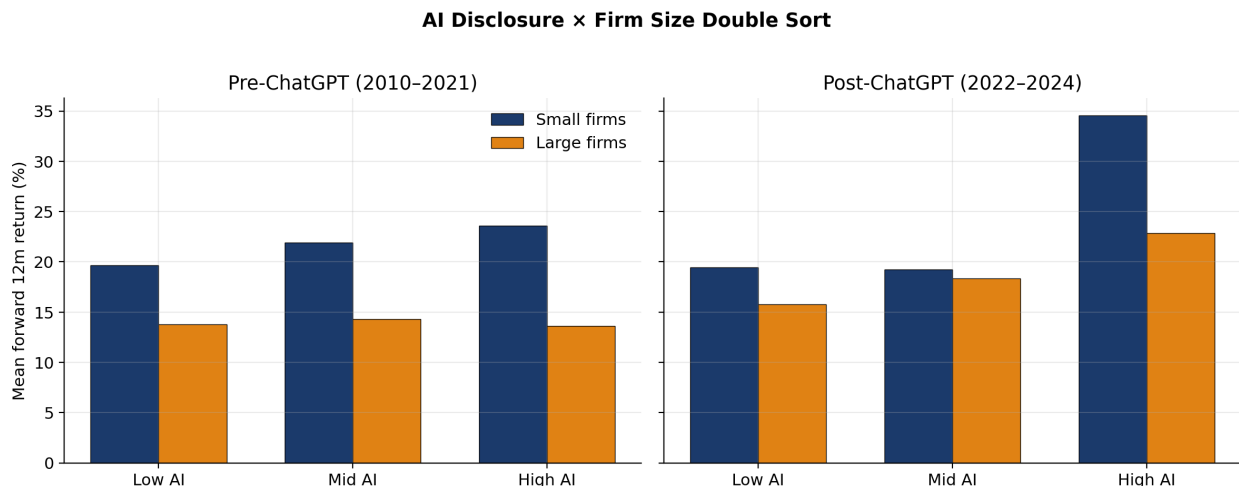


Figure 8. AI disclosure × firm-size double sort. The figure plots the mean forward twelve-month return by AI tercile within small and large firms, separately for the pre- and post-ChatGPT periods. The post-2022 AI gradient is steep in both size groups but dramatically steeper among small firms, where high-AI firms outperform low-AI firms by roughly 15 percentage points, consistent with an attention-based mechanism that operates most strongly in less-followed stocks.

portfolio—a direct gauge of return earned per unit of risk borne—rises from 0.15 in the pre-ChatGPT decade to 1.33 afterward; for the AI-disruption-risk spread it rises from 0.33 to 1.15. To put the post-2022 figure in perspective, the broad U.S. equity market has historically delivered a Sharpe ratio in the neighborhood of 0.4, so a value above one for a single-characteristic, self-financing strategy is exceptional, and its emergence precisely at the ChatGPT boundary—rather than gradually, as a slowly learned risk premium would—is hard to reconcile with a constant risk premium. The point estimates in the quintile sorts are equally vivid: a 14-point annual spread between the top and bottom AI quintiles is roughly three times the unconditional equity risk premium and an order of magnitude larger than the pre-2022 spread. Even acknowledging that a portion of this reflects the higher realized volatility of AI-exposed firms (documented in Section 6) and the broad technology rally, the magnitude of the post-2022 re-rating is among the largest single-theme cross-sectional effects in the recent asset-pricing record.

6. Robustness and Identification

The structural break is economically large—and large effects invite scrutiny, since they are also the ones most likely to be artifacts. This section therefore asks how much we should believe it, confronting the result with the alternative explanations a skeptical reader would raise: that the break is mechanical, that it is a mega-cap rally in disguise, that it is sensitive to how the signal is built, or that it merely tracks the technology sector. We subject it in turn to a placebo across non-AI

factors, an event study, a difference-in-differences design, and a battery of construction and universe robustness checks. The result withstands each.

6.1. Placebo: is the effect specific to AI? (H3)

If the post-2022 spread were a mechanical artifact of our pipeline—or a generic “high-disclosure firms did well” phenomenon—it should appear for any semantic factor. Table 9 and Figure 9 re-run the entire sub-period and interaction design on all eight factors. The result strongly supports AI-specificity. The three AI factors show large positive post-2022 spreads (+13.6 to +14.3 points). Macro-risk language shows essentially no spread (-0.5 , $t = -0.1$). And, strikingly, ESG-commitment and climate-risk disclosure earn *negative* post-2022 spreads (-8.8 and -5.0 points, both significant)—high-ESG firms *underperformed* after 2022, in line with the post-2021 reversal in the sustainable-investing literature (Pástor et al., 2022). The one non-AI factor that does show a large positive post-2022 spread is cybersecurity (+13.5); we read this not as a violation but as confirmation, since cybersecurity disclosure is tightly bundled with technology exposure and is itself increasingly AI-related. The cleanest contrast—AI rewarded, ESG and climate penalized—makes a powerful case that the post-ChatGPT re-rating reflects a thematic rotation *toward AI*, not a generic disclosure effect.

Table 9. Placebo Tests across Semantic Factors

Semantic factor	Q5–Q1 spread (%/yr)		Interaction
	Pre (2010–21)	Post (2022–24)	$z \times \text{Post}$
AI adoption [†]	+2.2 ($t=2.04$)	+14.3 ($t=2.15$)	+0.0199 ($t=1.17$)
AI infrastructure [†]	+1.4 ($t=0.91$)	+13.6 ($t=2.44$)	+0.0277 ($t=2.63$)
AI disruption risk [†]	+3.1 ($t=2.47$)	+14.1 ($t=2.07$)	+0.0108 ($t=0.60$)
ESG commitment	-2.5 ($t=-2.82$)	-8.8 ($t=-2.27$)	-0.0055 ($t=-0.76$)
Climate risk	-0.8 ($t=-1.38$)	-5.0 ($t=-2.48$)	+0.0008 ($t=0.17$)
Cybersecurity	+2.6 ($t=1.96$)	+13.5 ($t=2.12$)	+0.0090 ($t=0.64$)
Automation/labor	+2.1 ($t=1.91$)	+7.2 ($t=1.29$)	+0.0002 ($t=0.01$)
Macro risk	-1.6 ($t=-2.02$)	-0.5 ($t=-0.11$)	+0.0125 ($t=1.06$)

Notes: This table re-runs the structural-break design on all eight semantic factors in the database to test whether the post-2022 re-rating is specific to AI. For each factor it reports the equal-weighted Q5–Q1 forward-return spread (in percent per year, with Newey–West t -statistics) in the pre-period (2010–2021) and post-period (2022–2024), together with the coefficient on the interaction of the factor’s within-year z -score with the post-2022 indicator (size and momentum controls, year-clustered standard errors). The post-ChatGPT jump is concentrated in the AI-related factors, marked with a dagger (†); cybersecurity, which is bundled with technology exposure, also rises, whereas ESG-commitment and climate-risk disclosure earn *negative* post-2022 spreads and macro-risk disclosure shows essentially none.

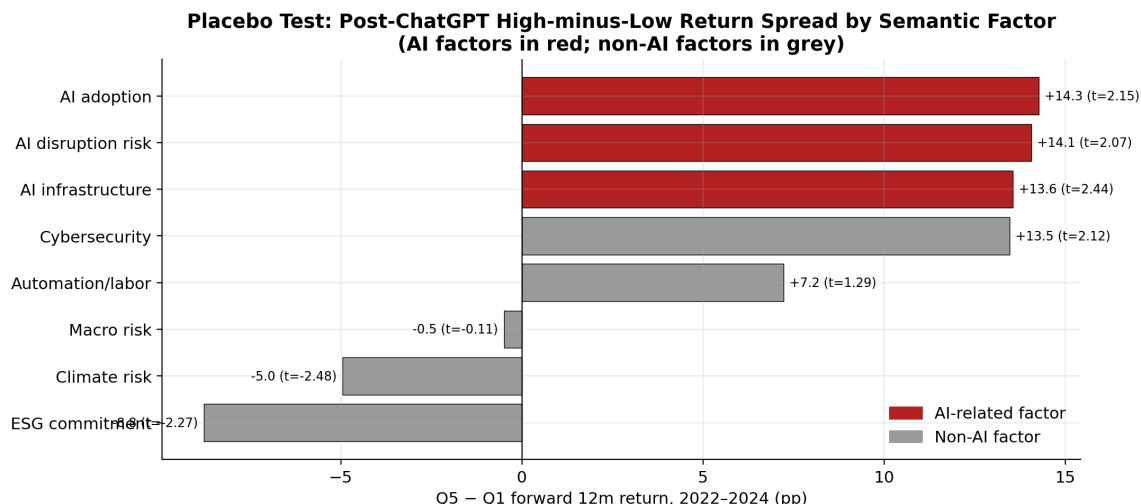


Figure 9. Placebo test across semantic factors. The figure plots the post-ChatGPT high-minus-low (Q5–Q1) return spread for each of the eight semantic factors. The AI-related factors (in red) cluster at the top with large positive spreads, while ESG-commitment and climate-risk disclosure earn *negative* spreads and macro-risk disclosure earns essentially none. The contrast shows the post-2022 re-rating is a thematic rotation toward AI, not a generic disclosure effect.

6.2. Event study around ChatGPT

The portfolio results use annual formation and forward returns; the event study provides a higher-frequency, cleanly identified complement. Table 10 and Figure 10 report market-model cumulative abnormal returns by pre-event AI tercile. The ordering is monotone at every horizon and widens steadily over time. The High-minus-Low CAR difference is +2.9 points at 60 trading days ($t=1.77$), +10.1 points at one year ($t=1.97$), and +14.6 points at eighteen months ($t=1.89$). Because these are abnormal returns—purged of each firm’s market beta—the spread cannot be attributed to high-AI firms simply being higher-beta in a recovering market. A cross-sectional regression of one-year CAR on pre-event AI exposure yields +3.1 points per standard deviation ($t=1.39$); adding sector fixed effects halves it to +2.1 points ($t=0.99$), again locating much—but not all—of the effect in a sector channel. Figure 11 shows the relation is smooth across the exposure distribution, not driven by a few outliers.

Table 10. ChatGPT Event Study

Horizon	High	Mid	Low	High–Low	<i>t</i>	<i>p</i>
+60 trading days	+1.89	+1.79	-1.04	+2.93	1.77	0.077
+6 months	-2.72	-5.17	-7.25	+4.53	1.60	0.110
+1 year	-6.38	-11.80	-16.46	+10.08	1.97	0.050
+18 months	-8.46	-16.03	-23.05	+14.59	1.89	0.060

Cross-sectional regression, CAR(+1y) on pre-event AI exposure (per SD):

No sector FE: +3.13% ($t=1.39$), $n = 483$, $R^2 = 0.160$ With sector FE: +2.09% ($t=0.99$)

Notes: This table reports market-model cumulative abnormal returns (CAR, in percent) around the launch of ChatGPT on 30 November 2022. For each stock the market model is estimated over the January–October 2022 window, and abnormal returns are cumulated over horizons of 60 trading days, six months, one year, and 18 months. Firms are grouped into terciles of *pre-event* AI disclosure measured from the FY2021 10-K—the last report filed before the event—and the table reports the mean CAR for the High, Mid, and Low terciles, the High–Low difference, and a Welch *t*-test (with *p*-value) of that difference. The bottom rows report a cross-sectional regression of one-year CAR on pre-event AI exposure (per standard deviation), with and without sector fixed effects. Because abnormal returns are purged of each firm’s market beta, the spread reflects a cross-sectional response to the AI-narrative shock rather than differential beta exposure.

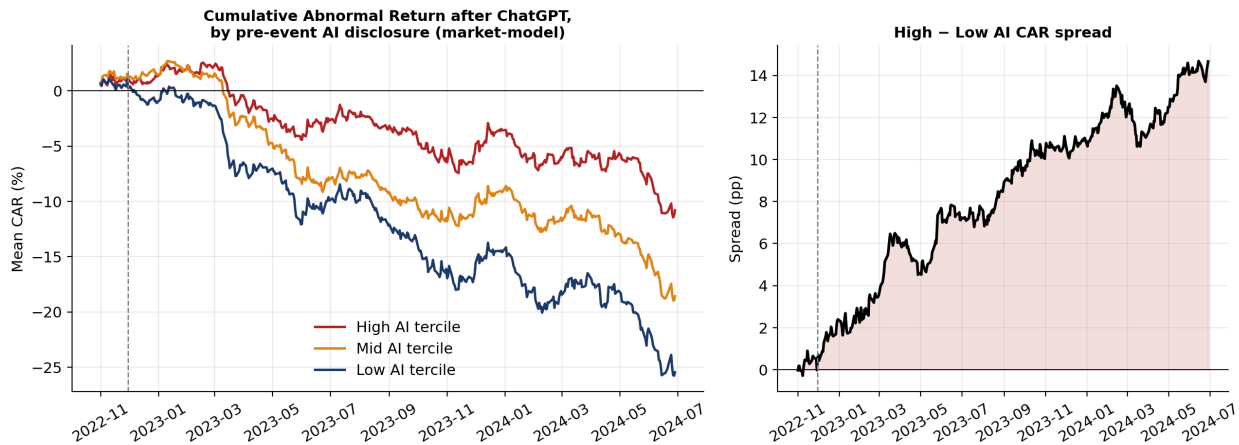


Figure 10. Cumulative abnormal returns around ChatGPT. The left panel plots the mean market-model cumulative abnormal return (CAR) by pre-event AI-disclosure tercile over the event window; the right panel plots the High–Low CAR spread. The tercile ordering is monotone at every horizon and widens steadily, reaching roughly +10 percentage points one year after the launch. Because abnormal returns are purged of market beta, the spread reflects a cross-sectional response to the AI-narrative shock.

6.3. Difference-in-differences and the cross-sectional nature of the effect (H4)

Table 11 reports difference-in-differences estimates in which treatment is the firm’s pre-determined (2019–2021) AI exposure interacted with the post-2022 indicator, within firm and year fixed effects.

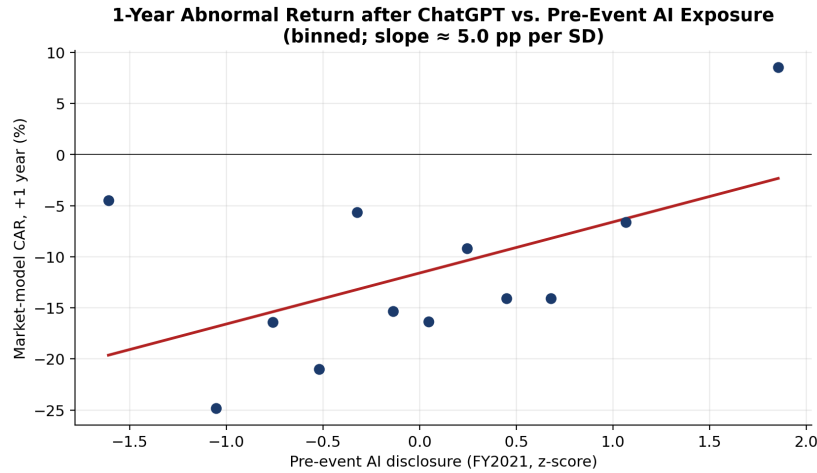


Figure 11. One-year abnormal return versus pre-event AI exposure. The figure plots a binned scatter, with fitted line, of each firm’s one-year cumulative abnormal return after ChatGPT against its pre-event AI exposure (measured from the FY2021 10-K). The relation is positive and smooth across the exposure distribution, confirming that the event-study spread is not driven by a handful of outliers.

The continuous-exposure DiD coefficient is +2.7 percentage points per standard deviation ($t=1.51$); the binary version is +1.7 points ($t=0.78$). These are positive and economically meaningful but weaker than the cross-sectional estimates—exactly the pattern predicted by H4. Once firm fixed effects absorb the permanent component of AI exposure, only the residual within-firm variation identifies the effect, and that variation is small. The event-time profile in Figure 12 reinforces this reading: the return premium on a firm’s *fixed* pre-event AI exposure is flat and near zero through 2021 and jumps after 2022, peaking in 2024. The market did not reward firms for adding AI sentences to their filings; it repriced the firms whose pre-existing, persistent exposure made them AI winners once the narrative turned.

Table 11. Difference-in-Differences around ChatGPT

	Treated×Post (binary)	Exposure×Post (continuous, per SD)
Coefficient	0.0170	0.0269
<i>t</i> -statistic	(0.78)	(1.51)
N (firm-years)	6790	6790
Firm & Year FE	Yes	Yes

Notes: This table reports difference-in-differences estimates of the effect of pre-determined AI exposure on returns around ChatGPT. The dependent variable is the winsorized forward 12-month return, and all specifications include firm and year fixed effects with standard errors double-clustered by firm and year. Treatment is fixed at each firm’s pre-event (2019–2021) average AI exposure, so it cannot respond to the event: the first column uses a binary treated indicator (above-median pre-event exposure) interacted with the post-2022 indicator, and the second uses continuous pre-event exposure (per standard deviation) interacted with Post. Both coefficients are positive but weaker than the cross-sectional estimates, the pattern expected once firm fixed effects absorb the persistent component of AI exposure and only within-firm variation remains to identify the effect.

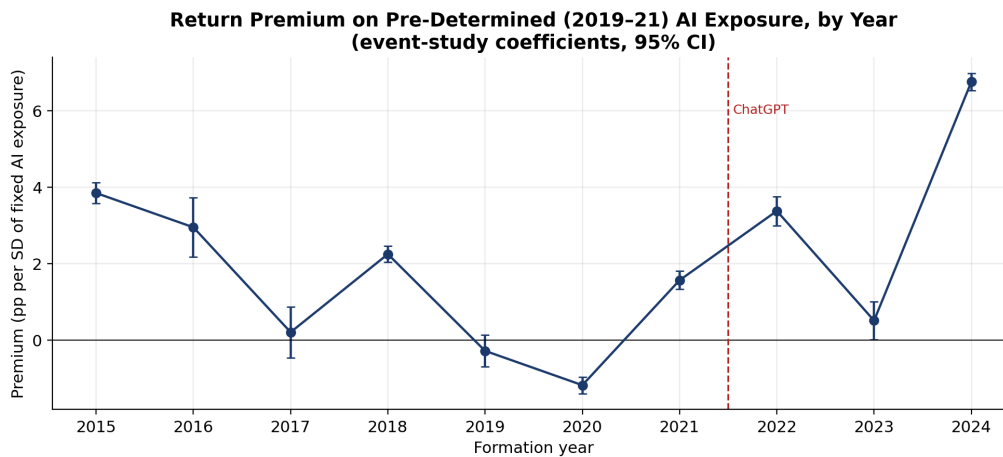


Figure 12. Event-time profile of the AI premium. The figure plots the estimated return premium on firms’ pre-determined (2019–2021) AI exposure by year, with 95% confidence intervals, measured relative to the 2021 base year and within firm and year fixed effects. The premium is flat and near zero through 2021 and activates after ChatGPT, peaking in 2024—evidence that the market repriced persistent pre-existing exposure rather than rewarding incremental wording.

6.4. Construction, horizon, and universe robustness

Table 12 collects the remaining checks. The post-2022 spread is essentially unchanged whether the signal is the section-weighted mean (base, +14.6), the simple mean (+13.4), or the top-5% tail (+14.9). It holds at six-, twelve-, and twenty-four-month horizons (growing with the horizon, as

a persistent premium should). It survives replacing the three AI signals with their first principal component—which alone explains 86.8% of their common variance, confirming the “single latent AI construct” interpretation—yielding a post-2022 PC1 spread of +13.4 points ($t=1.97$). Most importantly for the mega-cap concern, the spread is +14.1 points ($t=1.92$) when the Magnificent 7 are excluded and +14.5 points ($t=1.93$) when the ten largest firms are dropped (Figure 13). The AI re-rating is a broad-based feature of the index, not the mechanical footprint of a handful of trillion-dollar winners. Finally, the sector-neutral composite—which compares firms only to others *within* their GICS sector—produces a much smaller and insignificant post-2022 spread (+3.8 points, $t = 0.84$). This is not a failure of the result but a sharpening of it: a large share of the AI premium operates through a *between-sector* channel, as the market re-rated AI-exposed sectors (technology above all) relative to others, with a smaller within-sector residual.

Table 12. Robustness of the AI Return Spread

	Pre (2010–21)	Post (2022–24)
<i>Panel A. Alternative signal definition (composite)</i>		
Section-weighted (base)	+2.5 ($t=1.58$)	+14.6 ($t=2.07$)
Simple mean	+2.3 ($t=1.50$)	+13.4 ($t=2.23$)
Top-5% tail	+1.5 ($t=1.04$)	+14.9 ($t=2.47$)
<i>Panel B. Alternative holding horizon</i>		
6-month forward return	-0.1 ($t=-0.10$)	+1.1
24-month forward return	+9.3 ($t=2.86$)	+23.0
<i>Panel C. Construction / universe</i>		
Sector-neutral composite	+2.3 ($t=3.35$)	+3.8 ($t=0.84$)
First principal component	+2.6 ($t=2.25$)	+13.4 ($t=1.97$)
Excl. Magnificent-7	+1.7 ($t=1.07$)	+14.1 ($t=1.92$)
Excl. top-10 by market cap	—	+14.5 ($t=1.93$)

Notes: This table assesses the robustness of the AI composite Q5–Q1 forward-return spread (in percent per year, with Newey–West t -statistics) across alternative research-design choices, reported separately for the pre-period (2010–2021) and post-period (2022–2024). Panel A varies the signal definition (section-weighted mean, simple mean, top-5% tail); Panel B varies the holding horizon (6- and 24-month forward returns); and Panel C varies the construction and universe (sector-neutralization, replacing the three signals with their first principal component, and excluding the Magnificent 7 or the ten largest firms by market capitalization). The post-ChatGPT spread is stable across signal definitions, grows with the horizon, survives the exclusion of mega-cap winners, and shrinks markedly only under sector-neutralization—locating much of the premium in a between-sector channel.

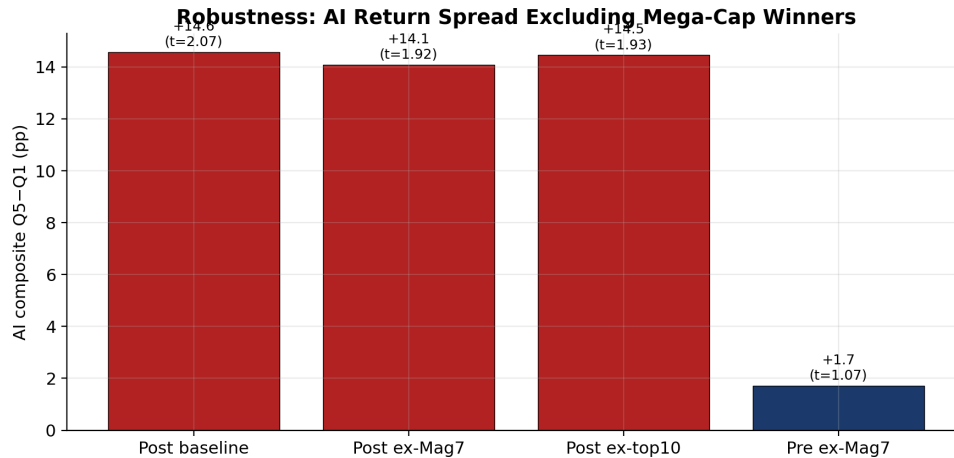


Figure 13. The AI spread is not a mega-cap artifact. The figure compares the post-2022 composite Q5–Q1 return spread for the full index against the spread computed after excluding the Magnificent 7 and after excluding the ten largest firms by market capitalization. The spread is essentially unchanged in both cases, showing that the AI re-rating is a broad-based feature of the index rather than the footprint of a handful of trillion-dollar winners.

6.5. An alternative performance outcome: risk

Performance is not only about mean returns. We also relate AI disclosure to subsequent realized volatility. High-AI firms exhibit modestly higher forward volatility, and the gap widens after 2022 (a Q5–Q1 difference of about 2.9 annualized volatility points, $t = 3.1$). The post-2022 AI premium therefore came bundled with somewhat higher risk—part, though by no means all, of the elevated returns is compensation for the greater volatility of AI-exposed firms during a period of rapid technological repricing.

7. Discussion: Mechanism and Interpretation

What economic mechanism best fits this constellation of facts? We find an attention- and narrative-driven repricing the most parsimonious explanation, and we read the evidence as follows.

The information pre-existed; the attention did not. Firms disclosed their AI engagement for years before it was priced. The semantic content of those disclosures was, in cross-section, only faintly related to returns through 2021. What changed in late 2022 was not the information but its *salience*: a consumer breakthrough (ChatGPT) made AI the dominant market narrative (Shiller, 2017), drawing investor attention (Da et al., 2011; Barber and Odean, 2008) to a body of disclosure that had been hiding in plain sight. The abrupt, near-simultaneous repricing across the cross-section—visible in the event study within months of the launch—is the signature of an attention shock rather than the gradual accrual of a risk premium.

It is a re-rating of levels, not of changes. The disappearance of the effect under firm fixed effects, the muted difference-in-differences estimates, and the event-time profile all point the same way: the market repriced the persistent *level* of a firm’s AI exposure, not the year-to-year change in its wording. This has a clear practical corollary—firms cannot mechanically earn the premium by adding AI boilerplate to their filings (and our top-5% and simple-mean robustness suggests the measure is not easily gamed)—and a clear methodological one: studies that difference out firm effects will understate the cross-sectional importance of disclosure content.

It operates substantially, but not entirely, through sectors. The sector-neutral and sector-fixed-effects results show that much of the AI premium reflects the re-rating of AI-exposed *sectors*—technology, communication services, and the industrial firms supplying the AI build-out—relative to the rest of the index. This is economically sensible: the generative-AI boom was, to first order, a sectoral reallocation of expected cash flows and valuations. Yet a positive within-sector residual remains in the event study, indicating that the firm’s own disclosure carried incremental information even among industry peers.

The thematic asymmetry is telling. Perhaps the most revealing single fact is the placebo: in the very same post-2022 window, high-AI firms were rewarded while high-ESG and high-climate firms were penalized. Markets did not simply reward verbose or virtuous-sounding disclosure; they rotated toward one theme (AI) and away from another (sustainability). This is difficult to reconcile with a purely mechanical or risk-based account and natural under a narrative-rotation view of investor attention.

Relation to the AI-asset-pricing literature. Our findings complement [Eisfeldt et al. \(2023\)](#), who show that an occupation-task measure of generative-AI exposure was priced around ChatGPT. We reach a convergent conclusion from an entirely different measurement—firms’ own disclosure language—and extend it along the time dimension, showing that the same exposure was *unpriced* for a decade beforehand. The agreement of two independent measures on the timing and sign of the effect strengthens confidence that a real AI repricing occurred.

7.1. Implications

Our findings carry distinct implications for three audiences. *For investors and quantitative strategists*, the results are a caution as much as an opportunity. The AI-disclosure signal delivered a striking post-2022 Sharpe ratio, but the same signal was nearly inert for the preceding decade, and the effect is absorbed by firm fixed effects—so a strategy that traded on year-to-year changes in AI verbosity would have captured little. The tradeable content lies in the persistent cross-sectional *level* of AI

exposure, and even that is partly a sector bet whose durability is unproven. The honest reading is that disclosure-based AI signals are best viewed as a lens on a regime-dependent theme rather than as a stable alpha source. *For corporate managers and disclosure policy*, the asymmetry between AI (rewarded) and ESG/climate (penalized) after 2022 illustrates how powerfully the prevailing market narrative conditions the valuation consequences of what firms choose to emphasize. Because the premium attaches to persistent, hard-to-fake exposure rather than to boilerplate, the result does not invite cheap disclosure manipulation; but it does suggest that the market reads filings for genuine strategic content and reprices it when attention turns. *For researchers*, the study demonstrates that embedding-based semantic measurement uncovers pricing dynamics invisible to both keyword counts (which cannot distinguish meaning) and firm-differenced panels (which discard the persistent cross-sectional variation that turns out to matter).

7.2. Summary of the evidence across designs

It is worth stepping back to see how the designs corroborate one another, since each has different weaknesses. The portfolio sorts establish the raw magnitude and its post-2022 timing but are model-free and could in principle reflect risk; the factor-model spanning regressions show the unconditional spread is not explained by standard factors yet is statistically weak over the full sample; the sub-period and interaction tests localize the effect to the post-ChatGPT regime; the placebo rules out a mechanical or generic-disclosure explanation by showing the opposite sign for ESG and climate; the event study, using daily abnormal returns purged of beta, provides cleanly-timed, cross-sectional confirmation within months of the catalyst; and the fixed-effects and difference-in-differences estimates pin down that the priced object is persistent firm- and sector-level exposure, not incremental wording. No single design is decisive, but their convergence—on the same sign, the same timing, and a consistent economic story—is what gives us confidence in the conclusion.

7.3. Limitations

Several caveats temper our conclusions. *First*, the post-2022 window contains only three formation cohorts, and the size-controlled premium is concentrated in the most recent (2024) cohort, which captures the 2024–2025 AI-infrastructure capital-expenditure boom. The regime result is economically large and consistent across designs, but its post-period rests on a short sample and could in principle reflect a single extraordinary episode rather than a durable new regime; only future data can settle this. *Second*, our universe is current S&P 500 membership, which introduces survivorship and look-ahead-at-the-index-level bias. This biases the level of returns upward but is less likely to drive the *cross-sectional* AI–return relation that is our focus. *Third*, the market-capitalization proxy

holds shares outstanding at their latest value, approximating historical size with error. *Fourth*, the semantic anchors, though carefully written, embed the authors’ definitional choices; different anchor wording would yield somewhat different scores, although the robustness to mean/tail aggregations and to PCA mitigates this concern. *Fifth*, disclosure intensity is correlated with sector, and while our within-year, size-, momentum-, and volatility-controlled tests and the sector-FE event study address this, we cannot fully separate “AI exposure” from “technology sector” as the priced characteristic. *Finally*, our designs establish a robust association and an event-study-based timing argument, but not strict causality; the attention-repricing interpretation is the most consistent with the evidence rather than the only logically possible one.

8. Conclusion

Using a semantic, embedding-based measure of artificial-intelligence language in the SEC filings of S&P 500 firms, linked to Yahoo Finance returns and standard factor data, we have shown that the pricing of AI disclosure is a *conditional, regime-dependent* phenomenon. For most of the 2010s, how intensively a firm wrote about adopting, building, or being threatened by AI bore only a faint relation to its subsequent returns, and a risk-adjusted long–short strategy earned no reliable alpha. After the launch of ChatGPT in late 2022, the same disclosure became strongly priced: the high-minus-low AI-disclosure return spread jumped from roughly +2 to roughly +15 percentage points per year, and the long–short Sharpe ratio rose from 0.15 to 1.33.

The break is robust to signal definition, horizon, dimension reduction, and the exclusion of mega-cap winners; it is specific to AI-related language (ESG and climate disclosure, tested identically, were *penalized* in the same window); it is corroborated by a market-model event study in which high-AI firms earned +10 points of additional abnormal return in the year after ChatGPT; and it operates as a cross-sectional re-rating of persistent firm- and sector-level exposure rather than a reward for year-to-year wording, fading under firm fixed effects.

Two messages emerge, one economic and one methodological. The economic message is that markets price the *language* of AI—but only once AI commands their attention; value-relevant information can sit unpriced in plain sight for years until a salient catalyst turns the market’s gaze toward it, after which the repricing is abrupt rather than gradual. The methodological message is that embedding-based measurement of disclosure content opens a productive avenue for asset pricing, capturing economically meaningful, multi-faceted exposures at scale and revealing pricing dynamics that keyword counts (which cannot read meaning) and firm-differenced designs (which discard the persistent cross-sectional variation that turns out to matter) would both miss. As generative AI continues to reshape the corporate landscape, the semantic footprint it leaves in mandatory disclosure offers a rich, and now demonstrably value-relevant, signal. Several next steps

follow naturally: extending the panel beyond the S&P 500 to the mid- and small-cap firms where attention frictions are larger and the effect may be stronger; incorporating earnings-call transcripts and patent text as complementary, higher-frequency AI-exposure measures; and—perhaps most consequentially for investors—tracking whether the post-2022 premium persists, decays, or reverses as the AI narrative matures and the market’s attention becomes fully impounded in prices.

References

- Acemoglu, Daron and Pascual Restrepo**, “Robots and Jobs: Evidence from US Labor Markets,” *Journal of Political Economy*, 2020, 128 (6), 2188–2244.
- Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson**, “Artificial Intelligence, Firm Growth, and Product Innovation,” *Journal of Financial Economics*, 2024, 151, 103745.
- Baker, Malcolm and Jeffrey Wurgler**, “Investor Sentiment and the Cross-Section of Stock Returns,” *Journal of Finance*, 2006, 61 (4), 1645–1680.
- Barber, Brad M. and Terrance Odean**, “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors,” *Review of Financial Studies*, 2008, 21 (2), 785–818.
- Bolton, Patrick and Marcin Kacperczyk**, “Do Investors Care about Carbon Risk?,” *Journal of Financial Economics*, 2021, 142 (2), 517–549.
- Bybee, Leland, Bryan Kelly, Asaf Manela, and Dacheng Xiu**, “Business News and Business Cycles,” *Journal of Finance*, 2024, 79 (5), 3105–3147.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller**, “Robust Inference with Multiway Clustering,” *Journal of Business & Economic Statistics*, 2011, 29 (2), 238–249.
- Cohen, Lauren, Christopher Malloy, and Quoc Nguyen**, “Lazy Prices,” *Journal of Finance*, 2020, 75 (3), 1371–1415.
- Cong, Lin William, Tengyuan Liang, and Xiao Zhang**, “Textual Factors: A Scalable, Interpretable, and Data-Driven Approach to Analyzing Unstructured Information,” Technical Report, Working Paper 2022.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao**, “In Search of Attention,” *Journal of Finance*, 2011, 66 (5), 1461–1499.
- DellaVigna, Stefano and Joshua M. Pollet**, “Investor Inattention and Friday Earnings Announcements,” *Journal of Finance*, 2009, 64 (2), 709–749.
- Eisfeldt, Andrea L., Gregor Schubert, and Miao Ben Zhang**, “Generative AI and Firm Values,” NBER Working Paper 31222, National Bureau of Economic Research 2023.
- Fama, Eugene F. and James D. MacBeth**, “Risk, Return, and Equilibrium: Empirical Tests,” *Journal of Political Economy*, 1973, 81 (3), 607–636.

- and **Kenneth R. French**, “A Five-Factor Asset Pricing Model,” *Journal of Financial Economics*, 2015, 116 (1), 1–22.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy**, “Text as Data,” *Journal of Economic Literature*, 2019, 57 (3), 535–574.
- Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li**, “Innovative Efficiency and Stock Returns,” *Journal of Financial Economics*, 2013, 107 (3), 632–654.
- , **Sonya S. Lim, and Siew Hong Teoh**, “Driven to Distraction: Extraneous Events and Underreaction to Earnings News,” *Journal of Finance*, 2009, 64 (5), 2289–2325.
- Hoberg, Gerard and Gordon Phillips**, “Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis,” *Review of Financial Studies*, 2010, 23 (10), 3773–3811.
- and —, “Text-Based Network Industries and Endogenous Product Differentiation,” *Journal of Political Economy*, 2016, 124 (5), 1423–1465.
- Kelly, Bryan, Asaf Manela, and Alan Moreira**, “Text Selection,” *Journal of Business & Economic Statistics*, 2021, 39 (4), 859–879.
- Kogan, Leonid and Dimitris Papanikolaou**, “Growth Opportunities, Technology Shocks, and Asset Prices,” *Journal of Finance*, 2014, 69 (2), 675–718.
- , —, **Amit Seru, and Noah Stoffman**, “Technological Innovation, Resource Allocation, and Growth,” *Quarterly Journal of Economics*, 2017, 132 (2), 665–712.
- Lopez-Lira, Alejandro**, “Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns,” Technical Report, Working Paper 2023.
- Loughran, Tim and Bill McDonald**, “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” *Journal of Finance*, 2011, 66 (1), 35–65.
- and —, “Textual Analysis in Accounting and Finance: A Survey,” *Journal of Accounting Research*, 2016, 54 (4), 1187–1230.
- MacKinlay, A. Craig**, “Event Studies in Economics and Finance,” *Journal of Economic Literature*, 1997, 35 (1), 13–39.
- Newey, Whitney K. and Kenneth D. West**, “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 1987, 55 (3), 703–708.

- Pástor, L’uboš, Robert F. Stambaugh, and Lucian A. Taylor**, “Sustainable Investing in Equilibrium,” *Journal of Financial Economics*, 2021, 142 (2), 550–571.
- , —, and —, “Dissecting Green Returns,” *Journal of Financial Economics*, 2022, 146 (2), 403–424.
- Petersen, Mitchell A.**, “Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches,” *Review of Financial Studies*, 2009, 22 (1), 435–480.
- Reimers, Nils and Iryna Gurevych**, “Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks,” in “Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)” 2019.
- Shiller, Robert J.**, “Narrative Economics,” *American Economic Review*, 2017, 107 (4), 967–1004.
- Tetlock, Paul C.**, “Giving Content to Investor Sentiment: The Role of Media in the Stock Market,” *Journal of Finance*, 2007, 62 (3), 1139–1168.
- , **Maytal Saar-Tsechansky, and Sofus Macskassy**, “More Than Words: Quantifying Language to Measure Firms’ Fundamentals,” *Journal of Finance*, 2008, 63 (3), 1437–1467.
- Webb, Michael**, “The Impact of Artificial Intelligence on the Labor Market,” Technical Report, Stanford University Working Paper 2020.

Appendices

A. Variable Definitions

Table A1. Variable Definitions

Variable	Definition
AI Adopt (ai_adoption_wmean)	Section-weighted mean cosine similarity of a firm’s 10-K text chunks to the AI-internal-adoption anchor in fiscal year Y .
AI Infra (ai_infrastructure_wmean)	Section-weighted mean similarity to the AI-compute-infrastructure anchor.
AI Risk (ai_disruption_risk_wmean)	Section-weighted mean similarity to the AI-product-disruption-risk anchor.
AI Compos	$\frac{1}{2}[z(\text{AI Adopt}) + z(\text{AI Infra})]$, with z the within-year cross-sectional standardization.
Forward return	Buy-and-hold simple return from July of $Y+1$ to June of $Y+2$ (adjusted close, dividends reinvested).
Market-adjusted return	Forward return minus the contemporaneous SPY return.
Size	Natural logarithm of the market-cap proxy (latest shares outstanding $\times$ price) at formation.
Momentum	Own return over the 12 months ending at formation, skipping the last month.
Volatility	Annualized standard deviation of monthly returns over the trailing 12 months.
Post	Indicator equal to one for formation (fiscal) years ≥ 2022 .
CAR	Cumulative market-model abnormal return; market model estimated Jan–Oct 2022, event window Nov 2022–Jun 2024.

B. Data Sources and Reproducibility

Semantic factors are from the EDGAR Research Database (UQO), which embeds 10-K text with the all-MiniLM-L6-v2 sentence transformer and scores each chunk by cosine similarity to curated anchor passages. Equity prices (monthly and daily) are from Yahoo Finance; the Fama–French five factors, the momentum factor, and the risk-free rate are from the Kenneth R. French data library. All return tests use look-ahead-free July-formation forward returns, with signals taken from the 10-K of the preceding fiscal year. The analysis pipeline downloads prices, builds the merged firm-year panel, runs the portfolio sorts, factor regressions, panel and Fama–MacBeth estimators, the difference-in-differences specifications, and the market-model event study, and emits all figures and tables programmatically; numerical results are serialized to machine-readable result files so that every number reported in the paper is reproducible from source.