UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

DÉPARTEMENT DES SCIENCES ADMINISTRATIVES

WORKING PAPER NO. 0002

Research funded by the FIRC (Fonds institutionnel de recherche et de création), Université du Québec en Outaouais.

Does the Language of Artificial Intelligence Move Markets?

A Semantic Analysis of Russell 3000 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

October 28, 2026

Version 2.0 (Russell 3000 extension)

Abstract

We measure artificial-intelligence (AI) engagement in the SEC filings of some 2,500 U.S. listed firms—the current S&P 500 constituents and the holdings of the iShares Russell 3000 ETF—by embedding every passage of their Form 10-K, 10-Q and 8-K filings of 2010–2026 with a sentence transformer and scoring it against curated AI anchor concepts (internal adoption, infrastructure exposure, product-disruption risk). The signal of filing year Y is linked to the forward twelve-month return from July of $Y+1$, so that it predates the holding period by at least six months. The panel of 27,468 firm-years (2,379 firms) extends our earlier S&P 500 study, which rested on current reports only, to annual and quarterly reports and to the mid- and small-capitalization firms that make up most of the Russell 3000, and it adds a second, independent measure of AI engagement: the AI, data-center, cloud and automation projects that a large-language-model reader extracts from the same filings. Three results stand out. First, in the broad universe AI language carries a modest but persistent cross-sectional premium over the whole sample: the equal-weighted top-minus-bottom quintile spread is +5.5 percentage points per year ($t = 2.41$) and the Fama–MacBeth slope +3.07 points per standard deviation ($t = 2.80$), with a significant within-sector component. Second, the sharp post-ChatGPT increase documented for the S&P 500—which we replicate within the new panel (+5.0 to +15.0 points)—does not generalize: in the Russell 3000 the post-launch spread is +7.9 points against +4.9 before, the 2022 cohort is negative, a placebo over ten candidate break years ranks 2022 8th, and the monthly long–short earns no post-launch alpha. The increase is confined to the largest thousand firms (+2.9 to +12.8 points), whereas the smaller firms earned a premium of about +9.6 points before the launch and +8.4 after; the ChatGPT event study shows the same size gradient (+19.5 versus +9.9 points of one-year abnormal return). Third, disclosed AI projects carry no return premium, while the language of annual reports carries the strongest one. The evidence is consistent with a persistent, partly within-industry premium on AI exposure in the broad market and with a large-cap re-rating after 2022, not with a market-wide regime change.

Keywords: textual analysis, sentence embeddings, artificial intelligence, disclosure, cross-section of returns, Russell 3000, firm size, 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 three years following the public launch of ChatGPT in late 2022, "AI" became the most frequently invoked theme 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. This paper asks whether the *language* of AI in firms' mandatory disclosures carried information about their subsequent stock returns, when that information was priced, and for which firms. It is the second version of a study whose first version answered the question for the 500 current constituents of the S&P 500 and concluded that AI language was barely priced for a decade and strongly priced in one post-launch cohort, mostly through the technology sector. The present version asks whether that conclusion survives when the laboratory is enlarged in three directions at once: to some 2,500 firms spanning the capitalization spectrum of the Russell 3000, to the full text of annual and quarterly reports rather than current reports alone, and to a second measure of AI engagement that reads what firms say they *do* rather than how they say it.

Answering the question requires three ingredients. The first is a large, machine-readable corpus of corporate disclosure: the UQO EDGAR Research Database, extended in October 2026 to an issuer universe that merges the S&P 500 snapshot with the holdings of the iShares Russell 3000 ETF, holds 527,548 Form 10-K, 10-Q and 8-K filings of 2,499 issuers filed since January 2010, all of them parsed into sections and 512-word chunks. The second is a measurement technology that captures the meaning of text rather than the frequency of keywords: every chunk is embedded with the `all-MiniLM-L6-v2` sentence transformer (Reimers and Gurevych, 2019) and scored by cosine similarity to written anchor passages that define three facets of AI engagement—internal adoption, infrastructure exposure, and product-disruption risk—and five placebo concepts. The third is a look-ahead-free mapping from disclosure to tradeable returns: the signal of filing year Y forms portfolios at the end of June of $Y+1$, and the forward twelve-month return runs from July of $Y+1$ to June of $Y+2$. The resulting estimation sample has 27,468 firm-years for 2,379 firms over formation years 2010–2024, of which 6,793 firm-years belong to S&P 500 members, so that every result of the earlier study can be reproduced inside the new panel and placed next to its broad-universe counterpart.

Three findings emerge, and each revises the earlier conclusion. First, in the broad universe AI language carries a modest but persistent cross-sectional premium over the entire sample rather than a premium that appears only after 2022. The equal-weighted top-minus-bottom quintile spread on the AI composite is +5.5 percentage points per year ($t = 2.41$) over fifteen formation cohorts, the Fama–MacBeth slope is +3.07 points per standard deviation ($t = 2.80$), and a within-sector slope of +1.46 points ($t = 4.38$) survives GICS-sector dummies. Among the 1,403 firms outside the largest

thousand the spread is +9.4 points ($t = 2.54$) and it is as large before the launch (+9.6) as after (+8.4). Second, the post-ChatGPT increase documented for the S&P 500 replicates within the new panel—the spread among index members rises from +5.0 to +15.0 points, with the 2024 cohort again dominant—but it does not generalize to the market. In the Russell 3000 the post-launch spread is +7.9 points against +4.9 before (t of the post mean 1.34), the 2022 cohort earned a negative spread, dropping the 2024 cohort leaves a post-period mean below the pre-period one, a placebo over ten candidate break years ranks 2022 8th of 10 (the largest break candidate is 2016), and the monthly long–short portfolio earns a post-launch six-factor alpha of -0.3% per year. The increase is a large-cap phenomenon: among the largest thousand firms the spread rises from +2.9 to +12.8 points ($t = 2.09$), and in the ChatGPT event study the largest thousand firms in the top tercile of pre-launch AI language earned +19.5 points more one-year abnormal return than the bottom tercile ($t = 5.09$) against +9.9 ($t = 1.55$) for the remaining firms. Third, the two new measurement dimensions matter. The signal built from annual reports alone, which the earlier study could not compute, is the strongest of the three form-specific signals (full-sample spread +16.7 points after the launch, Fama–MacBeth slope +6.06 per standard deviation in a horse race against the quarterly- and current-report signals), whereas the count of AI, data-center, cloud and automation *projects* extracted from the same filings by the SPID database carries no premium at all (-1.9 and -1.3 points before and after the launch).

Taken together, the evidence is consistent with a persistent and partly within-industry premium on AI exposure in the broad market, present throughout the 2010s among smaller firms, together with a re-rating of the largest, most-followed AI-exposed firms after 2022 that the S&P 500 sample isolated and that an attention-based account of the post-launch period would not predict. The sector channel that dominated the earlier study is still visible—the long–short loads 0.38 on a technology-sector factor after the launch, sector-by-year fixed effects absorb the post-launch cross-sectional slope entirely, and cybersecurity language is rewarded as much as AI language—but in the broad universe it coexists with a within-sector premium that the index sample was too small and too homogeneous to detect.

Contribution. The paper makes three contributions. First, substantively, it provides the first look-ahead-free evidence on the pricing of AI disclosure language across the full capitalization spectrum of the U.S. market, and shows that the headline post-ChatGPT result of the S&P 500 literature is a large-cap result. Second, methodologically, it demonstrates that embedding-based disclosure measurement at the scale of 2,500 firms and 527,000 filings is feasible, documents how the signal depends on the form of the filing it is computed from, and shows that a language measure and a project-count measure extracted from the same text can have opposite pricing implications. Third, it provides a template for disciplining regime claims built on short post-event samples—

cohort-by-cohort and leave-one-out estimates, pooled firm-clustered tests, placebo break dates, sector-factor spanning, winsorization and dependence-robust event-study inference—and applies it symmetrically to the index sample and to the broad universe. The remainder of the paper is organized as follows. Section 2 situates the study and states the hypotheses. Section 3 describes the corpus, the universe, the semantic measures and the timing convention. Section 4 lays out the designs. Section 5 presents the main results for both universes, Section 6 the robustness and identification tests, Section 7 the interpretation 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 behavioural mechanism, and a comparison case—and a sixth, methodological strand disciplines the inference. 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. 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; Loughran and McDonald (2016) and Loughran and McDonald (2020) survey the field and its pitfalls, including the sensitivity of results to document selection and pre-processing. A second generation moved from counting words to modelling them: Gentzkow et al. (2019) review text-as-data methods in economics, Ke et al. (2019) develop a supervised topic model that predicts returns from news, Kelly et al. (2021) propose a sparse text-regression framework, and Bybee et al. (2024) recover a "state of the economy" from business news. A third generation uses distributed representations of meaning. Hanley and Hoberg (2019) apply word embeddings to bank 10-Ks to identify emerging risks before they are named in dictionaries; Li et al. (2021) measure corporate culture from earnings calls with word embeddings; Huang et al. (2023) show that a domain-adapted transformer outperforms dictionaries and bag-of-words classifiers on financial text; and Cong et al. (2024) build interpretable "textual factors" from embeddings. Our paper belongs to this embedding-based generation. The limitations of keyword counting are acute for a theme like AI: a count cannot distinguish "we do not currently use artificial intelligence" from "we are an artificial-intelligence leader," it misses the paraphrase of a firm that describes machine

learning without writing "AI," and it treats boilerplate and substance identically. We instead measure the semantic similarity of each passage to anchor concepts with a sentence transformer (Reimers and Gurevych, 2019), a design close in spirit to the similarity-to-reference-text constructions of Hanley and Hoberg (2019). The literature also warns, however, that text-based measures inherit the composition of the corpus they are computed on (Loughran and McDonald, 2016, 2020); Section 3 therefore documents that composition in detail.

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 annual reports relative to the prior year 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 that carry asset-pricing content. Lopez-Lira (2023) extracts priced risk exposures from 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 concept-level measure—AI engagement—computed from the annual, quarterly and current reports of a broad universe of U.S. listed firms, and we characterize how its relation to returns varies over time and across firm size, dimensions 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 postings and résumés and document faster growth among AI-adopting firms; Acemoglu et al. (2022) and Webb (2020) measure the exposure of occupations and establishments to AI; Acemoglu and Restrepo (2020) study the labour-market effects of industrial robots. Closest to us, Eisfeldt et al. (2023) construct a firm-level measure of *generative*-AI exposure from the task content of occupations and show that a long–short portfolio on that measure earned large returns in the weeks after ChatGPT's release; Jha et al. (2024) use ChatGPT itself to extract expected investment and employment policies from earnings calls; and Lopez-Lira and Tang (2026) show that a large language model can rank news headlines in a way that predicts next-day returns. Our contribution relative to this work is to measure AI exposure from the firm's own disclosure language, available for some 2,500 firms at annual frequency, and to characterize the full time profile of its

pricing across fifteen cohorts rather than an event window. Where Eisfeldt et al. (2023) document a sharp repricing in the weeks around the launch, we ask whether a persistent, text-based exposure measure carried a premium before the launch and whether the premium after it is distinguishable from a technology-sector effect.

2.4. Investor attention, salience, and narrative economics

Why might value-relevant disclosure be priced only after a catalyst? The behavioural literature offers a mechanism: limited investor attention. Barber and Odean (2008) show that individual investors are drawn to attention-grabbing stocks; Da et al. (2011) construct an 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. These mechanisms predict that a salient consumer event such as ChatGPT could draw attention to information that had been sitting in filings for years. They do not, however, distinguish a firm-level attention shock from a sector-level revision of expected cash flows, and our tests are designed to separate the two.

2.5. ESG and climate disclosure as a comparison case

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, and Sautner et al. (2023) build a firm-level climate-exposure measure from earnings-call language that is priced in options and returns. Consistent with the post-2021 reversal in this literature, we find that high-ESG and high-climate disclosure earned negative spreads after 2022, which helps to rule out a generic "more disclosure earns more" explanation of the AI result.

2.6. Short samples, data snooping, and event-study inference

Two methodological literatures bear directly on how much weight our post-2022 estimates can carry. The first concerns the credibility of cross-sectional return patterns. Harvey et al. (2016) argue that the multiplicity of tested predictors requires t-statistics well above conventional thresholds; Hou et al. (2020) find that most published anomalies fail to replicate under standard procedures; and McLean and Pontiff (2016) document post-publication decay. A spread estimated on three annual

observations, one of which is extreme, must be read in that light. The second concerns inference in event studies when all firms share a single event date, so that abnormal returns are cross-sectionally correlated (Brown and Warner, 1985; Kolari and Pynnönen, 2010) and event-induced variance may differ from estimation-window variance (Boehmer et al., 1991), and in long-horizon settings, where the choice of benchmark and the compounding of abnormal returns matter (Barber and Lyon, 1997; Fama, 1998; Mitchell and Stafford, 2000; Kothari and Warner, 2007). Our event-study inference follows the portfolio approach recommended in that literature. Finally, the use of a current-membership universe raises the survivorship concerns analysed by Brown et al. (1995), which we address with restrictions on listing history.

2.7. 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 2010–2024 formation 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, and the increase is visible in each post-launch cohort rather than in a single one.

H3 (AI-specificity).

If the post-2022 spread reflects an AI repricing rather than a mechanical or market-wide artefact, it appears for AI-related semantic factors but not for unrelated factors such as ESG, climate, or macro risk, and it is not spanned by a technology-sector factor.

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, attenuates under firm fixed effects, and displays no pre-trend when exposure is held fixed at its pre-launch level.

We add a fifth hypothesis that the broad universe makes testable.

H5 (Size and attention).

If the post-launch repricing reflects limited attention rather than a revision of sector cash flows, it is stronger among smaller, less-followed firms than among the largest thousand, and it is visible in the firm's own disclosed AI projects as well as in its language.

Section 7 summarizes what the Russell 3000 evidence says about each hypothesis.

3. Data and Variable Construction

3.1. The scored filing corpus

The backbone of the study is the UQO EDGAR Research Database, an infrastructure built by the authors that downloads, parses and semantically scores SEC filings. Its first release (June 2026) covered the 500 current S&P 500 constituents and had scored only their Form 8-K current reports and the Form 10-Q reports of 84 firms. On 27–28 October 2026 the database was rebuilt on a broad issuer universe (Section 3.4) with every Form 10-K, 10-Q and 8-K filed from January 2010 to 27 October 2026: 527,548 filings of 2,499 issuers, of which all but sixteen were parsed. The parsing layer splits each document into its canonical sections where a section template exists (the annual report into Items 1 to 9B, the quarterly report into its financial statements, MD&A, quantitative disclosures, controls, legal proceedings, risk factors and unregistered sales; the current report is kept as a single body) and then into overlapping chunks of about 500 words. The embedding layer maps every chunk—10720890 in all—to a 384-dimensional vector with the `all-MiniLM-L6-v2` sentence transformer (Reimers and Gurevych, 2019). Table 1 describes the corpus. Unlike the earlier panel, the firm-year signal now draws on all three forms for essentially every firm-year; a typical firm-year contains about fifteen filings and three hundred chunks. The companion methodological paper (UQO Working Paper No. 0001) documents the acquisition, parsing and extraction pipeline, the universe reconciliation and the SPID project layer in detail.

Table 1. Corpus composition. Filings downloaded and parsed by form for the issuer universe, and the share of estimation-sample firm-years whose signal includes at least one scored filing of that form.

Form	Filings 2010–2026	Issuers	Parsed	Firm-years with a scored filing (estimation sample)
10-K	31,775	2,468	31,772	96.8%
10-Q	95,167	2,494	95,155	99.6%
8-K	400,606	2,499	400,605	99.9%
Total	527,548	2,499	527,532	

3.2. Anchor concepts and semantic scores

Each of eight economic concepts is defined by three short paragraphs written in the register of a filing; the full texts are reproduced in Appendix B. The AI-adoption anchor, for instance, describes a firm that "internally uses artificial intelligence, machine learning, generative AI, large language models, automation, predictive analytics … to improve productivity, operations, efficiency, decision-making." The three paragraphs of a concept are embedded, averaged and normalized, so that each concept a is represented by a unit vector $\bar{\mathbf{a}}_a$. The score of chunk c on concept a is the cosine

similarity

$$s_{c,a} = \frac{\mathbf{e}_c' \bar{\mathbf{a}}_a}{\|\mathbf{e}_c\| \, \|\bar{\mathbf{a}}_a\|} \in [-1, 1], \tag{1}$$

where $\mathbf{e}_c$ is the embedding of the chunk. Because similarity is computed in the embedding space, a passage on "deploying neural networks to forecast demand" scores highly on AI adoption even though it shares no keyword with the anchor. The eight concepts comprise three AI facets—*AI adoption*, *AI infrastructure* and *AI disruption risk*—and five placebos: *ESG commitment*, *climate risk*, *cybersecurity*, *automation/labour* and *macro risk*.

Chunk scores are aggregated to the filing level in three ways. The simple mean averages all chunks of filing f; the top-5% tail averages the chunks at or above the 95th percentile of the filing's scores; and the section-weighted mean, our headline measure, weights chunks by the section in which they appear,

$$\bar{s}^{w}_{f,a} = \frac{\sum_{c \in f} w_{\sigma(c),a} \, s_{c,a}}{\sum_{c \in f} w_{\sigma(c),a}}, \tag{2}$$

where $\sigma(c)$ is the section of chunk c and the weights emphasize MD&A and risk-factor text for the relevant concept (for AI adoption and infrastructure, MD&A receives weight 2.0 and risk factors 1.2; for AI disruption risk, risk factors 2.0 and MD&A 1.5; financial statements 0.5; the undivided body of an 8-K 1.0). Filing-level scores are averaged, without weights, over all scored filings dated in calendar year Y:

$$\mathrm{AI}_{i,Y,a} = \frac{1}{|\mathscr{F}_{i,Y}|} \sum_{f \in \mathscr{F}_{i,Y}} \bar{s}^{w}_{f,a}. \tag{3}$$

Because a firm-year now contains annual, quarterly and current reports, equation (3) blends three kinds of text; Section 6.6 rebuilds the signal separately from each form and shows that they are far from interchangeable. Our three primary signals are `ai_adoption_wmean` (**AI Adopt**), `ai_infrastructure_wmean` (**AI Infra**) and `ai_disruption_risk_wmean` (**AI Risk**); we also form a standardized composite, $\text{AI Compos} = \frac{1}{2}[z(\text{AI Adopt}) + z(\text{AI Infra})]$.

3.3. A second measure: disclosed AI projects

The SPID project database, built on the same corpus, submits every eligible filing section to a large-language-model reader with a strict JSON schema and extracts discrete corporate projects—named initiatives with a type drawn from a 21-category taxonomy, a status, cited years, locations, partners and amounts. We use two firm-year counts from it: the number of mentions of technology projects (AI initiatives, data centers, cloud migrations, automation, digital transformation, ERP implementations and technology deployments; **SPID Tech**, entered as $\log(1+n)$ and standardized within year), and an indicator for at least one AI-initiative project or AI-tagged technology (**Any AI project**). The two measures capture different objects: the semantic score measures how much

a firm's text *resembles* AI language, the project count how many concrete AI-related initiatives it *describes*. Projects are sparse: only 2.0% of firm-years contain an AI-initiative mention (4.2% in the post-launch cohorts), and about a quarter contain any technology project, so the SPID Tech quintile sort effectively separates firms that describe such projects from those that do not.

3.4. Universe and survivorship

The universe is the union of the S&P 500 membership of May 2026 and the equity holdings of the iShares Russell 3000 ETF as of 23 October 2026, mapped to SEC registrants: 2,564 issuers, of which 2,499 file the three forms we use (the remainder are foreign private issuers filing 20-F or 40-F). The ETF holdings are a current-membership proxy for the Russell 3000, not a point-in-time constituent list, and the panel therefore carries the survivorship and look-ahead biases of a current universe (Brown et al., 1995): firms that were delisted before 2026 are absent, and 1,016 of the 2,379 firms in the estimation sample have price histories that begin after July 2009 (they earn +11.0 points more per year than the 1,363 seasoned firms). The bias is more severe than in the S&P 500 sample because small firms exit more often; Section 6.7 shows that entrants are spread evenly across the AI quintiles and that the results hold among seasoned firms. Within the universe we distinguish two overlapping partitions. The *index partition* separates the 490 S&P 500 members with a valid forward return (6,793 firm-years) from the 1,889 non-members, so that the earlier study is nested in the present one. The *size partition*, re-assigned each formation year, separates the largest thousand firms by market capitalization (15,000 firm-years, 1,593 distinct firms), a Russell 1000 proxy, from the remaining firms (10,300 firm-years, 1,403 firms), a Russell 2000 proxy.

3.5. Price, share and factor data

We download monthly adjusted-close prices for all tickers from Yahoo Finance over June 2009 to October 2026 (2,499 tickers with at least one monthly return), together with the SPDR S&P 500 ETF (SPY) and the iShares Russell 3000 ETF (IWV) as benchmarks, and the Fama–French five factors (Fama and French, 2015), the momentum factor (Carhart, 1997) and the risk-free rate from the Kenneth R. French data library (through September 2026). Monthly simple returns are computed from adjusted prices, which incorporate dividends and splits; a handful of non-positive prices and impossible returns are set to missing. Market capitalization at formation is the price times the number of shares outstanding most recently reported on the cover of a 10-K or 10-Q filed on or before the formation date, taken from the SEC's XBRL company-facts API; this replaces the constant-shares proxy of the earlier study and is available for every firm. Daily prices from June 2021 to June 2024 are used for the event study, with IWV as the market index.

3.6. Timing and the construction of tradeable returns

Every filing contributing to the signal of year Y is dated between 1 January and 31 December of Y and is public on its filing date. We follow the portfolio-formation convention of Fama and French (2015): signals of year Y form portfolios at the end of June of $Y+1$, and the forward return is the buy-and-hold return from July of $Y+1$ through June of $Y+2$,

$$r_{i,Y+1} = \prod_{m=\text{Jul}\,Y+1}^{\text{Jun}\,Y+2} \left(1 + r_{i,m}\right) - 1, \tag{4}$$

computed when at least six monthly returns are available. The gap between the last filing of year Y and portfolio formation is at least six months, which rules out look-ahead bias. The signal of the 2022 cohort includes filings dated after the ChatGPT launch of 30 November 2022; the difference-in-differences design of Section 6, which fixes exposure at its 2019–2021 level, is immune to this. The holding period of the 2021 cohort straddles the launch; we treat it as pre-launch and show that moving it does not matter. After requiring a valid forward return we obtain 27,468 firm-years (2,379 firms) over formation years 2010–2024; the forward return of the 2024 cohort runs to June 2026, and filing years 2025–2026 contribute signals but no forward return.

3.7. Controls and standardization

For each formation year we compute cross-sectional z-scores of every signal,

$$z(x_{i,Y}) = \frac{x_{i,Y} - \bar{x}_Y}{\text{sd}_Y(x)}, \tag{5}$$

so that coefficients are expressed per cross-sectional standard deviation and the secular rise of AI language is removed. In the S&P 500 analyses the standardization is redone within the index sample, so that each panel of the two-sample tables is a self-contained replication. We control for *size* (log market capitalization at formation), *momentum* (the own return over the twelve months ending at formation, skipping the last month) and *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; portfolio sorts use raw returns and report winsorized spreads as a check. Appendix A gives precise definitions.

3.8. Summary statistics

Table 2 reports summary statistics for the estimation sample and Table 3 the correlations of the standardized signals. The three AI signals are collinear (the correlation of adoption with infrastructure is 0.79 and with disruption risk 0.82), three lenses on one latent construct that we also

summarize with their first principal component (84.3% of common variance). The SPID project count is only weakly related to the language scores (correlation 0.07 with AI adoption): firms that describe concrete technology projects are not, in the main, the firms whose text most resembles the AI anchors. The signal is persistent: the year-over-year cross-sectional rank correlation of the composite is 0.87 and at a five-year lag 0.72, which explains why firm fixed effects absorb most of it and motivates the pre-determined-exposure design.

Table 2. Summary statistics, estimation sample. Firm-years with a valid forward return, formation years 2010–2024. Semantic signals are section-weighted cosine similarities; returns are simple twelve-month buy-and-hold returns.

Variable	N	Mean	SD	P5	Median	P95
AI adoption (wmean)	27,468	0.0506	0.0235	0.0136	0.0498	0.0902
AI infrastructure (wmean)	27,468	-0.0027	0.0264	-0.0445	-0.0038	0.0423
AI disruption risk (wmean)	27,468	0.0420	0.0273	0.0017	0.0395	0.0902
SPID technology-project mentions	27,468	0.5575	1.3904	0.0000	0.0000	3.0000
Any AI-initiative mention (0/1)	27,468	0.0197	0.1388	0.0000	0.0000	0.0000
Forward 12-m return	27,468	0.2129	1.2105	-0.4712	0.1038	1.0628
Forward return, SPY-adjusted	27,468	0.0609	1.1942	-0.5929	-0.0387	0.8365
Momentum (12 m)	27,325	0.1700	1.0445	-0.4751	0.0900	0.9394
Volatility (12 m, ann.)	27,369	0.4020	1.9074	0.1434	0.3148	0.8589
log market cap	25,232	21.5704	1.8896	18.7766	21.5007	24.7208
Chunks per firm-year	27,468	330.1790	153.5281	145.0000	305.0000	598.6500

Table 3. Correlations of within-year standardized variables.

	AI adoption	AI infrastructure	AI disruption risk	SPID tech projects	Size	Momentum	Fo
AI adoption	1.00						
AI infrastructure	0.79	1.00					
AI disruption risk	0.82	0.69	1.00				
SPID tech projects	0.07	0.04	0.04	1.00			
Size	0.12	0.10	0.04	0.04	1.00		
Momentum	0.03	0.02	0.02	-0.01	0.09	1.00	
Forward return	0.01	0.02	0.03	-0.00	-0.10	-0.04	

3.9. The secular rise of AI language and the firms that score highest

Figure 1 plots the cross-sectional mean of the AI-adoption score by filing year for all firms, for the largest thousand and for the rest. AI language rises through the 2010s and inflects after 2022 in every size group, and larger firms score higher throughout; the within-year design compares each firm to its contemporaneous peers and removes both facts. Table 4 lists the firms with the highest

average composite over at least five years: a mixture of enterprise-technology and software firms, data-intensive services, and industrial firms, with the smaller technology names of the Russell 2000 well represented.

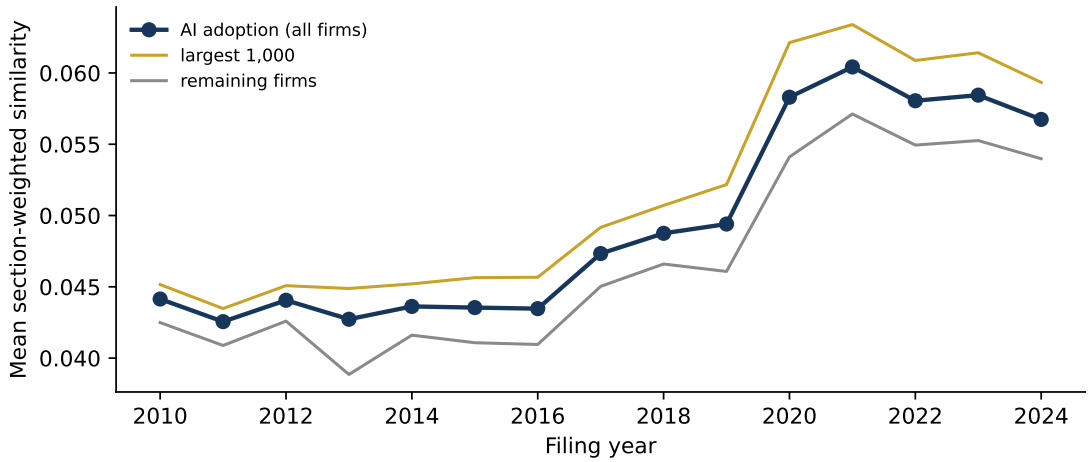


Figure 1. The secular rise of AI language. Cross-sectional mean of the section-weighted AI-adoption score by filing year, for all firms in the estimation sample, the largest thousand by market capitalization and the remaining firms.

Table 4. Firms with the highest average AI composite. Mean within-year standardized composite over the firm's formation years (at least five).

Ticker	Company	Sector	Mean composite (z)	Years
NET	Cloudflare Class A	Information Technology	3.45	6
HPE	Hewlett Packard Enterprise	Information Technology	3.34	10
IBM	Ibm	Information Technology	3.30	15
IDA	Idacorp	Utilities	3.26	15
INTC	Intel	Information Technology	3.24	15
AI	C3 Ai Class A	Information Technology	3.23	5
TYL	Tyler Technologies	Information Technology	2.79	15
ADSK	Autodesk	Information Technology	2.72	15
NSIT	Insight Enterprises Inc	Information Technology	2.66	15
UBER	Uber	Industrials	2.63	6
HE	Hawaiian Electric Industries	Utilities	2.61	15
CMI	Cummins	Industrials	2.53	15

4. Methodology

We bring six research designs to bear, ranging from portfolio sorts to a market-model event study, and we apply each of them to the full universe, to the largest thousand firms, and to the remaining smaller

firms. A seventh element, new to this version, is a second, independent measure of AI engagement: the number of technology projects (AI initiatives, data centers, cloud migrations, automation, digital transformation, ERP and technology deployments) that the SPID project database extracts from the same filings with a large-language-model reader, which we use both as a stand-alone signal and as a control. Each carries its own identifying assumptions and vulnerabilities: a sort is transparent but 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. A conclusion that survives all six is more credible than one resting on any one of them, but only if the inference within each design is sound, and this section is explicit about where it is fragile.

4.1. Portfolio sorts

Each formation year Y we sort firms into quintiles of a within-year standardized signal and compute the equal-weighted (EW) and value-weighted (VW) forward return of each quintile. The annual high-minus-low spread and its time-series average are

$$\mathrm{LS}_Y \;=\; \bar{r}_{Q5,Y} - \bar{r}_{Q1,Y}, \qquad \overline{\mathrm{LS}} \;=\; \frac{1}{N_Y}\sum_{Y} \mathrm{LS}_Y, \tag{6}$$

and we test $\overline{\mathrm{LS}} = 0$ with Newey–West standard errors with three lags (Newey and West, 1987) on the series of annual spreads. The appeal of the sort is that it imposes no functional form; the EW–VW comparison indicates whether the effect is concentrated in smaller index members. The EW–VW comparison indicates whether the effect is concentrated in smaller firms; because the universe now spans the largest 1,000 firms and some 1,500 smaller ones, we also report the sorts within the two size groups and a size-neutral sort in which quintiles are formed within size group and year. The limitation is the number of cohorts: the full sample has $N_Y = 15$ annual spreads, the pre-period twelve, and the post-period three. A Newey–West t-statistic computed on three observations has no useful sampling distribution, so for the post-period we report it only for completeness and rely instead on the cohort-by-cohort values, on leave-one-out means, on pooled firm-level regressions with standard errors clustered by firm, and on the monthly long–short series described next.

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

A raw spread could be compensation for bearing known risks. We build an annually rebalanced, equal-weighted, self-financing long–short portfolio (top minus bottom quintile), record its monthly

return LS_m, and estimate

$$\begin{aligned}\mathrm{LS}_m &= \alpha + \beta_M \,\mathrm{MKT}_m + \beta_S \,\mathrm{SMB}_m + \beta_H \,\mathrm{HML}_m + \beta_R \,\mathrm{RMW}_m \\ &\quad + \beta_C \,\mathrm{CMA}_m + \beta_U \,\mathrm{MOM}_m + \beta_T \,\mathrm{TECH}_m + u_m,\end{aligned} \tag{7}$$

with heteroskedasticity- and autocorrelation-consistent standard errors (three lags). We report the CAPM version, the Fama–French five factors plus momentum (FF6; Fama and French, 2015; Carhart, 1997), and FF6 augmented with a technology-sector factor TECH_m, the equal-weighted monthly return of the Information Technology constituents of the panel minus the equal-weighted return of all constituents. The sector factor is not a priced risk factor in the usual sense; it is a diagnostic that asks whether the AI long–short is distinguishable from a bet on the technology sector during the sample. The intercept α is the average monthly return beyond what the loadings explain. Because the monthly series has about 180 observations (36 in the post-period), it supports inference that the three annual post-period spreads cannot; we also report the Sharpe ratio of the series in each sub-period. A placebo on the break date complements the sub-period comparison: for every candidate first post-cohort between 2013 and 2022 we compute the difference between the mean spread of the cohorts from that year on and the mean of the earlier cohorts, and we report where 2022 ranks among the ten candidates.

4.3. Two-way fixed-effects panel regressions

We estimate

$$r_{i,Y+1} = \beta \,\mathrm{AI}_{i,Y} + \gamma' X_{i,Y} + \alpha_i + \delta_Y + \varepsilon_{i,Y}, \tag{8}$$

where $r_{i,Y+1}$ is the winsorized forward return of equation (4), $\mathrm{AI}_{i,Y}$ the standardized signal, X the standardized controls, and α_i, δ_Y firm and year fixed effects. Standard errors are double-clustered by firm and year (Petersen, 2009; Thompson, 2011; Cameron et al., 2011); with fifteen year clusters the time dimension of the clustering is thin, and we report firm-clustered versions where the distinction matters. Firm fixed effects absorb all time-invariant cross-firm differences, so β identifies whether within-firm changes in AI language predict returns; year 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 and average the slopes (Fama and MacBeth, 1973),

$$r_{i,Y+1} = a_Y + b_Y \,\mathrm{AI}_{i,Y} + c_Y' X_{i,Y} + e_{i,Y}, \qquad \hat{b} = \frac{1}{N_Y} \sum_Y \hat{b}_Y, \tag{9}$$

with Newey–West errors on the time series of $\hat{b}_Y$. 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, and adding GICS-sector dummies to the cross-sectional regression gives the within-sector version of the slope.

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

To test regime dependence (H2) we split the sample at the ChatGPT launch and compare sub-period spreads, and we estimate the pooled interaction model

$$r_{i,Y+1} = \beta_1 \,\mathrm{AI}_{i,Y} + \beta_2 \,(\mathrm{AI}_{i,Y} \times \mathrm{Post}_Y) + \gamma' X_{i,Y} + \delta \,\mathrm{Post}_Y + \varepsilon_{i,Y}, \qquad \mathrm{Post}_Y = \mathbb{1}[Y \geq 2022], \quad (10)$$

where β_2 is the change in the premium per standard deviation of the signal, with standard errors clustered by formation year. We complement this with a difference-in-differences design in which treatment is the firm's pre-determined exposure $\overline{\mathrm{AI}}_i^{19\text{–}21}$, the average of its composite over the filings of 2019–2021, standardized across firms, so that it cannot respond to the event:

$$r_{i,Y+1} = \theta \left(\overline{\mathrm{AI}}_i^{19\text{–}21} \times \mathrm{Post}_Y\right) + \gamma' X_{i,Y} + \alpha_i + \delta_Y + \varepsilon_{i,Y}, \quad (11)$$

estimated with firm and year fixed effects and double-clustered errors, and with a binary version in which treatment is above-median pre-determined exposure. An event-time version replaces the single interaction with interactions of the fixed exposure with each formation-year dummy relative to 2021,

$$r_{i,Y+1} = \sum_{\tau \neq 2021} \theta_\tau \left(\overline{\mathrm{AI}}_i^{19\text{–}21} \times \mathbb{1}[Y = \tau]\right) + \gamma' X_{i,Y} + \delta_Y + \varepsilon_{i,Y}, \quad (12)$$

which traces the dynamic response and, in the years before 2022, tests the parallel-trends assumption. Because each θ_τ is identified within a single year, standard errors for equation (12) are clustered by firm.

4.6. Market-model event study

Finally, we run an event study (MacKinlay, 1997) around the ChatGPT launch, which converts the annual-frequency portfolio evidence into a daily, cleanly dated test. The launch was a consumer-product event, unanticipated by equity markets, that raised the salience of AI without itself changing any firm's fundamentals overnight. For each stock we estimate the market model

$$r_{i,d} = a_i + b_i \, r_d^m + e_{i,d} \quad (13)$$

over the estimation window 3 January–31 October 2022 (209 trading days), with the iShares Russell 3000 ETF (IWV) as the market index, and compute abnormal and cumulative abnormal returns over the event window 1 November 2022–30 June 2024,

$$\mathrm{AR}_{i,d} = r_{i,d} - (\hat{a}_i + \hat{b}_i r_d^m), \qquad \mathrm{CAR}_i(T) = \sum_{d=1}^{T} \mathrm{AR}_{i,d}. \tag{14}$$

CARs are compared across terciles of pre-event AI exposure measured from the filings of calendar year 2021, all dated before the launch, within the full universe and within size groups, and firm-level CARs are regressed on exposure with size, beta, and sector controls. A firm-level Welch t-test treats the firms as independent, which they are not when they share an event date (Brown and Warner, 1985; Kolari and Pynnönen, 2010). We therefore also treat the equal-weighted High-minus-Low abnormal return as a single daily series AR_d^{HL} and compute

$$t_{\mathrm{CDA}}(T) = \frac{\sum_{d=1}^{T} \mathrm{AR}_d^{HL}}{\hat{\sigma}_{HL}\sqrt{T}}, \tag{15}$$

where $\hat{\sigma}_{HL}$ is the daily standard deviation of the same portfolio series in the estimation window (the crude dependence adjustment of Brown and Warner, 1985) or, to allow for event-induced variance (Boehmer et al., 1991), in the event window. We also report the average pairwise correlation of estimation-window residuals, which governs the overstatement of the independence-based test (Kolari and Pynnönen, 2010). Horizons of 250 and 395 trading days are long for an event study; at such horizons the market model's intercept, estimated in the 2022 bear market, makes all CARs drift negative, and only differences between terciles are interpretable (Fama, 1998; Kothari and Warner, 2007).

5. Results

Every table of this section reports the Russell 3000 universe and, where the comparison is informative, the S&P 500 members within it, with the standardization of the signal redone within each sample. The index panel reproduces the findings of the first version of this paper; the broad-universe panel is new.

5.1. Portfolio sorts: a modest, persistent relation

Table 5 reports quintile sorts over the fifteen formation cohorts. In the Russell 3000 the relation between AI language and future returns is positive and broadly monotone. The equal-weighted Q5−Q1 spread on the composite is +5.5 percentage points per year ($t = 2.41$), with a mean forward return of +17.5% in the bottom quintile and +23.0% in the top; the adoption, infrastructure

and disruption-risk spreads are of the same sign, the last the largest. Value-weighting *raises* the composite spread to +13.9 points ($t = 2.59$): the value-weighted bottom quintile earned only +0.4% per year, because the largest low-AI firms of the sample—energy majors, banks, utilities—lagged the market for most of the period, while the value-weighted top quintile is dominated by the technology leaders. In the S&P 500 panel the equal-weighted composite spread is +7.0 points ($t = 2.56$), close to the first version's +4.9. By the standards of Harvey et al. (2016) none of these spreads qualifies as a discovery, but in the broad universe the composite spread is significant at conventional levels over the full sample, which it was not in the index.

Table 5. Quintile sorts on AI language, Russell 3000 and S&P 500. Mean forward twelve-month return (%) by within-year quintile of each signal, equal- and value-weighted, averaged over formation cohorts 2010–2024, with the Q5−Q1 spread and its Newey–West t-statistic (three lags) on the series of annual spreads. Signals are standardized within year and within sample.

Signal	Weighting	Q1	Q2	Q3	Q4	Q5	Q5−Q1	t	Years
Panel A: Russell 3000 universe (N = 27,468 firm-years, 2,379 firms)									
AI adoption	EW	+19.6	+19.0	+20.4	+22.2	+22.0	+2.4	1.17	15
	VW	+0.5	+9.7	+3.8	+7.6	+17.2	+16.7	3.29	15
AI infrastructure	EW	+18.3	+19.4	+21.1	+21.4	+22.9	+4.6	1.99	15
	VW	-0.9	+8.2	+7.3	+11.9	+15.0	+15.9	3.43	15
AI disruption risk	EW	+17.7	+18.4	+19.5	+19.9	+27.6	+10.0	2.26	15
	VW	+8.0	+9.0	+6.1	+12.1	+12.4	+4.3	0.59	15
AI composite	EW	+17.5	+21.9	+19.3	+21.4	+23.0	+5.5	2.41	15
	VW	+0.4	+8.0	+5.3	+13.6	+14.3	+13.9	2.59	15
SPID tech projects	EW	+22.4	+18.9	+21.3	+20.0	+20.6	-1.8	-1.34	15
	VW	+9.9	+9.7	+12.5	+4.7	+11.4	+1.4	0.25	15
Panel B: S&P 500 members only (N = 6,793 firm-years, 490 firms)									
AI adoption	EW	+15.1	+18.2	+18.0	+19.0	+22.9	+7.8	2.44	15
	VW	+12.9	+10.8	+12.6	+10.3	+16.8	+3.9	1.16	15
AI infrastructure	EW	+16.2	+16.4	+18.1	+19.6	+22.8	+6.6	2.29	15
	VW	+10.7	+11.6	+13.2	+13.9	+13.9	+3.2	0.67	15
AI disruption risk	EW	+15.1	+17.5	+17.8	+18.5	+24.3	+9.2	3.61	15
	VW	+11.4	+13.1	+14.0	+15.5	+12.6	+1.2	0.30	15
AI composite	EW	+15.6	+17.2	+18.0	+19.8	+22.7	+7.0	2.56	15
	VW	+12.2	+10.5	+11.9	+15.3	+13.9	+1.7	0.38	15
SPID tech projects	EW	+18.1	+18.8	+20.1	+18.8	+17.3	-0.8	-0.66	15
	VW	+17.0	+13.5	+15.9	+13.0	+8.0	-9.0	-2.94	15

5.2. Factor-adjusted performance (H1)

Table 6 reports the monthly long–short portfolios. In the Russell 3000 the composite long–short earns +3.3% per year over 180 months (HAC $t = 1.58$, Sharpe ratio 0.43), a CAPM alpha of about the same size and a six-factor alpha of +2.0% per year ($t = 1.31$); adding the technology-sector factor, on which the portfolio loads 0.38, halves the alpha to +1.1% ($t = 0.85$). The adoption long–short earns nothing and the disruption-risk long–short the most. As in the index sample, unconditional risk-adjusted alpha is at best marginal, and part of what premium there is resembles a technology-sector tilt. The verdict on H1 is therefore the same in both universes: a positive but

economically modest and statistically fragile unconditional premium.

Table 6. Monthly long–short portfolios: mean, Sharpe ratio and factor alphas. Equal-weighted, annually rebalanced, self-financing Q5−Q1 portfolios; alphas from CAPM and Fama–French five-factor-plus-momentum (FF6) regressions with HAC standard errors (three lags), full sample and post-launch sub-period (from July 2023).

Signal	Sample	Months	Mean (%/yr)	Sharpe	FF6 α full (%/yr)	(t)	FF6 α post (%/yr) (t)
AI adoption	Russell 3000	180	-1.83	-0.09	-3.10	(-0.65)	-0.22 (-0.06)
	S&P 500	180	+4.99	0.66	+3.64	(2.12)	+2.38 (0.50)
AI infrastructure	Russell 3000	180	+2.56	0.36	+1.66	(1.12)	+0.43 (0.11)
	S&P 500	180	+4.32	0.58	+3.19	(2.04)	+3.16 (0.91)
AI disruption risk	Russell 3000	180	+2.29	0.10	+1.98	(0.40)	+12.35 (2.83)
	S&P 500	180	+6.20	0.68	+5.27	(3.07)	+2.66 (0.54)
AI composite	Russell 3000	180	+3.30	0.43	+1.98	(1.31)	-0.33 (-0.10)
	S&P 500	180	+4.53	0.58	+3.33	(2.01)	+0.94 (0.21)
SPID tech projects	Russell 3000	180	-5.28	-0.25	-6.12	(-1.30)	+2.92 (0.63)
	S&P 500	180	-1.14	-0.24	-1.56	(-1.36)	-0.57 (-0.43)

5.3. Panel and Fama–MacBeth estimates

Table 7 reports Fama–MacBeth slopes in both universes and Table 8 the two-way fixed-effects panel in the broad universe. The composite premium is +3.07 points per standard deviation in the Russell 3000 ($t = 2.80$) and +2.07 ($t = 2.17$) in the S&P 500. When the three collinear facets are entered jointly, adoption takes a negative sign and infrastructure and disruption risk positive ones in both samples: it is the compute-exposure and threat-recognition vocabulary, not the "we use AI internally" vocabulary, that carries the premium. The SPID project count has no cross-sectional premium in either sample. Size carries a strongly negative premium in the broad universe (-27.28 points per standard deviation in the panel, $t = -5.99$), momentum a negative one and volatility a positive one, as expected in a sample that reaches far into small capitalizations. In the two-way fixed-effects panel the AI coefficients are small: the composite earns +0.56 points per standard deviation within firm ($t = 0.94$). The signal lives across firms, not within them.

5.4. The post-ChatGPT change: an index result, not a market result (H2)

Table 9 is the central table of the paper. It splits the cohorts at the ChatGPT launch and reports, for each signal and each universe, the pre- and post-launch spreads, the three post-launch cohorts, the pooled firm-clustered test of the post-period spread, its winsorized version and the post-period Sharpe ratio of the monthly long–short. In the S&P 500 panel the first version's result reappears almost exactly: the composite spread rises from +5.0 to +15.0 points ($t = 2.21$), the cohorts are 2022: +13.5, 2023: -6.9, 2024: +38.4, the pooled post-period spread is +15.1 points ($t = 2.22$; +9.2 winsorized) and the post-period Sharpe ratio is 0.66. In the Russell 3000 the picture is different. The

Table 7. Fama–MacBeth regressions, Russell 3000 and S&P 500. Average of annual cross-sectional slopes of the forward return on the standardized signals and controls, in percentage points per standard deviation, with Newey–West t-statistics (three lags) on the series of annual slopes. "Joint" coefficients come from a regression that enters the three AI facets together.

Fama–MacBeth slope (pp per SD, controls: size, momentum, volatility)	Russell 3000	(t)	S&P 500	(t)
AI composite	+3.07	(2.80)	+2.07	(2.17)
AI adoption (joint)	-1.99	(-1.62)	-1.65	(-2.25)
AI infrastructure (joint)	+3.21	(1.49)	+1.84	(1.79)
AI disruption risk (joint)	+2.22	(1.22)	+2.19	(2.11)
SPID tech projects	-0.22	(-0.24)	-0.87	(-0.94)
Size	-7.49	(-6.71)	-4.01	(-11.89)
Momentum	-4.14	(-2.47)	+0.78	(1.18)
Volatility	+9.91	(2.28)	+5.79	(2.98)

Table 8. Two-way fixed-effects panel regressions, Russell 3000. Dependent variable: winsorized forward twelve-month return. All regressions include firm and year fixed effects; standard errors are double-clustered by firm and year; t-statistics in parentheses; coefficients in percentage points per standard deviation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AI adoption (z)	-0.63	+0.24			-0.33		
	(-0.99)	(0.39)			(-0.22)		
AI infrastructure (z)			+0.68		+1.69		
			(1.32)		(1.65)		
AI disruption risk (z)				-0.10	-1.24		
				(-0.13)	(-0.80)		
AI composite (z)							+0.56
							(0.94)
SPID tech projects (z)						-0.40	-0.41
						(-1.08)	(-1.11)
Size (z)		-27.28	-27.29	-27.27	-27.29	-27.26	-27.28
		(-5.99)	(-5.99)	(-5.98)	(-6.00)	(-5.98)	(-5.99)
Momentum (z)		-2.55	-2.55	-2.55	-2.56	-2.56	-2.56
		(-2.76)	(-2.76)	(-2.76)	(-2.76)	(-2.77)	(-2.77)
Volatility (z)		+3.74	+3.75	+3.74	+3.77	+3.74	+3.75
		(2.65)	(2.65)	(2.65)	(2.66)	(2.65)	(2.65)
Observations	25,135	25,135	25,135	25,135	25,135	25,135	25,135
Within R^2	0.0001	0.0595	0.0596	0.0596	0.0598	0.0596	0.0596

composite spread is +4.9 points before the launch and +7.9 after ($t = 1.34$); the 2022 cohort, the first held after the launch, earned a *negative* spread, and the post-period mean rests on 2024 (2022: -8.4, 2023: +11.8, 2024: +20.4); the pooled post-period spread is +8.2 points ($t = 1.80$) and +3.0 winsorized ($t = 1.25$); the difference between the post- and pre-period pooled spreads is +2.8 points ($t = 0.58$); and the monthly long–short earns +2.1% per year after July 2023 with a Sharpe ratio of 0.23, *lower* than its pre-launch Sharpe ratio of 0.50, and a six-factor alpha of -0.3%. Figure 2 plots the cohort spreads of the two universes side by side and Figure 3 the cumulative long–short returns: the index series rises after the launch, the broad-universe series does not.

Table 9. Before and after ChatGPT, Russell 3000 and S&P 500. Equal-weighted Q5−Q1 spreads averaged over the pre-launch cohorts (2010–2021) and the post-launch cohorts (2022–2024), with Newey–West t-statistics (the post-period t rests on three observations); the three post-launch cohort spreads; for the composite, the pooled regression of the forward return on a top-quintile indicator with year dummies and firm-clustered standard errors in the post-period sample (raw and winsorized); and the annualized Sharpe ratio of the monthly long–short from July 2023.

Signal	Sample	Pre Q5−Q1	(t)	Post Q5−Q1	(t)	Post cohorts 2022 / 2023 / 2024	Pooled post (firm-cl. t)	Winsorized	Post Sharpe
AI adoption	Russell 3000	+1.9	(0.91)	+4.5	(1.11)	-7.8 / +10.3 / +10.9			0.07
	S&P 500	+4.9	(2.42)	+19.5	(2.19)	+14.8 / -6.2 / +50.1			0.82
AI infrastructure	Russell 3000	+3.9	(1.47)	+7.5	(1.65)	-5.0 / +10.4 / +17.1			0.39
	S&P 500	+4.2	(1.68)	+16.1	(2.74)	+10.5 / +2.1 / +35.5			0.87
AI disruption risk	Russell 3000	+6.8	(2.09)	+22.8	(1.95)	-2.1 / +15.8 / +54.6			1.05
	S&P 500	+7.2	(3.20)	+17.3	(2.96)	+16.8 / -2.5 / +37.5			0.69
AI composite	Russell 3000	+4.9	(2.00)	+7.9	(1.34)	-8.3 / +11.8 / +20.4	+8.2 (1.80)	+3.0 (1.25)	0.23
	S&P 500	+5.0	(1.98)	+15.0	(2.21)	+13.5 / -6.9 / +38.4	+15.1 (2.22)	+9.2 (1.96)	0.66
SPID tech projects	Russell 3000	-1.9	(-2.64)	-1.3	(-0.18)	+1.6 / +20.1 / -25.6			0.09
	S&P 500	-0.1	(-0.10)	-3.6	(-0.75)	+5.5 / +1.2 / -17.6			-0.70

Table 10 and Figure 4 look at the broad-universe cohorts one by one. 75% of the twelve pre-launch spreads are positive, 12 of them exceed the 2022 spread, and the largest cohort spread of the sample is 2018 (+7.6 points on the pre-determined-exposure measure as well). Dropping the 2024 cohort leaves a post-period mean of +16.1, +6.0, +1.7 (leave-one-out means for 2022, 2023 and 2024 respectively), that is, below the pre-period mean of +4.9 when 2024 is removed. The placebo on the break date (Figure 5) makes the point formally: for each candidate first post-cohort between 2013 and 2022 we compute the post-minus-pre difference in mean spreads; 2022 produces a difference of +3.0 points and ranks 8th of 10 candidates, behind every candidate from 2013 to 2018. Whatever changed in the cross-section of AI language returns in the broad market, it did not change in 2022. The year-by-year Fama–MacBeth slopes of Figure 6 tell the same story: the slope is positive in most years, large in 2018–2019 and 2024, and not systematically larger after the launch.

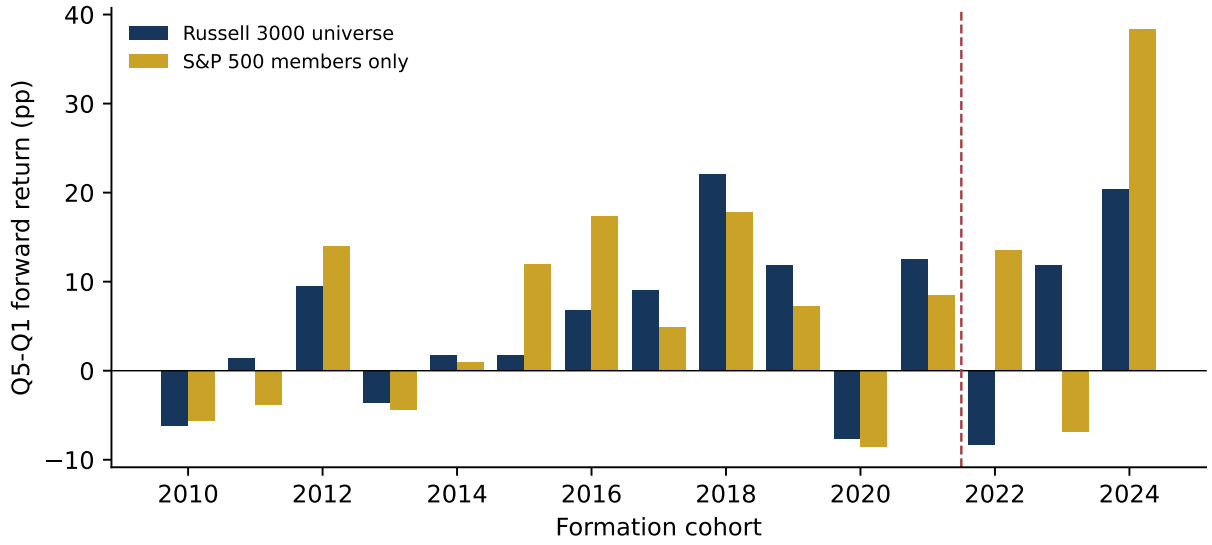


Figure 2. Annual Q5−Q1 spreads by formation cohort, Russell 3000 versus S&P 500. The dashed line marks the ChatGPT launch; cohorts to its right are held after the launch. In the broad universe the pre-launch spreads are frequently positive and the 2022 cohort is negative; in the index the post-launch average is driven by the 2024 cohort.

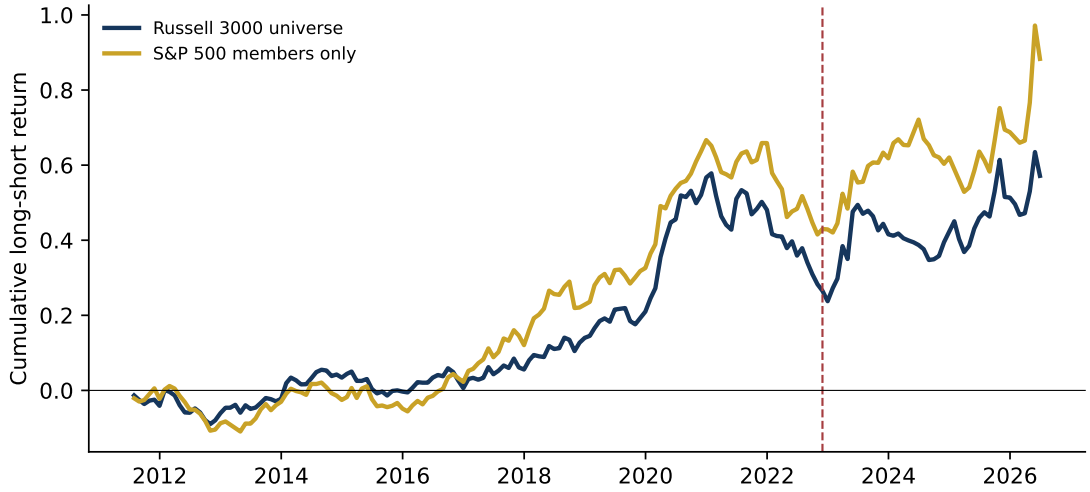


Figure 3. Cumulative long–short return on the AI composite. Equal-weighted, annually rebalanced Q5−Q1 portfolio in the Russell 3000 universe and among S&P 500 members.

Table 10. Cohort spreads and pooled tests, Russell 3000. Left: equal-weighted Q5−Q1 spread on the composite by formation cohort. Right: pooled regressions of the forward return of top- and bottom-quintile firm-years on a top-quintile indicator with year dummies and firm-clustered standard errors, and leave-one-out means of the post-period cohorts.

Cohort	Q5−Q1 (pp)	Pooled test	Estimate
2010	-6.3	Post cohorts, Q5 vs Q1 (firm-clustered t)	+8.2 (1.80), N=2,792
2011	+1.4	Post, winsorized	+3.0 (1.25)
2012	+9.5	Post 2022–2023 only	+1.8 (0.33)
2013	-3.6	Pre cohorts	+5.4 (3.90)
2014	+1.7	Difference post − pre	+2.8 (0.58)
2015	+1.6	Post mean / leave-one-out	+7.9 / +16.1, +6.0, +1.7
2016	+6.8		
2017	+9.0		
2018	+22.1		
2019	+11.8		
2020	-7.7		
2021	+12.5		
2022	-8.4		
2023	+11.8		
2024	+20.4		

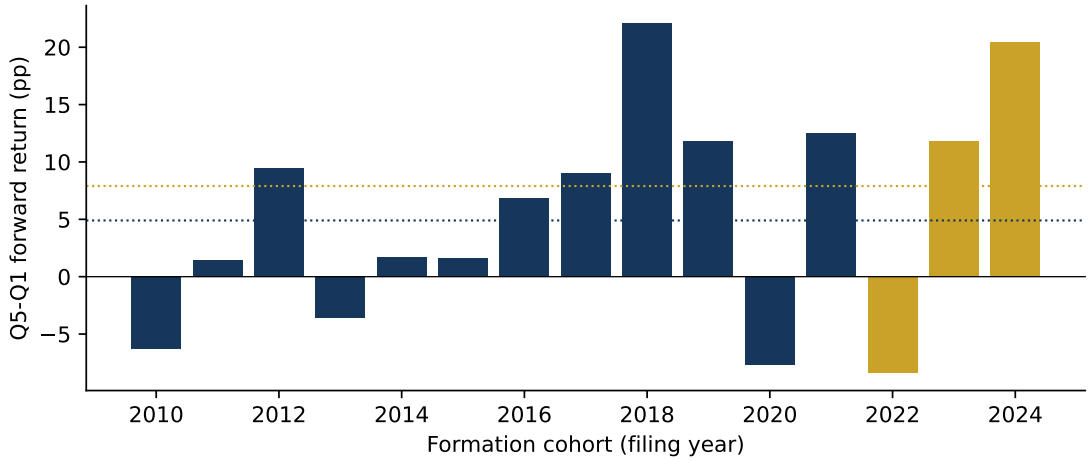


Figure 4. Annual AI-composite spreads by cohort, Russell 3000. Post-launch cohorts in gold; dotted lines mark the pre- and post-period means.

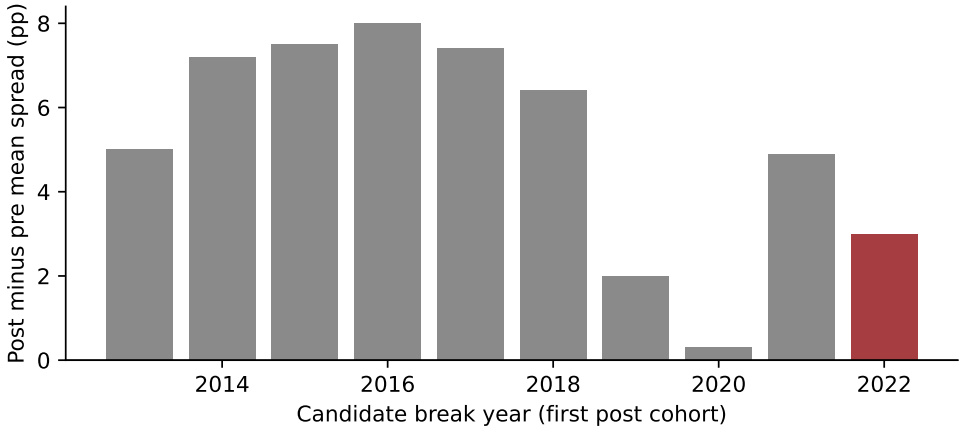


Figure 5. Placebo break dates. Difference between the mean Q5−Q1 spread of the cohorts from the candidate break year onward and the mean of the earlier cohorts, for every candidate between 2013 and 2022 (2022 in red).

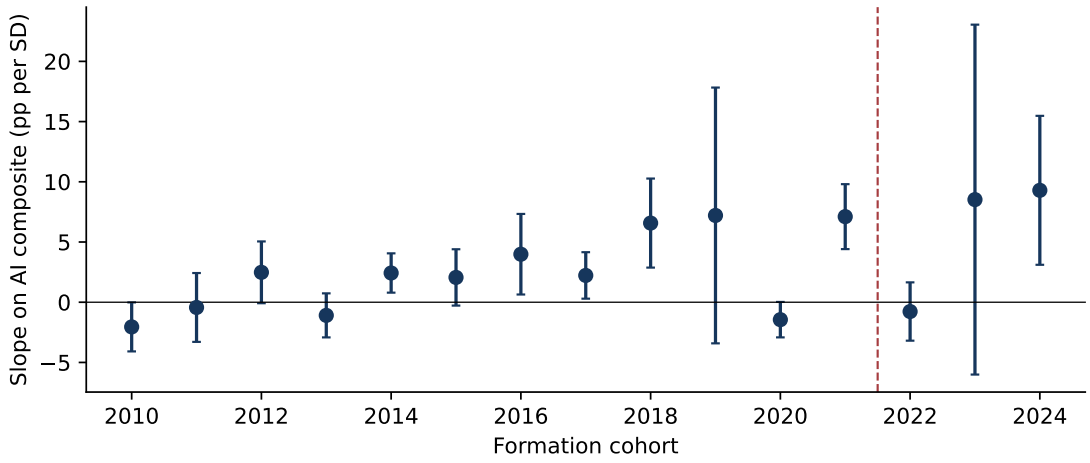


Figure 6. Year-by-year Fama–MacBeth slopes on the AI composite, Russell 3000. Annual cross-sectional slope with size, momentum and volatility controls and 95% confidence bands (heteroskedasticity-robust, within year).

5.5. Size: where the two universes part ways (H5)

Why does the index show a post-launch jump and the market not? Table 11 and Figure 7 split the universe by size and by index membership. Among the largest thousand firms—the Russell 1000 proxy—the composite spread rises from +2.9 points before the launch to +12.8 after ($t = 2.09$; cohorts 2022: +5.0, 2023: +1.0, 2024: +32.6), with a post-period Sharpe ratio of 0.61 against 0.43 before: the index result is a large-cap result, and it holds for the 500 largest and the 1,000 largest alike. Among the remaining firms the spread is +9.6 points before the launch and +8.4 after (cohorts 2022: -13.8, 2023: +26.0, 2024: +13.1); the full-sample spread of +9.4 points ($t = 2.54$) is the largest of any sub-sample, and its monthly long–short Sharpe ratio is 0.57 before the launch and 0.32 after. The same contrast appears between S&P 500 members (+5.1 to +15.0) and non-members (+4.6 to +5.7, $t = 0.90$). A size-neutral sort, in which quintiles are formed within size group and year, yields a full-sample spread of +7.2 points ($t = 2.73$), +6.2 before the launch and +11.3 after. The AI $\times$ size double sort of Table 12 and Figure 8 shows the gradient in finer detail: after the launch the high-minus-low AI gap is +17.4 points in the smallest size tercile and +12.8 in the largest, but before the launch it was already +11.7 in the smallest tercile against +2.6 in the largest. Smaller firms whose filings resembled the AI anchors earned a premium throughout the 2010s; the largest firms earned one only after 2022.

Table 11. AI-composite spreads by size group and index membership. Equal-weighted Q5−Q1 spreads with quintiles formed within each sub-sample and year; full-sample, pre- and post-launch means with Newey–West t-statistics; FF6 alpha of the sub-sample's monthly long–short after July 2023. The size-neutral row forms quintiles within size group and year in the full sample.

Sample	Firm-years	Firms	Full Q5−Q1	(t)	Pre	Post	(t)	FF6 α post (%/yr)	(t)
Largest 1,000 by market cap	15,000	1,593	+4.9	(1.92)	+2.9	+12.8	(2.09)	+2.06	(0.58)
Remaining firms	10,300	1,403	+9.4	(2.54)	+9.6	+8.4	(1.29)	+7.29	(1.37)
S&P 500 members	6,793	490	+7.0	(2.57)	+5.1	+15.0	(2.21)	+0.94	(0.21)
Non-members	20,675	1,889	+4.8	(1.88)	+4.6	+5.7	(0.90)	+1.77	(0.38)
Size-neutral quintiles (all)	27,468	2,379	+7.2	(2.73)	+6.2	+11.3	(1.97)		

Table 12. AI language $\times$ size double sort, Russell 3000. Mean forward twelve-month return (%) by AI-composite tercile within size tercile (both formed within year), before and after the launch.

Period	Size tercile	Low AI	Mid	High AI	High−Low
Pre-ChatGPT (2010–2021)	Small	21.5	38.3	33.2	+11.7
	Mid	13.1	12.6	15.4	+2.3
	Large	10.7	10.7	13.3	+2.6
Post-ChatGPT (2022–2024)	Small	49.2	40.4	66.6	+17.4
	Mid	17.7	13.9	24.1	+6.3
	Large	11.9	14.2	24.7	+12.8

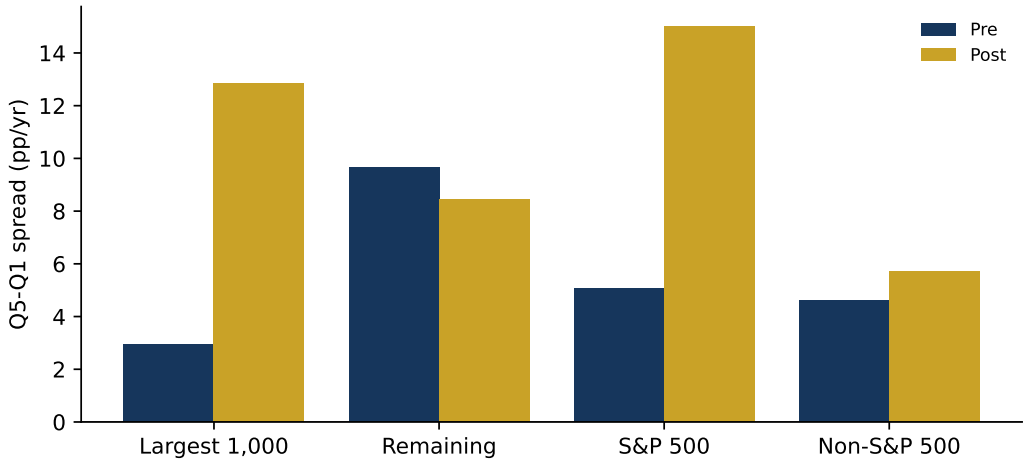


Figure 7. Pre- and post-launch spreads by size group and index membership.

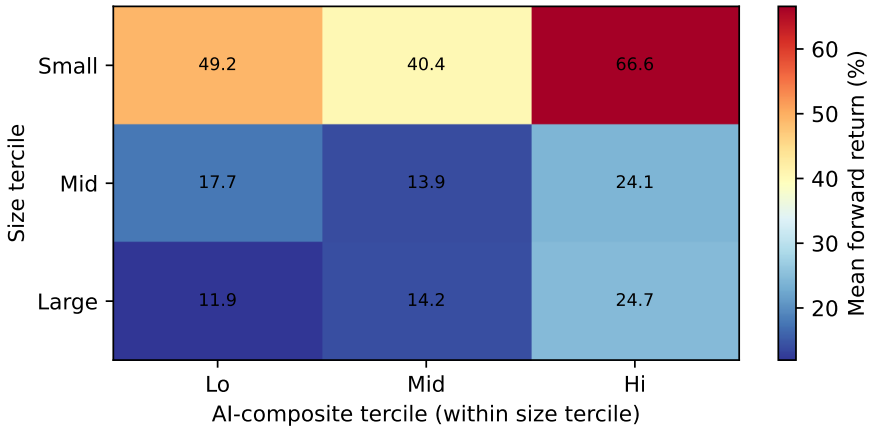


Figure 8. AI language × size double sort, post-launch cohorts. Mean forward return by AI tercile within size tercile.

5.6. Economic significance

Three numbers summarize the economic content. In the broad market a one-standard-deviation increase in AI language relative to same-year peers is associated with about three percentage points of additional forward return per year on average over fifteen cohorts, with a within-sector component of about one and a half points; this is the order of magnitude of a modest characteristic premium, not of an anomaly. The headline post-launch spread of fifteen to twenty points exists only in the largest firms and, within them, mostly in the 2024 cohort. And the monthly long–short on the composite, which rebalances every month and therefore does not let extreme winners run, earned a Sharpe ratio of 0.43 over the full sample in the broad universe, 0.50 before the launch and 0.23 after; the "Sharpe ratio above one" of the first version is a large-cap, post-2022 statistic that the broad universe does not reproduce.

6. Robustness and Identification

This section confronts the broad-universe results with the alternative explanations a sceptical reader would raise: that the premium is a sector effect, that the post-launch pattern is mechanical or specific to how the signal is built, that it reflects the composition of a current-membership universe, that the event-study inference is overstated, and that a different measure of AI engagement would tell a different story. Unless stated otherwise, all tests use the Russell 3000 universe.

6.1. Sector controls and the technology factor

Table 13 addresses the concern that AI language proxies for the technology sector. In annual cross-sectional regressions with size, momentum and volatility controls, the full-sample slope on the composite is +3.12 points per standard deviation ($t = 2.79$) and +1.46 ($t = 4.38$) with GICS-sector dummies: unlike the index sample, the broad universe shows a within-sector premium that is statistically robust over the full sample. In pooled regressions of the winsorized forward return with firm-clustered errors, sector-by-year fixed effects leave a full-sample slope of +0.88 points ($t = 2.48$). The post-launch period is different: the post-period slope of +5.81 ($t = 2.68$) falls to +1.88 with sector dummies ($t = 2.57$) and to +0.10 ($t = 0.11$) with sector-by-year fixed effects, so that the post-launch cross-sectional relation is entirely a between-sector relation, as it was in the index. In the monthly series the long–short loads 0.38 on the technology-sector factor over the full sample and 0.38 after the launch; adding the factor reduces the full-sample FF6 alpha from +2.0% to +1.1% per year and the post-launch alpha to -2.1%. The sector-neutral composite sort gives +1.5 points after the launch ($t = 0.72$), a fifth of the baseline. The within-sector premium on AI language in the broad market is therefore a feature of the whole sample, not of the post-launch regime.

Table 13. Sector controls and technology-factor spanning, Russell 3000. Top: Fama–MacBeth slopes on the composite with and without GICS-sector dummies, and pooled firm-clustered regressions with year or sector-by-year fixed effects (percentage points per standard deviation, t in parentheses). Bottom: the monthly composite long–short; TECH is the equal-weighted return of the Information Technology constituents of the panel minus the equal-weighted return of all constituents.

	Pre	Post	Full
FM slope, no sector dummies (pp/SD)	+2.44 (2.25)	+5.81 (2.68)	+3.12 (2.79)
FM slope, sector dummies	+1.35 (3.95)	+1.88 (2.57)	+1.46 (4.38)
Pooled, year FE (firm-cl.)	+1.89 (5.67)	+2.57 (3.13)	+2.00 (6.13)
Pooled, sector×year FE	+1.20 (3.12)	+0.10 (0.11)	+0.88 (2.48)
Monthly LS mean (%/yr)	3.61	2.08	3.30
Sharpe	0.50	0.23	0.43
FF6 α (%/yr)	2.19	-0.33	1.98
FF6 t	1.28	-0.10	1.31
FF6+TECH α (%/yr)	1.05	-2.10	1.06
FF6+TECH t	0.77	-0.69	0.85
β_{TECH}	0.41	0.38	0.38

6.2. Placebo: is the change specific to AI? (H3)

Table 14 and Figure 9 run all eight semantic factors through the sub-period design. After the launch the AI-disruption-risk factor earns the largest spread (+22.8 points, $t = 1.95$; +6.8 before), AI infrastructure +7.5 and AI adoption +4.5. ESG-commitment, climate-risk and macro-risk language earn large negative spreads (-15.9, -8.6 and -16.3 points), in line with the post-2021 reversal in the sustainable-investing literature (Pástor et al., 2022) and with the underperformance of rate- and cycle-sensitive firms. But cybersecurity (+13.5 points, the most significant of all) and automation/labour (+7.8) also rise, as in the index. The placebo rules out a generic "more disclosure earns more" effect and a pipeline artefact; it does not separate AI from the broader technology theme with which AI, cybersecurity and automation language travel. The interaction regressions, now estimated on 27,000 firm-years with firm-clustered errors, give the same ordering: a significant post-launch increment for disruption risk and cybersecurity, none for adoption, and significant decrements for ESG and macro language.

6.3. Event study around ChatGPT

Table 15 and Figure 10 report market-model cumulative abnormal returns (IWV as the market) by pre-event AI tercile for 2,226 firms. The High-minus-Low difference is +0.9 points at sixty trading days, +14.0 at one year and +12.2 at eighteen months; the independence-based Welch t at one year is 3.94. Table 16 re-examines the inference with the portfolio statistic of equation (15): the average pairwise correlation of estimation-window residuals is 0.037, and the dependence-

Table 14. Placebo across semantic factors, Russell 3000. Equal-weighted Q5−Q1 spreads for each factor over the full sample and the two sub-periods, and the coefficient on the factor × post interaction in the pooled regression of equation (10) with firm-clustered standard errors (percentage points per standard deviation).

Semantic factor	Full Q5−Q1	Pre	Post	(t)	Interaction	(t)
AI adoption	+2.4	+1.9	+4.5	(1.11)	-0.08	(-0.10)
AI infrastructure	+4.6	+3.9	+7.5	(1.65)	+0.87	(1.02)
AI disruption risk	+10.0	+6.8	+22.8	(1.95)	+2.36	(2.76)
ESG commitment	-5.4	-2.8	-15.9	(-3.86)	-2.71	(-3.13)
Climate risk	-0.1	+2.0	-8.6	(-6.03)	-1.16	(-1.42)
Cybersecurity	+3.1	+0.5	+13.5	(5.27)	+3.30	(4.06)
Automation/labor	+3.4	+2.3	+7.8	(2.45)	-0.95	(-1.22)
Macro risk	-5.8	-3.1	-16.3	(-2.20)	-2.33	(-2.81)

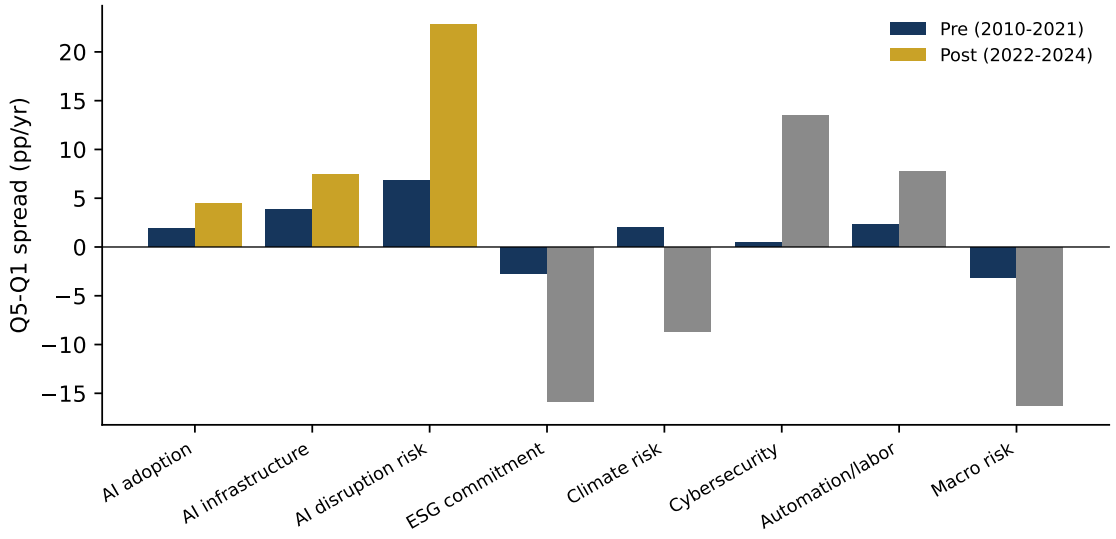


Figure 9. Placebo test across semantic factors. Pre- and post-launch Q5−Q1 spreads for the eight factors; AI factors in gold in the post period.

robust t is 0.30 at sixty days, 2.21 at one year and 1.53 at eighteen months. The cross-sectional regression of the one-year CAR on pre-event exposure gives +7.3 points per standard deviation ($t = 4.50$) and +5.8 ($t = 3.26$) with sector fixed effects. The size gradient is the new result: the one-year High-minus-Low difference is +19.5 points among the largest thousand firms ($t = 5.09$) and +9.9 among the rest ($t = 1.55$); +11.3 among S&P 500 members ($t = 1.92$) and +13.5 among non-members ($t = 3.15$). As in the index, nothing happens within sixty days of the launch and the difference accumulates over a year, which fits the gradual re-rating of large AI-exposed firms during the 2023–2024 capital-expenditure cycle better than a discrete attention shock; and the re-rating was concentrated in the firms that investors already followed most closely, the opposite of what an attention-friction account predicts.

Table 15. ChatGPT event study, Russell 3000. Mean market-model cumulative abnormal return (%) by pre-event AI tercile at four horizons, the High−Low difference, the Welch t (firms treated as independent) and the dependence-robust CDA t of equation (15); last row: one-year High−Low difference in sub-samples.

Horizon	High	Mid	Low	High−Low	Welch t	CDA t
60 trading days	+2.68	+1.36	+1.74	+0.94	0.65	0.30
6 months	-4.59	-8.54	-12.31	+7.72	3.28	1.73
1 year	-6.50	-16.27	-20.46	+13.96	3.94	2.21
18 months	-9.65	-20.63	-21.83	+12.18	2.29	1.53
Sub-samples (1-year H−L)	largest 1,000: +19.5 (5.09)	remaining: +9.9 (1.55)	S&P 500: +11.3 (1.92)			

Table 16. Event-study inference under cross-sectional dependence. High−Low portfolio abnormal return and t-statistics using the estimation-window and event-window standard deviation of the daily portfolio series; $\bar{\rho}$ is the average pairwise correlation of estimation-window residuals.

Horizon	Days	High−Low CAR (%)	t (estimation-window SD)	t (event-window SD)	$\bar{\rho}$
60 da years	60	+0.94	0.30	0.31	0.0371
6 months	125	+7.72	1.73	1.80	0.0371
1 year	250	+13.96	2.21	2.30	0.0371
18 months	395	+12.18	1.53	1.59	0.0371
High tercile: 741 firms; Low tercile: 743 firms; Kolari–Pynnonen deflation factor of the independence t: 0.131					

6.4. Difference-in-differences and the event-time profile (H4)

Table 17 reports the difference-in-differences estimates of equation (11) with treatment fixed at the firm's 2019–2021 exposure, within firm and year fixed effects. The continuous coefficient is +1.11 points per standard deviation ($t = 0.69$) and the binary one +2.54 ($t = 0.98$); neither is significant. The event-time profile of Figure 11 is flat around the launch: relative to 2021, the premium on pre-determined exposure is +0.0 points in 2022, +5.0 in 2024, and +7.6 in 2018, with several positive pre-launch years. Once exposure is held fixed, the broad universe shows no post-launch

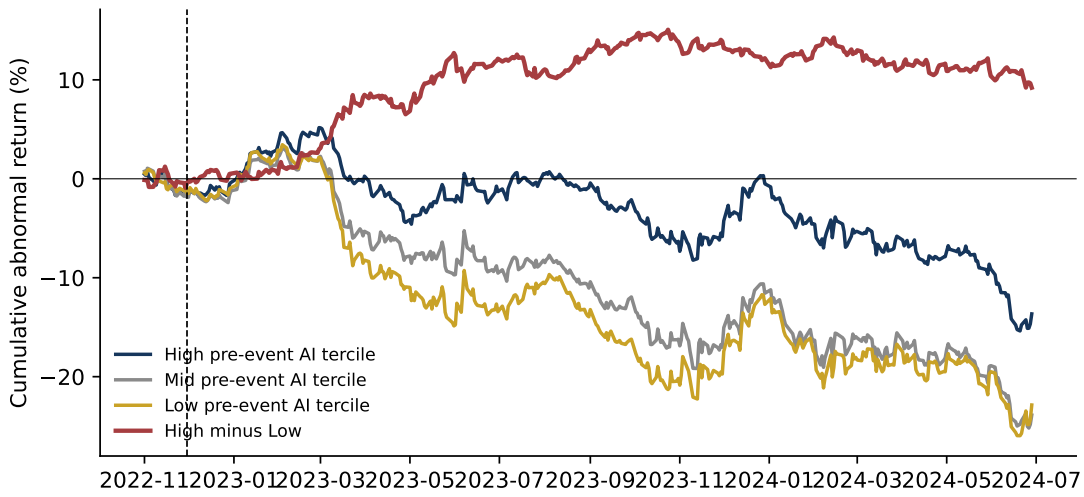


Figure 10. Cumulative abnormal returns around ChatGPT by pre-event AI tercile. All terciles drift negative because the market model was estimated in the 2022 bear market; only differences between terciles are interpretable.

break at all; firms with persistently AI-intensive filings earned premia scattered across the sample. The parallel-trends assumption is not satisfied, and the design confirms, rather than qualifies, the cohort evidence of Section 5.4.

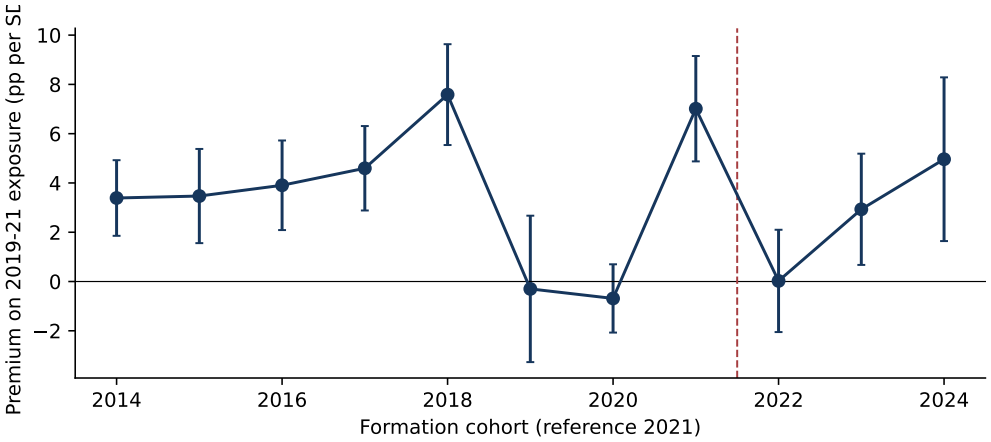


Figure 11. Event-time profile of the premium on pre-determined AI exposure, Russell 3000. Premium by formation year relative to 2021, 95% confidence bands from firm-clustered standard errors.

6.5. Disclosed projects versus disclosed language

The SPID project count offers an independent check on what the semantic score measures. If the AI language premium reflected concrete AI investment, firms that describe AI, data-center, cloud and automation projects should earn it too. They do not. The SPID Tech quintile spread is -1.9 points

Table 17. Difference-in-differences and the SPID project signal, Russell 3000. Top: firm and year fixed effects, double-clustered errors, treatment is pre-determined 2019–2021 exposure. Bottom: pooled regressions of the winsorized forward return on the SPID signal, the AI composite, their post-launch interactions and controls, with firm-clustered errors (percentage points per standard deviation or per unit of the indicator).

Specification	Coefficient (pp)	t
DiD: treated (above-median 2019–21 exposure) × post	+2.54	0.98
DiD: continuous exposure × post	+1.11	0.69
SPID technology projects (z)	+0.11	0.00
SPID technology projects (z) × post	-1.66	-0.00
AI composite (z)	+2.23	0.00
AI composite × post	+0.59	0.00
Any AI-initiative project (0/1)	-1.86	-0.00
Any AI-initiative project (0/1) × post	+5.08	0.00
AI composite (z)	+2.26	0.00
AI composite × post	+0.33	0.00

before the launch and -1.3 after (Table 9), its Fama–MacBeth slope is indistinguishable from zero in both universes (Table 7), and in the pooled regression of Table 17 neither the project count nor its post-launch interaction is significant while the composite retains its coefficient. Firms with at least one AI-initiative project earned -4.1 points less than other firms before the launch and +6.6 more after ($t = 1.68$), on a small and sparse sample. Two readings are possible. Disclosed projects may be a noisier measure than language because they are rare (only 2.0% of firm-years), or the market may price the *framing* of a firm as AI-exposed—which the embedding captures—rather than its announced AI capital commitments. Either way, the two measures extracted from the same text are not substitutes, and the premium documented here is a premium on AI language.

6.6. Composition of the signal by filing form

Because every firm-year now contains annual, quarterly and current reports, we can ask which text carries the signal (Table 18, Figure 12). The composite rebuilt from Form 10-K filings alone has a correlation of 0.61 with the baseline and earns +6.9 points before the launch and +16.7 after ($t = 2.61$), with a full-sample Fama–MacBeth slope of +6.06 points per standard deviation in a horse race against the 10-Q-only (-2.00) and 8-K-only (+0.90) composites. The 8-K-only composite, which is the closest analogue of the first version's signal, is the weakest: +1.1 before and +6.6 after. The annual narrative of the business, risk-factor and MD&A items is where the priced AI language lives; the earlier study, which could not read annual reports, measured the weakest component of the signal, which helps to explain why it found so little before 2022.

Table 18. Form-specific AI composites, Russell 3000. Composite rebuilt from the filing-level scores of one form at a time: coverage, correlation with the baseline, spreads, and Fama–MacBeth slope; last row, slopes when the three form-specific composites are entered jointly.

Signal built from	Firm-years covered	Corr. w/ baseline	Full Q5−Q1	Pre	Post	(t)	FM slope (pp/SD)
Form 10-K only	96.8%	0.61	+8.8	+6.9	+16.7	(2.61)	+4.66 (2.60)
Form 10-Q only	99.6%	0.61	+7.3	+5.4	+14.8	(1.91)	+3.30 (2.72)
Form 8-K only	99.9%	0.87	+2.2	+1.1	+6.6	(2.79)	+2.11 (2.96)
Horse race (FM, all three):							10-K: +6.06 (1.71); 10-Q: -2.00 (-0.78); 8-K: +0.90 (1.60)

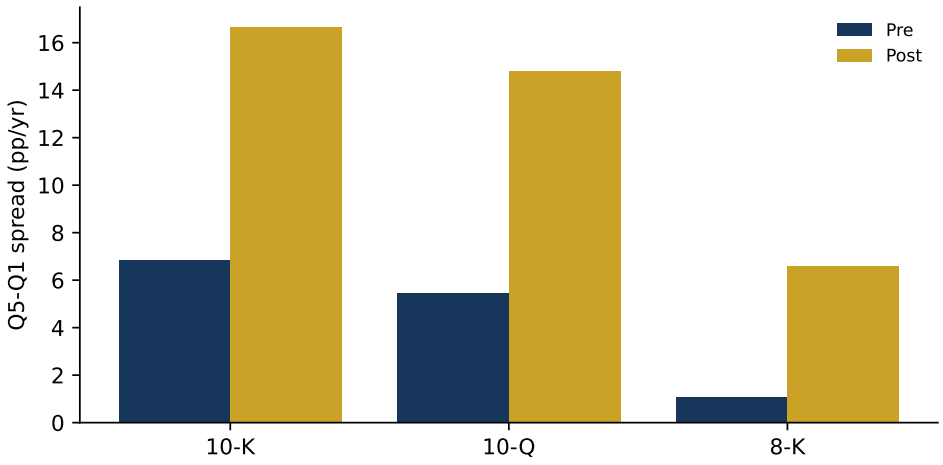


Figure 12. Pre- and post-launch spreads of the form-specific composites.

6.7. Construction, horizon, benchmark and universe robustness

Table 19 collects the remaining checks. Excluding the Magnificent 7 (+7.9 points after the launch) or the 500 largest firms (+8.1) leaves the spreads unchanged, as a broad-universe result should. The unweighted-mean and top-5% constructions of the signal, the first principal component (+11.6 post), the IWV-adjusted outcome (identical to the raw spread by construction of a common benchmark) and the 24-month horizon give the same picture; at a six-month horizon the spread is +2.7 points after the launch and nothing before. Among seasoned firms (listed before July 2009) the spread is +4.3 before and +12.6 after the launch, and among firm-years listed for at least 36 months +6.2 and +9.4; entrants, who earn +11.0 points more per year than seasoned firms, are spread evenly across AI quintiles. Finally, high-AI firms carry higher forward realized volatility (Q5−Q1 difference of +6.8 annualized points after the launch, $t = 3.51$, against +3.9 before), so part of the return differential is compensation for risk during a period of rapid technological repricing.

Table 19. Construction, horizon, benchmark and universe checks, Russell 3000. Equal-weighted Q5−Q1 composite spreads before and after the launch under alternative specifications; Newey–West t-statistics in parentheses.

Specification	Pre Q5−Q1	(t)	Post Q5−Q1	(t)
Baseline (EW quintiles)	+4.9	(2.00)	+7.9	(1.34)
Sector-neutral composite	-0.7	(-0.25)	+1.5	(0.72)
Excluding Magnificent-7	+4.6	(1.83)	+7.9	(1.30)
Excluding 500 largest	+4.9	(2.00)	+8.1	(1.28)
Signal: unweighted mean	+4.5	(1.84)	+7.3	(1.21)
Signal: top-5% chunks	+4.4	(1.74)	+5.6	(2.15)
Outcome: IWV-adjusted return	+4.9	(2.00)	+7.9	(1.34)
Horizon 6 months	-0.1	(-0.11)	+2.7	(0.94)
Horizon 24 months	+11.9	(2.41)	+2.4	(–)
PCA first component	+6.2	(2.19)	+11.6	(1.45)
Seasoned firms (listed before July 2009)	+4.3	(2.08)	+12.6	(1.96)
Listed $\geq$ 36 months	+6.2	(2.40)	+9.4	(2.14)

6.8. Definition of the break

Ending the pre-period in 2020 or starting the post-period in 2021 does not change the conclusion (results in `results/r3k/results2.json`, block `breaks`): in the broad universe the 2021 cohort earned a spread of +12.5 points, larger than two of the three post-launch cohorts, so moving it into the post-period raises the post mean and lowers the difference with the pre-period.

7. Discussion: Mechanism and Interpretation

What economic mechanism fits this constellation of facts? The first version of this paper, on the index alone, settled on a sector-level re-rating of persistent AI exposure after 2022. The broad universe keeps that element and adds two that the index could not reveal.

A persistent, partly within-industry premium in the broad market. In the Russell 3000 the relation between AI language and subsequent returns is not a post-2022 phenomenon. The composite spread is positive in three quarters of the pre-launch cohorts, the full-sample Fama–MacBeth slope is significant with and without sector dummies, the sector-by-year fixed-effects slope is significant over the full sample, and the premium is largest among the smaller firms that the index excludes. The 2018–2019 cohorts, held through the first wave of enterprise machine learning and cloud adoption, earned spreads as large as the post-launch ones. A reader of the S&P 500 study would have concluded that AI language was "dormant for a decade"; the broad market says that it was priced, modestly and intermittently, throughout the decade, among the firms whose AI exposure was least visible to the market at large. This is consistent with the disclosure-pricing literature's finding that qualitative filing content is processed incompletely and with delay (Cohen et al., 2020; Hoberg and Phillips, 2016), and with the attention literature's prediction that such processing is slowest for small, lightly covered firms (Hirshleifer et al., 2009; DellaVigna and Pollet, 2009).

A large-cap re-rating after 2022, not a market-wide regime change. The post-launch increase that the index sample displays is real within the largest thousand firms and absent outside them. Among large firms the spread triples after the launch, the one-year event-study difference is +19.5 points and highly significant, and the 2024 cohort—held through the memory, storage and data-center rally of 2025–2026—dominates. Among smaller firms the spread does not rise, the event-study difference is half as large and insignificant, and the long–short Sharpe ratio falls after the launch. The generative-AI boom was, to first order, a sectoral and large-cap reallocation of expected cash flows: the firms that were re-rated were the suppliers of the AI build-out and the largest users, which are disproportionately large, and the re-rating operated through sectors, as the sector-by-year fixed effects show. An attention-based account of the post-launch period would predict the opposite size gradient—the largest effect where attention frictions are largest—and is not supported.

Language, not projects. The sharpest new fact is that the premium attaches to the *language* of AI exposure and not to disclosed AI *projects*. Firms that announce AI initiatives, data centers or cloud migrations earn no premium on that account, before or after the launch, while firms whose

filings resemble the AI anchors do. One interpretation is that the embedding measures a firm's position in the AI value chain—its vocabulary of compute, infrastructure and disruption—which is what the market priced, whereas project announcements are a mixture of leaders and laggards catching up. Another is that project counts are too sparse to carry a signal. Both interpretations warn against reading the semantic premium as a premium on AI investment: it is a premium on being, or sounding like, an AI-exposed firm.

Where the signal lives. The annual report carries the strongest form of the signal and the current report the weakest, by a wide margin. The first version of this paper, which read current reports only, therefore understated the pre-launch premium and overstated the contrast with the post-launch period. Embedding-based measures inherit the composition of the corpus they are computed on; documenting that composition, and re-estimating on each form, is part of the method.

Relation to the AI-asset-pricing literature. The findings are complementary to Eisfeldt et al. (2023), who document a sharp repricing of an occupation-based generative-AI exposure measure in the weeks around the launch, and to Babina et al. (2024), who measure AI investment from labour data. Our disclosure-language measure shows no repricing within sixty days, a gradual one over a year among large firms, and a persistent premium among small firms that predates the launch by years. The three measures capture different objects—labour exposure, investment, and narrative positioning—and the differences in their timing and in their size gradients are themselves informative about how the market priced the AI theme.

7.1. What the evidence says about the hypotheses

H1 (unconditional pricing) is weakly supported in the broad universe: the full-sample spread and slope are positive and significant at conventional levels, but the risk-adjusted alpha is marginal and partly a technology tilt. H2 (regime dependence) is rejected for the market and supported only within the largest firms, where it rests on one cohort. H3 (AI specificity) holds against ESG, climate and macro language but not against cybersecurity and automation language. H4 (cross-sectional, not within-firm) is supported: the premium is absorbed by firm fixed effects and the pre-determined-exposure design shows no post-launch break. H5 (size and attention) is rejected in its attention form—the post-launch re-rating is larger among large firms—but the broad universe does reveal a persistent small-firm premium that the index concealed, and the project-based measure carries no premium.

7.2. Implications

For investors, the AI-language signal is a modest, persistent characteristic in the broad market, strongest among small firms and inside industries, and a large-cap, sector-driven bet after 2022; it is not a stable source of alpha, and the high post-launch Sharpe ratios of the index sample are a large-cap episode. For corporate managers, the asymmetry between AI and sustainability language illustrates how the prevailing narrative conditions the valuation consequences of what firms emphasize, while the absence of any within-firm effect and of any premium on announced projects means that neither emphasis nor announcement is rewarded as such. For researchers, the study shows that a conclusion drawn from an index sample—here, a dormant-then-priced regime story—can be an artefact of the index's size composition, and that measurement choices about which filings to read can change the sign of a comparison.

7.3. Limitations

First, the post-launch period contains three cohorts and the last ends in June 2026; the large-cap result rests mainly on the 2024 cohort. *Second*, the universe is a current holdings snapshot, which omits delisted firms and biases the level of small-firm returns upward; restrictions on listing history leave the results intact, but a point-in-time Russell membership has not been reconstructed. *Third*, AI language is correlated with sector, and after 2022 every sector control leaves nothing within sectors; the within-sector premium is a full-sample fact, not a post-launch one. *Fourth*, the semantic anchors embed the authors' definitional choices and no alternative anchor set has been tried. *Fifth*, the SPID project counts are extracted by a language model whose precision and recall have not yet been validated against human labels (UQO Working Paper No. 0001), so the null result on projects is a null result on a noisy measure. *Sixth*, the event-study horizons are long for a market-model design and the parallel-trends assumption is not met. *Finally*, the designs establish associations and their timing, not causality.

8. Conclusion

Using a semantic, embedding-based measure of artificial-intelligence language in the annual, quarterly and current SEC reports of some 2,500 U.S. listed firms over fifteen formation cohorts, linked to look-ahead-free forward returns, factor data and an independent project-based measure of AI engagement, we have re-examined whether and when AI language was priced in the cross-section of returns. The answer differs from the one we gave for the S&P 500 alone, and the difference is instructive. In the broad market, AI language carried a modest premium throughout 2010–2024: about +5.5 percentage points per year between the top and bottom quintiles, +3.07 points per

standard deviation in Fama–MacBeth regressions, with a within-sector component that survives sector-by-year fixed effects over the full sample and a concentration among the smaller firms that the index excludes. The sharp post-ChatGPT increase that the index displays—and that we replicate within the new panel—is a large-cap phenomenon: among the largest thousand firms the spread rises from +2.9 to +12.8 points and the event-study re-rating reaches +19.5 points at one year, whereas among the remaining firms the spread was already +9.6 points before the launch and did not rise, the 2022 cohort was negative, a placebo over ten candidate break years ranks 2022 8th, and the monthly long–short earned no post-launch alpha. Disclosed AI projects carry no premium at all; the language of annual reports carries the strongest.

Two messages emerge. The economic message is that the pricing of AI exposure in U.S. equities is better described as a persistent, partly within-industry characteristic premium concentrated in smaller firms, overlaid after 2022 by a sector-driven re-rating of the largest AI-exposed firms, than as a market-wide regime change triggered by a consumer event. The methodological message is that an index sample can manufacture a regime: the S&P 500 is the one part of the market where the premium was absent before 2022 and large afterward, and a study confined to it reads a large-cap episode as a change in how the market prices disclosure. Reading the full universe, and reading all three forms of filing, reverses that reading.

Several next steps follow. The most important is to add post-2024 cohorts as they accrue, which is the only way to learn whether the large-cap re-rating of 2024 was the start of a regime or an episode. The second is a point-in-time Russell membership, which would remove the survivorship that is more severe among small firms. The third is to validate the SPID project extraction against human labels so that the null result on disclosed projects can be attributed to the market rather than to measurement. Beyond these, earnings-call transcripts and patent text would sharpen the distinction between narrative positioning and real AI activity that the contrast between language and projects has opened.

References

Acemoglu, Daron and Pascual Restrepo, "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

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

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 John D. Lyon, "Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics," *Journal of Financial Economics*, 1997, *43* (3), 341–372.

_ 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.

Boehmer, Ekkehart, Jim Musumeci, and Annette B. Poulsen, "Event-Study Methodology under Conditions of Event-Induced Variance," *Journal of Financial Economics*, 1991, *30* (2), 253–272.

Bolton, Patrick and Marcin Kacperczyk, "Do Investors Care about Carbon Risk?," *Journal of Financial Economics*, 2021, *142* (2), 517–549.

Brown, Stephen J. and Jerold B. Warner, "Using Daily Stock Returns: The Case of Event Studies," *Journal of Financial Economics*, 1985, *14* (1), 3–31.

_ , William N. Goetzmann, and Stephen A. Ross, "Survival," *Journal of Finance*, 1995, *50* (3), 853–873.

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.

Carhart, Mark M., "On Persistence in Mutual Fund Performance," *Journal of Finance*, 1997, *52* (1), 57–82.

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

Cong, Lin William, Tengyuan Liang, Xiao Zhang, and Wu Zhu, "Textual Factors: A Scalable, Interpretable, and Data-Driven Approach to Analyzing Unstructured Information," NBER Working Paper 33168, National Bureau of Economic Research 2024.

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., "Market Efficiency, Long-Term Returns, and Behavioral Finance," *Journal of Financial Economics*, 1998, *49* (3), 283–306.

_ **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.

Hanley, Kathleen Weiss and Gerard Hoberg, "Dynamic Interpretation of Emerging Risks in the Financial Sector," *Review of Financial Studies*, 2019, *32* (12), 4543–4603.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, "... and the Cross-Section of Expected Returns," *Review of Financial Studies*, 2016, *29* (1), 5–68.

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.

Hou, Kewei, Chen Xue, and Lu Zhang, "Replicating Anomalies," *Review of Financial Studies*, 2020, *33* (5), 2019–2133.

Huang, Allen H., Hui Wang, and Yi Yang, "FinBERT: A Large Language Model for Extracting Information from Financial Text," *Contemporary Accounting Research*, 2023, *40* (2), 806–841.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, "ChatGPT and Corporate Policies," NBER Working Paper 32161, National Bureau of Economic Research 2024.

Ke, Zheng Tracy, Bryan T. Kelly, and Dacheng Xiu, "Predicting Returns with Text Data," NBER Working Paper 26186, National Bureau of Economic Research 2019.

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.

Kolari, James W. and Seppo Pynnönen, "Event Study Testing with Cross-Sectional Correlation of Abnormal Returns," *Review of Financial Studies*, 2010, *23* (11), 3996–4025.

Kothari, S. P. and Jerold B. Warner, "Econometrics of Event Studies," in B. Espen Eckbo, ed., *Handbook of Empirical Corporate Finance*, Vol. 1, Amsterdam: Elsevier, 2007, pp. 3–36.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, "Measuring Corporate Culture Using Machine Learning," *Review of Financial Studies*, 2021, *34* (7), 3265–3315.

Lopez-Lira, Alejandro, "Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns," Working Paper, University of Florida 2023. SSRN 3313663.

_ **and Yuehua Tang**, "Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models," *Journal of Financial Economics*, 2026, *184*, 104335.

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.

_ **and** _ , "Textual Analysis in Finance," *Annual Review of Financial Economics*, 2020, *12*, 357–375.

MacKinlay, A. Craig, "Event Studies in Economics and Finance," *Journal of Economic Literature*, 1997, *35* (1), 13–39. JSTOR 2729691.

McLean, R. David and Jeffrey Pontiff, "Does Academic Research Destroy Stock Return Predictability?," *Journal of Finance*, 2016, *71* (1), 5–32.

Mitchell, Mark L. and Erik Stafford, "Managerial Decisions and Long-Term Stock Price Performance," *Journal of Business*, 2000, *73* (3), 287–329.

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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)" 2019, pp. 3980–3990.

Sautner, Zacharias, Laurence van Lent, Grigory Vilkov, and Ruishen Zhang, "Firm-Level Climate Change Exposure," *Journal of Finance*, 2023, *78* (3), 1449–1498.

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.

Thompson, Samuel B., "Simple Formulas for Standard Errors That Cluster by Both Firm and Time," *Journal of Financial Economics*, 2011, *99* (1), 1–10.

Webb, Michael, "The Impact of Artificial Intelligence on the Labor Market," Working Paper, Stanford University 2020. SSRN 3482150.

Appendices

A. Variable Definitions

B. Anchor Texts

Each concept is defined by the three paragraphs below (`src/factors/anchors.py` of the EDGAR Research Database). The three paragraphs are embedded separately with `all-MiniLM-L6-v2`, averaged, and normalized to a unit vector; chunk scores are cosine similarities to that vector. The anchors were written by the authors in June 2026 and have not been revised since; no alternative anchor set has been tested.

AI adoption (operational weighting). (1) "The company internally uses artificial intelligence, machine learning, generative AI, large language models, automation, predictive analytics, intelligent agents, and data-driven systems to improve productivity, operations, efficiency, decision-making, customer service, risk management, compliance, cybersecurity, supply chain, software development, finance, human resources, and other business processes." (2) "We have implemented generative AI, large language models, and natural language processing capabilities across our enterprise. Our AI strategy includes deploying intelligent automation, computer vision, and predictive analytics systems to transform operations." (3) "Investment in artificial intelligence research and development, including partnerships with AI technology providers, hiring of machine learning engineers, and deployment of AI infrastructure including GPU computing clusters for internal business applications."

AI disruption risk (risk weighting). (1) "Artificial intelligence may reduce the scarcity, defensibility, differentiation, pricing power, or competitive advantage of the company's products or services by enabling rapid replication, automation, synthetic generation, low-cost substitution, or commoditization of capabilities that were previously difficult, expensive, or specialized." (2) "The rapid advancement of AI technologies may render our current offerings less competitive. New market entrants leveraging AI could fundamentally alter industry dynamics, reduce barriers to entry, and erode our market position through AI-native products and services." (3) "Risks associated with artificial intelligence include potential competitive displacement by AI-powered alternatives, obsolescence of our existing products, regulatory uncertainty around AI governance, algorithmic bias, intellectual property challenges, and emergence of AI-native competitors."

Table A1. Variable Definitions

Variable	Definition
AI Adopt (`ai_adoption_wmean`)	Average over the scored filings of firm i dated in calendar year Y of the section-weighted mean cosine similarity of the filing's text chunks to the AI-internal-adoption anchor, equations (1)–(3).
AI Infra (`ai_infrastructure_wmean`)	Same construction for the AI-compute-infrastructure anchor.
AI Risk (`ai_disruption_risk_wmean`)	Same construction for the AI-product-disruption-risk anchor.
Simple mean, top-5% tail	Filing-level unweighted mean of chunk scores, and mean of the chunks at or above the filing's 95th percentile, averaged over filings as above.
AI Compos	$\frac{1}{2}[z(\text{AI Adopt}) + z(\text{AI Infra})]$, with z the within-formation-year cross-sectional standardization of equation (5).
Form-specific composites	AI Compos rebuilt from the filing-level scores of Form 10-K, Form 10-Q or Form 8-K filings only (`data/raw_russell3000/filing_scores`).
SPID technology projects	$\log(1+n)$, where n is the number of project mentions of types AI initiative, data center, cloud migration, automation, digital transformation, ERP implementation or technology deployment extracted by the SPID database from the firm's filings of year Y; standardized within year. Any-AI-project is an indicator for at least one AI-initiative mention or AI-tagged technology in year Y.
Pre-determined exposure	Firm average of AI Compos over filing years 2019–2021, standardized across firms; binary treatment equals one above the median.
Forward return	Buy-and-hold simple return from July of $Y+1$ to June of $Y+2$, equation (4), from adjusted closes (dividends reinvested), requiring at least six monthly returns.
Market-adjusted return	Forward return minus the contemporaneous SPY return (IWV-adjusted version in robustness).
Size	Natural logarithm of market capitalization at formation: price times the shares outstanding most recently reported on the cover of a 10-K or 10-Q filed on or before the formation date (SEC XBRL `dei:EntityCommonStockSharesOutstanding`).
Size group	Largest 1,000 firms by market capitalization at formation (Russell 1000 proxy) versus the remaining firms (Russell 2000 proxy), re-assigned each year.
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 (filing) years ≥ 2022.
TECH	Equal-weighted monthly return of the Information Technology constituents of the panel minus the equal-weighted return of all constituents.
CAR	Cumulative market-model abnormal return, equations (13)–(14); market model estimated 3 January–31 October 2022 on the iShares Russell 3000 ETF (IWV), event window 1 November 2022–30 June 2024.

AI infrastructure (operational weighting). (1) "Investment in AI computing infrastructure including data centers, GPU and TPU clusters, cloud computing resources, semiconductor chips, accelerator hardware, and high-performance computing systems dedicated to artificial intelligence and machine learning workloads." (2) "Capital expenditures for AI infrastructure, including networking equipment for distributed AI training, specialized computing platforms for large-scale model development, edge computing for AI inference, and data pipeline infrastructure for machine learning operations." (3) "Cloud infrastructure spending for AI workloads, including partnerships with hyperscale cloud providers, deployment of AI-optimized hardware, GPU supply chain management, and investment in next-generation computing architectures for artificial intelligence applications."

Placebo concepts. The ESG-commitment anchor describes sustainability reporting, corporate social responsibility, diversity and inclusion, stakeholder engagement, adherence to SASB, GRI, and TCFD frameworks, emission reductions, and board-level oversight of environmental and social risk. The climate-risk anchor describes physical and transition risks, carbon pricing, stranded assets, water scarcity, TCFD scenario analysis, and Scope 1–3 emissions targets. The cybersecurity anchor describes data breaches, ransomware, GDPR and CCPA compliance, security operations, zero-trust architecture, cyber insurance, and "AI-powered threat detection." The automation/labour anchor describes robotic process automation, workforce restructuring, reskilling, human–machine collaboration, and "the future of work transformation driven by AI and robotics." The macro-risk anchor describes inflation, interest and exchange rates, recession, geopolitical tensions, tariffs, commodity prices, and monetary and fiscal policy. The full texts are in the source file cited above. Two of the placebo anchors (cybersecurity and automation/labour) mention AI explicitly, which bears on the interpretation of Table 14.

C. Corpus Composition

The EDGAR Research Database was extended on 27–28 October 2026 from the 500 S&P 500 constituents to an issuer universe that merges the S&P 500 snapshot of May 2026 with the holdings of the iShares Russell 3000 ETF as of 23 October 2026 (2,564 distinct SEC registrants). For every issuer it downloaded the primary document of every Form 10-K, 10-Q and 8-K filed since January 2010 (527548 filings), parsed all of them into sections and 512-word chunks (1342283 sections, 10720890 chunks), embedded every chunk and scored it against the eight anchors. Unlike the June 2026 release, in which only the 8-K filings and the 10-Q filings of 84 firms had been scored, the present panel is built from all three forms for every firm; Table 18 in the text reports the composition of the firm-year signal by form. The year variable of the panel is the calendar year of the filing date. The construction of the universe, its reconciliation and the extraction of the SPID project signals are

documented in the companion methodological paper (UQO Working Paper No. 0001).

D. Data Sources and Reproducibility

Semantic scores and project mentions are from the EDGAR Research Database and the SPID project database (UQO), which are not redistributed; the firm-year factor panel, the filing-level scores by form, the SPID firm-year counts, the monthly and daily prices, the share-count history, the factor series and the long–short series needed to reproduce every table and figure are kept in `data/raw_russell3000/` and `data/processed_r3k/`. Equity prices are from Yahoo Finance (downloaded 28 October 2026); shares outstanding are from the SEC XBRL company-facts API; the Fama–French five factors, the momentum factor and the risk-free rate are from the Kenneth R. French data library (through September 2026). The pipeline in `scripts/r3k_*.py` runs in numerical order: `00` downloads the factors, `01` the prices and `01c` the share counts, `02` builds the look-ahead-free panel, `03` runs the sorts, factor regressions, panel and Fama–MacBeth estimators, the break interaction and the size-group analyses, `04` the placebo, sector-neutral and double sorts, exclusions, alternative signals, horizons and benchmarks, principal components, difference-in-differences and event-time profile, cohort and pooled tests, entrants, sector controls, technology-factor spanning, filing-type composition, placebo break dates and the SPID interactions, `05` the event study, `06` the figures and `07` the LaTeX tables and the number macros used in the text. All numerical results are serialized to `results/r3k/*.json`; every number quoted in the text is a macro generated by `07` or a table cell, so that the manuscript can be regenerated from the data without manual editing. The June 2026 and October 2026 (v1.1) versions of the paper, which covered the S&P 500 only, are archived in `docs/paper_v1.1_sp500_20261026/`.