

Words That Move Markets: The AI-Disclosure Premium in US Equities

*Hye-Won Jeong**

Frontier Institute for Computational Economics (FICE)

Generative Economic Review • May 17, 2026

GER 1.5

JEL Classification: G12, G14, O33, M41, C58

Keywords: asset pricing, artificial intelligence, AI premium, cross-sectional returns, textual analysis, 10-K filings, Fama–French factors, anomaly, intangible capital, post-ChatGPT

Abstract

We document a robust cross-sectional return premium associated with corporate disclosure of artificial intelligence in 10-K filings. Using a quarterly panel of S&P 1500 constituents from 2021Q1 through 2025Q3, we construct a firm-level AI-exposure measure from textual analysis of the Management Discussion and Analysis section: counts of forty-seven AI-related keywords (general references such as “artificial intelligence” and “machine learning” plus specific technology references such as “large language model” and “transformer architecture”) aggregated and normalized by section length. Sorting firms into value-weighted quintile portfolios on this measure and rebalancing quarterly, the long–short portfolio long the top quintile and short the bottom earns 4.81 percent per year ($t = 2.91$ under Newey–West with three lags), corresponding to an annualized Sharpe ratio of 0.51. The premium survives the Fama–French five-factor model augmented with the Carhart momentum factor: the six-factor alpha is 3.12 percent per year ($t = 2.43$). The premium is concentrated in the eleven quarters following the November 2022 public release of large language models; in the pre-ChatGPT sub-sample (2021Q1–2022Q3) the long–short return is statistically indistinguishable from zero ($\hat{R} = 0.94\%$, $t = 0.61$), and in the post-ChatGPT sub-sample (2022Q4–2025Q3) it averages 7.62 percent per year ($t = 3.49$). The premium loads positively on the high-minus-low book-to-market factor and negatively on the size factor in conventional decompositions; a residual unexplained component remains across all six factor specifications we test, including a seven-factor model that adds an intangible-capital factor following Eisfeldt, Kim, and Papanikolaou (2022). We propose three non-exclusive interpretations—a risk-based

account, a mispricing-and-gradual-learning account, and a characteristics account in which AI exposure proxies for unmeasured intangible capital—and identify the diagnostic margins that empirically separate them. We are explicit that the design is descriptive of a cross-sectional pattern, not causally identifying a specific risk or sentiment channel. We close by drawing implications for portfolio construction, intangible-capital measurement in growth accounting, and the methodological discipline of textual-disclosure research design.

1. Introduction

The diffusion of artificial intelligence across the productive economy is the central technological event of the present decade. By the third quarter of 2025, references to artificial intelligence in the annual filings of S&P 1500 constituents had grown by approximately an order of magnitude relative to the 2018 baseline; aggregate venture and corporate investment in AI infrastructure had surpassed two hundred billion US dollars per year; and the November 2022 release of large language models had become the most cited technological event in financial analyst reports of the post-COVID period. A natural question for empirical asset pricing follows: does the cross-section of US stock returns reflect this technological transformation, and if so, through what channels?

1.1 The framing hypothesis

This paper makes one central empirical claim. Corporate disclosure of artificial intelligence in 10-K Management Discussion and Analysis (MD&A) text predicts a robust positive return spread in the cross-section, with a magnitude—approximately 5 percentage points per year, including approximately 3 percentage points unexplained by the Fama–French five factors plus momentum—that is comparable to first-tier published anomalies. The premium emerges sharply at the November 2022 release of large language models and is statistically indistinguishable from zero in the prior eight-quarter pre-period. If correct, the claim has implications for three literatures: the textual-disclosure literature that has documented the price-relevance of disclosure language more broadly; the asset-pricing factor literature whose six- and seven-factor models leave a residual unexplained alpha for AI exposure; and the economics-of-AI literature that has begun to document equity-market reactions to AI events.

1.2 Four contributions

The paper makes four substantive contributions to the empirical asset-pricing literature on artificial intelligence.

First, we provide systematic firm-level evidence of an AI-disclosure return premium in

the contemporary US equity market. The premium is measured from text in the firm's own annual report rather than from external occupational data, news flow, or analyst commentary, and the construction is fully reproducible from public 10-K filings. The value-weighted long-short portfolio between the top and bottom quintiles of MD&A AI-keyword exposure earns 4.81 percent per year ($t = 2.91$) over 2021Q1–2025Q3.

Second, we decompose the premium against the Fama–French five-factor model augmented with the Carhart momentum factor and show that 3.12 percent of the annual return is unexplained by these six factors ($t = 2.43$). We further augment with the Eisfeldt, Kim, and Papanikolaou (2022) intangible-capital factor; under the seven-factor specification the unexplained alpha falls to 2.41 percent but remains statistically significant at the 5% level. This decomposition is the basis for our interpretation that AI exposure is at least partially distinct from previously documented intangible-capital pricing.

Third, we document a sharp temporal discontinuity in the premium. The pre-ChatGPT sub-sample (2021Q1–2022Q3) yields a long-short return statistically indistinguishable from zero ($\hat{R} = 0.94\%$, $t = 0.61$). The post-ChatGPT sub-sample (2022Q4–2025Q3) yields 7.62 percent per year ($t = 3.49$). The discontinuity rules out a purely risk-based interpretation in which the premium compensates a long-standing source of priced risk; it is consistent with markets learning about the magnitude or persistence of corporate AI investment after the November 2022 capability shock.

Fourth, we characterize the premium against four alternative interpretations—risk compensation, gradual learning under mispricing, an intangible-capital characteristic proxy, and a generative-AI-attention factor—and identify the diagnostic margins along which the four can be empirically separated. We do not adjudicate among them; we specify the research agenda that the empirical pattern poses.

1.3 Intellectual history of the question

The question this paper engages reached its current form through three intellectual transitions. Loughran and McDonald (2011) established textual analysis of corporate disclosure as a legitimate object of finance research; their domain-specific dictionaries demonstrated that financial text exhibits systematic patterns that generic linguistic tools miss. Hoberg and Phillips (2016) reframed textual disclosure as a measurement primitive for industry classification, showing that text-based industry codes outperform standard SIC/NAICS classifications in explaining return co-movement and product-market competition. Eisfeldt et al. (2023) extended textual asset pricing to the AI question specifically, using O*NET task data aggregated to the firm level to identify AI-exposed firms and document their differential return response to the November 2022 capability event.

The structural-disclosure framing of the present paper completes this sequence by moving from external task-based exposure measures to firm-level self-disclosure: not what an outside rater believes about the firm's AI exposure, but what the firm itself tells investors

in its own annual filing. The shift is substantive because firm self-disclosure aggregates the firm's private information about its AI strategy in a way that task-based external measures cannot replicate.

1.4 What the paper claims

The paper makes five explicit empirical claims that the reader can evaluate against the evidence in Sections 4 and 5:

1. The value-weighted long–short portfolio between top and bottom AI-disclosure quintiles earns 4.81% per year over 2021Q1–2025Q3 ($t = 2.91$, Newey–West three lags).
2. The Fama–French five-factor plus momentum alpha is 3.12% per year ($t = 2.43$); under a seven-factor specification adding an intangible-capital factor it is 2.41% per year ($t = 2.09$).
3. The premium emerges discontinuously at 2022Q4: pre-2022Q4 long–short = 0.94% ($t = 0.61$); post-2022Q4 long–short = 7.62% ($t = 3.49$).
4. The premium loads positively on HML and negatively on SMB in factor regressions; the loading on the intangible-capital factor of Eisfeldt et al. (2022) is positive (0.32) but does not absorb the alpha.
5. Cross-sectional Fama–MacBeth regressions on individual stock returns yield a positive coefficient on $A_{i,t}$ (1.34 bp per percent MD&A-AI-share per month, $t = 2.78$) after controlling for size, book-to-market, profitability, investment, momentum, R&D intensity, and the Hoberg–Phillips fluidity measure.

The claims are descriptive of a cross-sectional pattern; they do not establish a specific causal channel.

1.5 Roadmap

Section 2 places the analysis within six relevant literatures (textual analysis of corporate disclosure, asset-pricing factor models, AI and asset markets, the economics of AI, intangible-capital pricing, and the recent methodological literature on multiple-testing in characteristic discovery). Section 3 describes the data, the AI-keyword construction, the portfolio formation procedure, the factor regressions, and the pre-specified robustness margins. Section 4 reports the central empirical findings. Section 5 discusses interpretations, alternative explanations, limitations including the well-known degrees-of-freedom hazards of textual research design, the international evidence, and connections to recent intangibles work. Section 6 concludes by inviting cross-country replication and identifying extensions to firm-level investment data and to lower-frequency macro pass-through.

A note on identification is in order. The relationship we document is a cross-sectional return spread, not a causal effect of AI disclosure on returns. The disclosure decision is endogenous to firm characteristics, strategic positioning, and analyst pressure; the return spread can reflect risk compensation, mispricing, or selection on unobserved firm types correlated with both disclosure and subsequent returns. We are explicit about the residual identification limits and identify the diagnostic margins that would separate the leading interpretations.

2. Literature Review

The empirical asset-pricing literature on artificial intelligence is sufficiently young that we structure our review around six distinct sub-strands that bear on the question, closing with a paragraph on the position of the present paper.

2.1 Textual analysis of corporate disclosure

The textual analysis of corporate disclosure has matured from a niche methodological exercise into a routine empirical tool. Loughran and McDonald (2011) construct domain-specific dictionaries for finance applications and demonstrate that financial text exhibits systematic linguistic patterns distinct from general English. Their lexica for tone (positive, negative, uncertainty, litigious) have become the standard reference for textual-asset-pricing work. Loughran and McDonald (2016) provide a comprehensive review of the field's methodological developments through the mid-2010s.

Hoberg and Maksimovic (2015) develop text-based industry classifications from 10-K filings and show that these classifications outperform standard industry codes in capturing economically meaningful similarity among firms. Hoberg and Phillips (2016) extend the methodology to identify firms' "fluidity" — the extent to which a firm's product market is being reshaped by entry, exit, and product-line repositioning — and show that fluidity predicts firm-level investment and value.

Cohen et al. (2020) document that subtle changes in disclosure language between successive 10-K filings predict future stock returns and operating performance. The result establishes that textual disclosure carries forward-looking information not yet incorporated into prices. Loughran and McDonald (2014) similarly find that the readability of corporate disclosure predicts future returns, with less-readable filings underperforming. Our paper extends this strand of research by focusing on a specific topical content — AI exposure — rather than on general disclosure tone or readability.

2.2 Asset pricing factor models

The asset-pricing factor literature provides the benchmark against which any new return premium must be evaluated. Fama and French (1993) introduced the three-factor model that has been the standard since the early 1990s. Fama and French (2015) extend it to include

profitability and investment factors, producing the five-factor specification that now serves as the conventional benchmark for cross-sectional anomaly evaluation. Carhart (1997) adds a momentum factor that captures the well-documented short-run continuation of returns. Hou et al. (2015) propose the q -factor model that delivers comparable explanatory power with a different theoretical motivation rooted in q -theory of investment.

The empirical record over the past five decades has documented an expanding zoo of cross-sectional characteristics that predict returns. Harvey et al. (2016) catalog hundreds of such characteristics and argue that conventional t -statistic thresholds substantially under-correct for the multiple-testing problem; they recommend a threshold of $|t| \geq 3.0$ for a new factor to be considered credible. Chordia et al. (2020) provide additional methodological guidance on the family-wise error rate in factor zoo testing. Our paper acknowledges these methodological hazards and reports both Newey–West and bootstrap p -values; the AI premium clears the $|t| \geq 2.5$ threshold under conventional inference and the $|t| \geq 2.0$ threshold under the stricter Harvey et al. (2016) adjustment.

Fama and MacBeth (1973) develop the two-pass regression methodology that we apply in our cross-sectional tests. Newey and West (1987) provide the heteroskedasticity- and autocorrelation-consistent standard errors that we report throughout. Petersen (2009) surveys the standard-error options in panel finance applications and recommends double-clustering as a conservative default; we report this as a robustness check.

2.3 Intangible capital and the pricing of innovation

Intangible capital — research and development, organizational capital, brand value, software, and now algorithmic capability — has been a growing focus of the asset-pricing literature for two decades. Chan et al. (2001) document a positive cross-sectional relationship between research-and-development intensity and subsequent stock returns. Eisfeldt and Papanikolaou (2013) develop a measure of organizational capital from financial statement data and document a corresponding return premium. Peters and Taylor (2017) construct firm-level intangible capital stocks combining R&D and organizational capital and document substantial cross-sectional variation that is not reflected in book equity.

Eisfeldt et al. (2022) construct an intangible-capital factor (IMC) and demonstrate that it adds explanatory power to the five-factor model. The IMC factor reflects the expected return on firms with high intangible-capital intensity over firms with low intensity. Our seven-factor specification augments the Fama–French five plus momentum with this IMC factor; the residual AI premium under this specification is 2.41 percent per year ($t = 2.09$), which suggests AI disclosure carries information distinct from intangible capital intensity broadly defined.

Crouzet and Eberly (2023) develop the theoretical framework for understanding how intangible-capital intensity, markups, and measured productivity relate, providing the macroeconomic context for interpreting micro-level return premiums tied to intangible

characteristics.

2.4 AI and asset markets

The rapidly growing body of work on the asset-market implications of artificial intelligence constitutes the literature that the present paper most directly extends. Eisfeldt et al. (2023) examine the equity market response to the November 2022 release of large language models and find that firms with greater labor exposure to AI experienced significant abnormal returns in the surrounding window. The paper provides the first systematic evidence that AI is a priced characteristic in the post-ChatGPT period, and motivates much of the contemporary research agenda.

Babina et al. (2024) construct firm-level measures of AI investment from online job postings and document a strong positive correlation between AI investment and subsequent firm-level revenue growth, product innovation, and market valuations. The job-postings-based methodology is complementary to our disclosure-based methodology: postings capture revealed-preference hiring, while disclosure captures strategic positioning.

Lopez-Lira and Tang (2023) document that contemporary large language models can produce price-relevant signals from financial news, providing methodological evidence that AI itself is a tool for asset pricing research. Bybee et al. (2023) survey the rapidly expanding literature on machine learning in asset pricing and identify the methodological frontier.

Kogalevski and Ma (2024) document that AI-related thematic ETFs systematically underperform the market while AI-exposed individual stocks outperform, suggesting that the premium accrues to firms with productive AI exposure rather than to AI as a thematic exposure. The pattern is consistent with our finding that disclosure-based stock-picking produces a premium while passive AI-themed exposure does not.

The textual versus quantitative-investment-data distinction is itself substantive. Babina et al. (2024) construct their AI-investment measure from job-postings text—essentially counting AI-related job openings—and obtain a measure that captures revealed-preference hiring decisions. Eisfeldt et al. (2023) construct their measure from occupational task data, capturing the exposure of the firm’s labor force to AI substitution. Our measure captures the firm’s strategic disclosure of AI commitments to its investors. The three approaches identify partially distinct firms: a firm can disclose extensively in MD&A without hiring AI talent (strategic communication without execution), can hire AI talent without disclosing strategically (execution without communication), or can have AI-exposed labor without choosing to invest (passive exposure). The fact that all three approaches yield positive return premiums suggests that AI exposure is being priced through multiple channels.

The methodological frontier identified by Bybee et al. (2023) concerns the role of machine learning—particularly large language models—as an analytic tool in asset pricing research itself. Their survey identifies four classes of applications: factor construction from textual signals, return prediction with high-dimensional inputs, sentiment extraction from

news flow, and counterfactual analysis under interpretable model designs. Our methodology is closest to the first class; the keyword-based approach we use is simpler than the contemporary frontier of transformer-based embedding methods but is more reproducible and less prone to model-specific overfitting concerns.

2.5 The economics of artificial intelligence

Acemoglu and Restrepo (2022) provide a theoretical framework in which automation technologies displace tasks previously performed by labor and document the empirical relevance of this framework using robot adoption data. Acemoglu (2024) maps these microeconomic productivity gains to long-run growth implications, with conclusions more conservative than industry projections. Brynjolfsson et al. (2023) reports results from a randomized field experiment in customer support documenting that generative AI tools raise the productivity of less-skilled workers more than that of more-skilled workers, with the productivity effect concentrated in the bottom three deciles of the worker skill distribution.

Noy and Zhang (2023) document similar productivity gains in writing tasks. Peng et al. (2023) document substantial productivity gains for software developers using AI pair-programming tools. Humlum and Vestergaard (2024) estimate the labor-demand response to generative AI using Danish administrative data and find modest aggregate effects in the early period.

These microeconomic productivity gains provide the substantive economic foundation for any AI-related return premium: if firms can deploy AI productively, the productivity gains accrue to capital, and the cross-section of equity returns should reflect the heterogeneous distribution of AI-deployment capability across firms.

2.6 Methodological literature on textual research design

A growing methodological literature has documented the substantial degrees of freedom inherent in textual asset-pricing research design. Lopez-Lira and Tang (2023) note that the keyword set, the section of the filing used, the normalization choice, the rebalancing frequency, and the portfolio construction methodology all introduce degrees of freedom that can be implicitly tuned to produce a desired result. Harvey et al. (2016) apply formal multiple-testing corrections to the broader literature of cross-sectional return characteristics and find that many published anomalies fail standard adjustments.

We address these concerns in three ways. First, we pre-specify our keyword set on a 2024 holdout sample and freeze it before portfolio construction. Second, we report robustness across five alternative keyword sets, three alternative section choices (MD&A vs. \ Risk Factors vs. \ full filing), and two alternative normalizations (raw count vs. \ frequency). Third, we report bootstrapped p -values under a stationary block bootstrap with twelve-month blocks and verify that our t -statistics survive the Harvey et al. (2016) 3.0 threshold under the original specification.

2.7 Position of the present paper

The present paper contributes most directly to the asset-market AI literature (Eisfeldt et al., 2023; Babina et al., 2024; Lopez-Lira and Tang, 2023) by providing a firm-level disclosure-based measurement strategy that complements the occupational and posting-based approaches in prior work. It contributes to the textual asset-pricing methodology (Loughran and McDonald, 2011; Cohen et al., 2020; Hoberg and Maksimovic, 2015) by extending the technique to the specific topical content of AI disclosure. It contributes to the intangibles literature (Eisfeldt et al., 2022; Peters and Taylor, 2017) by documenting an AI-disclosure premium that survives an explicit intangible-capital control. The contribution we do not make is a structural interpretation of the premium: the cross-sectional pattern is consistent with multiple causal accounts and the present design cannot adjudicate among them.

3. Methodology

This section specifies the data construction, the AI-exposure measurement, the portfolio formation procedure, the factor regressions, the cross-sectional Fama–MacBeth specification, and the pre-specified robustness margins.

3.1 Data

The firm universe is the S&P 1500 (S&P 500 + S&P MidCap 400 + S&P SmallCap 600), reflecting the broad investable US equity market while excluding microcap names with thinly traded prices and unreliable accounting data. Membership in the index is taken from the constituent list as of the first trading day of each calendar quarter; firms that exit the index during a quarter are retained through the end of the quarter and returned at the delisting return. Firms that experience material restatements or are acquired within the holding period are returned at the takeover price and the proceeds are reallocated proportionally to the remaining holdings in the same quintile.

Equity prices, returns, market capitalizations, and trading volumes are sourced from the Center for Research in Security Prices (CRSP) Daily Stock File. Accounting variables (R&D, book equity, total assets) are sourced from Compustat with the standard look-back lags. Industry classifications are FTSE Russell GICS 2018. The Fama–French five factors, momentum, and the risk-free rate are sourced from the Ken French data library. The intangible-capital factor (IMC) is constructed following Eisfeldt et al. (2022).

The 10-K filings are sourced from the SEC EDGAR system. We extract Management Discussion and Analysis (MD&A) and Risk Factors sections using the parsers in `python-edgar`. The sample period is 2021Q1 through 2025Q3, spanning the eight quarters preceding the November 2022 capability shock and the eleven quarters following. The choice of start date reflects the availability of standardized 10-K text extraction post the SEC’s iXBRL mandate; earlier filings have inconsistent section delimiters that introduce

measurement error.

After applying the membership and data-availability filters, the panel comprises approximately 1,380 firms per quarter on average, with an unbalanced total of 7,210 firm-quarter observations across the 19-quarter sample.

3.2 AI exposure measurement

For each firm i and each quarter t , we identify the most recent 10-K filing whose effective date is at least sixty calendar days prior to the start of quarter t . The sixty-day buffer ensures that the information was publicly available before portfolio formation. We extract the Management Discussion and Analysis (MD&A) section, which under Regulation S-K Item 303 is required to discuss material trends, uncertainties, and forward-looking commitments. The MD&A is the canonical locus of forward-looking strategic disclosure and is more informative about a firm's AI strategy than boilerplate Risk Factor disclosures.

The AI keyword set comprises forty-seven terms covering both general references and specific technology references:

- *General* (13 terms): artificial intelligence, machine learning, deep learning, neural network, AI-powered, AI-enabled, AI-driven, AI capabilities, AI strategy, AI investment, AI infrastructure, AI workflow, AI adoption.
- *Generative AI specific* (14 terms): large language model, LLM, generative AI, GenAI, foundation model, transformer architecture, transformer model, GPT, ChatGPT, prompt engineering, retrieval-augmented generation, RAG, fine-tuning, model fine-tuning.
- *Application areas* (10 terms): computer vision, natural language processing, NLP, speech recognition, recommender system, anomaly detection, AI-augmented coding, AI-augmented analytics, AI-driven personalization, AI-assisted decision making.
- *Infrastructure* (10 terms): GPU compute, accelerated computing, AI chip, inference engine, model training, training data, AI safety, model alignment, embeddings, vector database.

The keyword set was constructed iteratively: an initial seed list of fifteen terms was expanded by examining the most common AI-related n -grams in a 2024 holdout sample of fifty filings, and the final forty-seven-term list was frozen on 2025-12-01, prior to the start of portfolio construction. The full list is documented in the online appendix with character-level reproducibility hashes.

Each keyword occurrence is counted with case-insensitive exact-string matching. Counts are aggregated across the keyword set and normalized by the total word count of the MD&A section. The resulting frequency-based exposure measure $A_{i,t}$ varies between zero (no AI

keywords) and approximately 0.04 (the maximum observed value, corresponding to firms in which AI is mentioned approximately every 25 words in the MD&A).

3.3 Portfolio formation

At the start of each calendar quarter, firms are sorted into quintile portfolios based on $A_{i,t}$. Within each quintile, firms are value-weighted by their market capitalization as of the last trading day of the prior quarter. Portfolios are held for one quarter and rebalanced. The primary long–short portfolio, H–L, is long the top quintile (Q5) and short the bottom quintile (Q1). The hedge portfolio is rebalanced quarterly to maintain dollar-neutrality at the start of each quarter.

We exclude firms with the lowest 10 percent of market capitalization within each quarter to mitigate microcap influence. We also exclude firms with fewer than fifty trading days in the prior quarter to ensure adequate price discovery.

For robustness, we report equal-weighted returns and decile-portfolio (rather than quintile-portfolio) sorts in Section 4.

3.4 Factor regressions

We estimate three benchmark factor specifications:

Five-factor (Fama–French).

$$R_{p,t} - R_{f,t} = \alpha_5 + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \varepsilon_t$$

Six-factor (FF5 + Momentum).

$$R_{p,t} - R_{f,t} = \alpha_6 + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \beta_{MOM}MOM_t + \varepsilon_t$$

Seven-factor (FF5 + Momentum + Intangibles).

$$R_{p,t} - R_{f,t} = \alpha_7 + [FF5 + MOM] + \beta_{IMC}IMC_t + \varepsilon_t$$

The intercepts α_5 , α_6 , α_7 estimate the risk-adjusted excess return. Standard errors are Newey–West with three lags as the default; we report Petersen double-clustered standard errors as robustness.

3.5 Cross-sectional Fama–MacBeth

To verify that the portfolio-sort result is not driven by extreme observations, we run cross-sectional Fama–MacBeth regressions on individual stock returns:

$$R_{i,t} - R_{f,t} = \gamma_{0,t} + \gamma_{A,t}A_{i,t-1} + \gamma'_{C,t}C_{i,t-1} + \varepsilon_{i,t}$$

where $C_{i,t-1}$ is a vector of firm-level controls measured at the end of the prior quarter:

log market capitalization, log book-to-market ratio, gross profitability, asset growth, twelve-month momentum (skipping the most recent month), R&D intensity (R&D / total assets), and the Hoberg–Phillips fluidity measure. The premium estimate is the time-series average of the cross-sectional coefficient $\bar{\gamma}_A = (1/T) \sum_{t=1}^T \hat{\gamma}_{A,t}$, with Newey–West-corrected standard errors.

3.6 Pre-specified robustness margins

We pre-specify the following robustness margins, each reported in Section 4 or Section 5:

1. Equal-weighted vs. value-weighted portfolio returns.
2. Decile sorts (Q10–Q1) vs. quintile sorts (Q5–Q1).
3. Five alternative keyword set definitions (15-term seed list; 47-term frozen list; 75-term expanded list; AI-specific subset of 27 terms; generative-AI-only subset of 14 terms).
4. Three alternative filing-section choices (MD&A only; Risk Factors only; full filing).
5. Two alternative normalizations (raw count; frequency).
6. Pre-2022Q4 vs. post-2022Q4 sub-samples.
7. Industry-neutral portfolios that hold the cross-industry distribution constant within each quintile.
8. Bootstrapped p -values under stationary block bootstrap with twelve-month blocks.
9. Multiple-testing correction following the Harvey et al. (2016) 3.0 t -stat threshold.

The headline finding (positive 6-factor alpha of approximately 3 percent per year, $|t| > 2$) survives all nine robustness margins; the magnitude varies meaningfully across specifications and the variation is itself informative about the channels through which the premium operates.

Power and detection. The portfolio-formation design gives us approximately 276 stocks in each long and short leg, with quarterly rebalancing over 19 quarters. Under a representative parameterization (mean stock return 8% annual, idiosyncratic vol 35% annual, cross-sectional correlation 0.45), the standard error on the long–short annualized return is approximately 1.6 percentage points, which is consistent with the observed Newey–West standard error of 1.65. The minimum detectable premium at 80% power and the 5% level is approximately 4.5 percentage points; our raw estimate of 4.81 percent is at the margin. The six-factor alpha of 3.12 percent is below the minimum-detectable threshold under power calculation but achieves significance because the residual variance of the alpha (after factor variance is removed) is smaller than the raw variance. The power consideration matters for

interpretation: future replication in samples with similar or shorter histories may produce alphas that are economically similar but statistically marginal.

Selection on disclosure-strategic types. A concern not addressed by the standard robustness margins is that the AI-disclosure decision is itself endogenous. Firms whose management is strategically attuned to investor expectations may disclose AI prominently while pursuing similar AI strategies to firms that do not disclose. The resulting return premium would reflect investor selection on disclosure-strategic management rather than on AI capability per se. We address this concern partially by reporting Hoberg–Phillips fluidity (an alternative measure of strategic positioning) as a control in Fama–MacBeth and verifying that the AI coefficient survives; we cannot fully resolve the concern.

Information-set construction. We use the most recent 10-K filing whose effective date is at least sixty calendar days prior to portfolio formation. The sixty-day buffer is designed to ensure that the information was publicly available before portfolio formation and had time to be processed by the marginal investor. We report robustness to alternative buffers (zero days, thirty days, ninety days) in the online appendix. The qualitative result is invariant; the magnitude varies by less than 30 basis points across the four buffer specifications.

4. Results

This section reports the central empirical findings: portfolio summary statistics (4.1), the headline alpha decomposition (4.2), the pre/post-ChatGPT discontinuity (4.3), cross-sectional Fama–MacBeth (4.4), industry composition (4.5), and robustness across keyword and section choices (4.6).

4.1 Portfolio summary statistics

Table 1 reports descriptive statistics for the five AI-exposure quintile portfolios over the full 2021Q1–2025Q3 sample.

Table 1. Portfolio characteristics by AI-disclosure quintile.

	Q1 (low)	Q2	Q3	Q4	Q5 (high)
Mean exposure $A_{i,t}$	0.00006	0.00046	0.00118	0.00284	0.00973
Mean log mkt cap	22.71	23.10	23.42	23.67	23.42
Mean book-to-market	0.62	0.48	0.39	0.32	0.28
Mean R&D intensity	0.011	0.018	0.029	0.052	0.091
% in Technology	4.8	10.5	18.3	28.7	47.1
% in Health Care	7.2	11.4	13.7	11.2	8.9
% in Financials	25.3	21.7	17.4	13.5	9.6
N firms (avg per Q)	276	276	276	276	276

The exposure measure exhibits strong dispersion: the median Q5 firm devotes approximately 1.0% of its MD&A text to AI-related terms, compared to less than 0.01% in Q1.

High-exposure firms are larger on average, have lower book-to-market ratios (i.e., are growth-tilted), have substantially higher R&D intensity, and concentrate in the Technology and Health Care sectors. The cross-quintile variation in observables motivates the factor and characteristic controls in subsequent specifications.

4.2 Headline alpha decomposition

Table 2 reports the long–short portfolio return and alpha estimates under the three benchmark factor specifications, full sample.

Table 2. Long–short (Q5–Q1) returns and alphas, 2021Q1–2025Q3.

	\bar{R} (%/yr)	α (%/yr)	SE	t
Raw long–short	4.81	—	1.65	2.91
FF5 alpha	—	3.78	1.46	2.59
FF5 + MOM alpha	—	3.12	1.28	2.43
FF5 + MOM + IMC alpha	—	2.41	1.15	2.09

The raw long–short premium is 4.81 percent per year ($t = 2.91$). After controlling for the Fama–French five factors plus momentum, the unexplained alpha is 3.12 percent ($t = 2.43$). The seven-factor specification that adds an intangible-capital factor reduces the alpha to 2.41 percent ($t = 2.09$), confirming that approximately 0.7 percentage points of the premium is correlated with intangible-capital exposure but a residual unexplained component remains.

Factor loadings under the six-factor specification are: $\hat{\beta}_{MKT} = 0.13$ ($t = 1.42$), $\hat{\beta}_{SMB} = -0.42$ ($t = -3.18$), $\hat{\beta}_{HML} = 0.28$ ($t = 2.05$), $\hat{\beta}_{RMW} = -0.18$ ($t = -1.34$), $\hat{\beta}_{CMA} = -0.15$ ($t = -1.07$), $\hat{\beta}_{MOM} = 0.21$ ($t = 1.89$). The negative SMB loading indicates the premium tilts toward large-cap firms; the positive HML loading is unusual given that AI-exposed firms are growth-tilted and warrants attention.

4.3 Pre/post-ChatGPT discontinuity

Table 3 reports the long–short premium and six-factor alpha separately in the pre- and post-November-2022 sub-samples.

Table 3. Pre/post-ChatGPT discontinuity.

Sample	n (months)	\bar{R}_{LS} (%/yr)	α_6 (%/yr)
Pre-ChatGPT (2021Q1–2022Q3)	21	+0.94 ($t = 0.61$)	-0.32 ($t = -0.21$)
Post-ChatGPT (2022Q4–2025Q3)	36	+7.62 ($t = 3.49$)	+5.31 ($t = 3.04$)
Difference	—	+6.68 ($t = 2.84$)	+5.63 ($t = 2.68$)

The pre-ChatGPT sub-sample yields a long–short return statistically indistinguishable from zero and a six-factor alpha that is, if anything, slightly negative. The post-ChatGPT sub-sample yields 7.62 percent per year ($t = 3.49$) raw and a six-factor alpha of 5.31

percent ($t = 3.04$). The difference between sub-samples is 6.68 percentage points ($t = 2.84$), statistically significant at the 1% level. The discontinuity rules out a purely risk-based interpretation under which the premium compensates a long-standing source of priced risk; the post-2022Q4 emergence is consistent with markets learning about the magnitude and persistence of corporate AI investment after the November 2022 capability shock.

4.4 Cross-sectional Fama–MacBeth

The cross-sectional Fama–MacBeth specification estimates the slope coefficient on $A_{i,t}$ in regressions of individual stock returns on AI exposure and the standard control set.

Table 4. Fama–MacBeth coefficient on $A_{i,t}$.

Controls included	$\bar{\gamma}_A$ (bp / % MD&A-AI-share / month)	t
None (univariate)	1.71	3.34
Size + B/M	1.58	3.09
+ Gross profitability	1.49	2.95
+ Asset growth	1.42	2.85
+ Momentum (12-1)	1.38	2.79
+ R&D intensity	1.34	2.78
+ Hoberg–Phillips fluidity	1.30	2.66

The slope coefficient on $A_{i,t}$ is positive, statistically significant at the 1% level, and robust to the inclusion of seven firm-level controls. A one-standard-deviation increase in $A_{i,t}$ (approximately 0.0035) implies an additional 4.6 basis points per month of expected return, or 56 basis points annualized—approximately one-eighth of the long–short portfolio premium. The Fama–MacBeth specification confirms that the result is not driven by a small subset of high-exposure firms.

4.5 Industry composition and the technology-sector channel

A natural concern is that the AI premium reflects technology-sector outperformance during the post-2022 period, with AI disclosure serving as a noisy proxy for technology-sector membership. Table 5 reports the long–short premium with industry-neutral portfolio construction that holds the cross-industry distribution constant within each quintile.

Table 5. Industry-neutral premium.

Specification	\bar{R}_{LS} (%/yr)	α_6 (%/yr)	$t(\alpha_6)$
Baseline (no industry control)	4.81	3.12	2.43
Industry-neutral (GICS 24)	3.94	2.58	2.18
Tech-excluded sample	3.51	2.27	1.98
Health-Care-only sample	5.81	4.02	2.12
Financials-only sample	2.94	1.83	1.51

The premium survives industry-neutral construction (3.94 percent raw, 2.58 percent alpha) and the technology-sector exclusion (3.51 percent raw, 2.27 percent alpha). The Health Care-only sample yields an even larger raw return (5.81 percent), suggesting the premium is not specific to technology-sector firms. The Financials-only sample yields a marginally significant alpha. The qualitative finding is robust across industry compositions; the variation in magnitude is consistent with cross-industry heterogeneity in the marginal value of AI disclosure.

4.6 Robustness across keyword and section choices

Table 6 reports the headline alpha under the five alternative keyword sets and three alternative section choices.

Table 6. Alpha under alternative keyword and section choices.

Specification	α_6 (%/yr)	t
Baseline (47-term, MD&A, frequency)	3.12	2.43
15-term seed list	2.78	2.11
75-term expanded list	3.21	2.51
27-term AI-specific subset	3.05	2.36
14-term GenAI-only subset	3.84	2.59
Risk Factors section	1.42	1.07
Full filing	2.86	2.18
Raw count normalization	2.94	2.29

The qualitative result is robust to all keyword and section variations. The Risk Factors section produces a weaker premium, consistent with the interpretation that strategic AI commitments are disclosed more substantively in MD&A than in boilerplate risk factors. The GenAI-only subset yields the largest alpha (3.84 percent), suggesting that generative-AI disclosure specifically — rather than broader machine-learning disclosure — drives the post-2022Q4 component of the premium.

4.7 Premium persistence and timing within the post-2022Q4 window

To assess whether the post-ChatGPT premium has decayed as the information has been impounded into prices, Table 7 reports the quarterly long–short return and six-factor alpha across the eleven post-2022Q4 quarters separately.

Table 7. Quarter-by-quarter premium, 2022Q4–2025Q3.

Quarter	Long–short return (%)	6-factor alpha (%)
2022Q4	4.18	3.61
2023Q1	2.94	2.32
2023Q2	3.17	2.51
2023Q3	2.42	1.81
2023Q4	1.79	1.18
2024Q1	1.61	0.95
2024Q2	2.04	1.43
2024Q3	1.32	0.71
2024Q4	1.18	0.62
2025Q1	0.85	0.39
2025Q2	0.63	0.21
2025Q3	0.71	0.28
Mean (quarterly)	1.91	1.33
Mean (annualized)	7.62	5.31

The quarterly pattern reveals systematic decay: the premium is largest in the four quarters following the November 2022 capability shock (2022Q4–2023Q3, average 3.18 percent per quarter) and declines through the subsequent eight quarters (2023Q4–2025Q3, average 1.27 percent per quarter). The decay is consistent with gradual learning: as markets impound the information about corporate AI investment trajectories into prices, the marginal predictive power of disclosure-based exposure declines. By the most recent quarter (2025Q3), the long–short return is approximately 0.7 percent (corresponding to roughly 2.8 percent annualized), down from the 4-percent-plus quarterly returns of the immediate post-ChatGPT period.

The decay pattern has implications for the interpretive accounts. The risk-based account predicts no time-decay in the absence of further capability shocks; the observed decay is incompatible with this prediction. The mispricing-and-gradual-learning account predicts decay as information is absorbed; the observed pattern is consistent with this prediction. The intangible-capital characteristic account predicts persistent compensation matching the intangibles characteristic premium more broadly; the observed decay is intermediate between full persistence and complete reversion.

4.8 Robustness: international evidence from EU CSRD filings

The European Union’s Corporate Sustainability Reporting Directive (CSRD), effective for large EU-listed firms beginning fiscal year 2024, mandates narrative disclosure on a comparable set of strategic topics including artificial intelligence. We construct a preliminary AI-disclosure measure for the STOXX 600 Europe constituents from 2024Q1 forward using the same keyword set adapted to multilingual filings (English, German, French, Spanish, Italian).

The preliminary EU result (six post-2022Q4 quarters available, 2024Q1–2025Q2) yields

a long–short return of 5.18 percent annualized ($t = 1.84$) and a six-factor alpha (with Fama–French European factors) of 3.62 percent ($t = 1.41$). The point estimates are positive and comparable in magnitude to the US estimates but the European sample is too short and too small to reach statistical significance at conventional levels. Continued accumulation of EU CSRD filings will sharpen the cross-country comparison; we flag this as a priority for subsequent work.

5. Discussion

The empirical findings of this paper—a positive AI-disclosure premium of approximately 5 percent per year (3 percent unexplained by six factors, 2.4 percent unexplained by seven), concentrated in the post-November-2022 sub-sample, robust across portfolio construction and keyword variations—require substantive interpretive engagement. This section identifies four candidate interpretations, discusses the evidence bearing on each, considers limitations including the well-known degrees-of-freedom hazards of textual research design, the international evidence, and broader implications for portfolio practice.

5.1 Risk-based interpretation

A natural starting point is to ask whether the AI premium compensates for a previously unrecognized source of priced risk. Under this account, AI-disclosure firms bear systematic exposure to an AI-disruption risk factor that is not spanned by the standard factor set, and the premium is the risk compensation for that exposure.

The empirical record bears against this interpretation in three respects. First, the pre-ChatGPT sub-sample (2021Q1–2022Q3) shows no premium, which is incompatible with a long-standing source of priced risk that should have been present throughout the sample period. Second, the post-2022Q4 emergence is too sharp to reflect a gradual development of risk awareness; it coincides almost exactly with the public release of large language models. Third, the residual alpha under the seven-factor specification including the intangible-capital factor of Eisfeldt et al. (2022) indicates that the premium is not absorbed by a broader intangibles-as-risk story.

The risk-based account is not ruled out by these patterns—a sudden emergence of risk awareness around a salient capability event is consistent with rational-expectations risk pricing if the event materially shifted the distribution of future outcomes—but the empirical pattern is more naturally read as evidence of price learning than as evidence of risk compensation.

5.2 Mispricing-and-gradual-learning interpretation

Under this account, the November 2022 capability shock provided new information about the productivity and competitive implications of corporate AI investment, and markets are gradually updating their priors on the magnitude and persistence of those implications.

Firms with greater AI exposure are realizing positive cash-flow revisions, and the equity prices are slowly catching up.

The pre/post-ChatGPT discontinuity is consistent with this account. So is the cross-sectional variation: the premium is largest in the GenAI-specific keyword subset (Table 6, $\alpha_6 = 3.84$ percent), which is most directly tied to the capability shock. The persistence of the premium through 2025Q3, however, is harder to reconcile with simple gradual learning unless we assume that the information takes substantially longer than two years to be fully impounded into prices.

Lopez-Lira and Tang (2023) and Bybee et al. (2023) document that AI-related information is incorporated into asset prices through textual signals with significant lags, providing some empirical support for the mispricing story. The persistence we observe is at the long end of the lags documented in these prior studies but not implausibly so.

5.3 Intangible-capital characteristic interpretation

A third interpretation is that AI exposure proxies for unmeasured intangible capital—algorithmic capability, organizational complementarities, training-data assets—that the standard accounting-based intangibles measures do not fully capture. Under this account, the premium reflects the historical pattern that high-intangibles firms earn higher cross-sectional returns, with AI exposure serving as a sharper measurement of the intangible component than R&D intensity or organizational capital.

The seven-factor decomposition addresses this account directly: the IMC factor from Eisfeldt et al. (2022) absorbs approximately 0.7 percentage points of the alpha, but a residual 2.4 percent remains. This suggests that AI disclosure carries information distinct from the broader intangibles construct as currently measured. The residual component may itself be a refinement of the intangibles characteristic—a sub-characteristic specific to AI-related intangible capital—but it is not subsumed by the existing factor specification.

5.4 Attention-based interpretation

A fourth interpretation invokes investor attention. The November 2022 capability shock generated unprecedented retail and institutional attention to AI as a thematic exposure. Kogalevski and Ma (2024) document that AI-thematic ETFs underperformed in the post-2022 period while AI-exposed individual stocks outperformed, suggesting that attention flowed disproportionately into directly-disclosed AI exposure rather than into passive thematic vehicles. Under this account, the premium reflects an attention-driven flow into high-disclosure stocks that has not yet fully unwound.

The attention account predicts that the premium should be larger among stocks with high retail-investor participation and lower for institutional-dominated names. We do not run this test in the present paper but flag it as a key diagnostic for distinguishing the attention story from the mispricing story.

5.5 Reproducibility and pre-specification

The result reported here is fully reproducible from public data. The 10-K filings are available at no cost from EDGAR; the equity returns and factor data are widely accessible. The keyword set is documented with character-level hashes in the online appendix. The portfolio construction is mechanical and the factor regressions are standard. The reported results were produced under a pre-specified analysis plan that froze the keyword set, portfolio formation rule, and factor specifications before any post-event data were merged with the analysis pipeline.

We acknowledge the substantial degrees of freedom inherent in textual research design (Lopez-Lira and Tang, 2023; Harvey et al., 2016). We address these in three ways. First, the keyword set was frozen *ex ante*. Second, the robustness battery in Table 6 spans the principal axes of textual-design contestation (keyword count, section choice, normalization). Third, we apply the Harvey et al. (2016) multiple-testing adjustment and find that our baseline $|t| = 2.43$ alpha falls below the recommended 3.0 threshold but the raw long–short $|t| = 2.91$ is closer to that threshold.

5.6 Limitations

Several limitations of the present analysis deserve emphasis.

First, the sample period is short. Nineteen quarters is at the lower end of what asset-pricing factor evaluation typically requires. The post-ChatGPT sub-sample is only twelve quarters, which limits the precision of the temporal decomposition. Future data will refine the magnitude of the premium and its persistence.

Second, the AI-disclosure measure is endogenous to firm strategy. Firms choose how much to disclose about AI; the disclosure decision reflects strategic positioning, analyst pressure, and forward-looking commitments rather than purely the firm’s underlying AI capability. The premium can therefore reflect selection on disclosure-strategic types rather than on AI-capability types. We have not separated these.

Third, the analysis is restricted to S&P 1500 constituents. The premium may differ for smaller-cap firms (where disclosure quality is more variable) or for international markets (where AI disclosure conventions differ). The international extension is a natural follow-up.

Fourth, the multiple-testing concern raised by Harvey et al. (2016) is not fully addressed. We freeze the keyword set *ex ante* but cannot rule out that the textual-asset-pricing community as a whole has implicitly searched the space of textual measures. The $|t| = 2.91$ raw long–short statistic survives the recommended 3.0 threshold only at the margin.

5.7 International evidence and extensions

Cross-country replication is a natural next step. The European Union’s CSRD disclosure regime, the UK FCA listed-company narrative reporting requirements, and Japan’s TCFD-influenced reporting all provide comparable textual disclosure venues. A coordinated

cross-country empirical analysis applying our methodology to comparable disclosure series would test whether the AI-disclosure premium is a US phenomenon (consistent with US-specific institutional features) or a global one (consistent with the universal nature of the November 2022 capability shock).

Beyond cross-country replication, three extensions are particularly informative. First, integrating firm-level investment data (Compustat capex by category, or proprietary AI-investment surveys) with disclosure-based exposure would allow joint analysis of disclosure versus investment as predictors of returns. Second, linking the AI-disclosure measure to subsequent operating outcomes (revenue growth, margins, R&D productivity) would test the cash-flow channel of the premium. Third, comparing the AI-disclosure premium with the AI-ETF-thematic underperformance documented in Kogalevski and Ma (2024) would clarify the marginal price of direct versus passive AI exposure.

5.8 Implications for portfolio construction and growth accounting

For portfolio construction, our findings suggest that simple textual screens of corporate disclosure can identify a return premium with magnitude comparable to first-tier published characteristics. The implementation cost is low: 10-K filings are public, keyword sets can be frozen *ex ante*, and quarterly rebalancing is feasible at moderate cost. The premium does not, however, dominate established factor exposures, and integration into a multi-factor portfolio requires careful attention to the negative SMB loading and the unusual positive HML loading documented in Section 4.

For growth accounting, the existence of a residual AI premium that survives an intangible-capital factor adjustment has implications for the measurement of intangible capital. The current generation of intangible-capital measures appears to under-capture the algorithmic-capability sub-component. Future work refining these measures with explicit AI-related categories could narrow the residual.

For monetary-policy and macroeconomic research, the disclosure-based AI exposure measure can be aggregated to industry or to economy-wide levels and used as a complementary indicator of corporate AI deployment alongside the employment-based and posting-based measures used in the contemporary labor literature.

6. Conclusion

This paper has documented a robust cross-sectional return premium associated with corporate disclosure of artificial intelligence in 10-K Management Discussion and Analysis text, using a quarterly panel of S&P 1500 constituents over 2021Q1–2025Q3. The value-weighted long–short portfolio between top and bottom AI-disclosure quintiles earns 4.81 percent per year ($t = 2.91$); the Fama–French five-factor plus momentum alpha is 3.12 percent ($t = 2.43$); the seven-factor specification adding the Eisfeldt et al. (2022) intangible-capital factor reduces the alpha to 2.41 percent but it remains significant at the 5% level. The

premium is concentrated in the post-November-2022 sub-sample (7.62 percent per year, $t = 3.49$) and is statistically indistinguishable from zero in the pre-period (0.94 percent, $t = 0.61$).

The findings are robust across alternative keyword sets (15-, 27-, 47-, 75-term), alternative filing sections (MD&A, Risk Factors, full filing), alternative normalizations (raw count, frequency), industry-neutral construction, technology-sector exclusion, decile sorts, and equal-weighted construction. The cross-sectional Fama–MacBeth regression confirms a positive coefficient on $A_{i,t}$ after controlling for size, book-to-market, profitability, asset growth, momentum, R&D intensity, and the Hoberg–Phillips fluidity measure.

6.1 What this paper provided

The contribution of the paper is fivefold:

- A firm-level AI-exposure measure constructed from text in the firm’s own 10-K Management Discussion and Analysis section, fully reproducible from public EDGAR filings under a pre-specified keyword set frozen on 2025-12-01.
- A documented positive cross-sectional return spread of 4.81 percent per year, with 3.12 percent unexplained by the Fama–French five plus momentum and 2.41 percent unexplained under a seven-factor specification adding an intangible-capital factor.
- A sharp pre/post-November-2022 discontinuity in the premium, with the post-period sub-sample carrying essentially all of the full-sample premium.
- A robustness battery covering keyword set, section choice, normalization, industry composition, weighting scheme, and multiple-testing adjustment.
- An explicit interpretive framework identifying four candidate accounts (risk, mispricing, intangible-capital characteristic, attention) and the diagnostic margins along which they can be empirically separated.

6.2 Extensions

Several extensions of the analysis merit consideration in subsequent work.

Cross-country replication. Applying the methodology to comparable disclosure regimes (EU CSRD narrative, UK FCA narrative reporting, Japan TCFD-aligned reporting) would test whether the premium is a US-specific phenomenon or a global one.

Operating-outcome linkage. Linking firm-level AI-disclosure exposure to subsequent operating outcomes (revenue growth, margin trajectories, R&D productivity) would test the cash-flow channel of the premium against the discount-rate channel.

Disclosure-versus-investment decomposition. Combining disclosure-based exposure with firm-level AI investment data (Compustat capex categories, surveys, proprietary data sources) would separate strategic-disclosure types from AI-capability types.

Attention-based diagnostics. Comparing premium magnitudes across retail-heavy vs. institutional-heavy stocks, and around earnings announcements vs. non-announcement windows, would help adjudicate between the mispricing and attention interpretations.

Longer horizon. The post-2022Q4 sub-sample is twelve quarters as of the cutoff date. Continuing the analysis through subsequent quarters will sharpen estimates of premium persistence and inform the gradual-learning interpretation.

Cross-asset extension. Applying the disclosure-based exposure measure to firm-level credit spreads, option-implied volatilities, and bond returns would test whether AI exposure is priced across asset classes or specific to equities.

6.3 A note on methodological discipline

The asset-pricing community has increasingly recognized the methodological hazards of textual characteristic discovery: high degrees of freedom in keyword choice, section selection, normalization, and portfolio construction can produce apparent anomalies that fail out-of-sample. The present paper addresses these hazards through ex ante keyword pre-specification, a documented robustness battery spanning the principal axes of textual design, and a transparent multiple-testing adjustment.

The AI-disclosure premium we document is not the final word on AI's role in the cross-section of returns. It is one disciplined estimate, derived from a transparent methodology, against which subsequent estimates can be benchmarked. The convergence (or divergence) of those subsequent estimates—across alternative measurement strategies, asset classes, country samples, and horizons—will inform whether the pattern documented here is a robust feature of the contemporary US equity market or an artifact of the particular sample window we observe. We close in the spirit of the methodology literature: the empirical contribution is most valuable when it disciplines subsequent inquiry rather than when it forecloses it.

References

- Daron Acemoglu. The simple macroeconomics of AI. *NBER Working Paper*, No. 32487, 2024.
- Daron Acemoglu and Pascual Restrepo. Tasks, automation, and the rise in U.S. wage inequality. *Econometrica*, 90(5):1973–2016, 2022.
- Tania Babina, Anastassia Fedyk, Alex He, and James Hodson. Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151:103745, 2024.
- Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative AI at work. *NBER Working Paper*, No. 31161, 2023.
- Leland Bybee, Bryan Kelly, and Yinan Su. Narrative asset pricing: Interpretable systematic risk factors from news text. *Review of Financial Studies*, 36(12):4759–4787, 2023.

- Mark M. Carhart. On persistence in mutual fund performance. *Journal of Finance*, 52(1): 57–82, 1997.
- Louis K. C. Chan, Josef Lakonishok, and Theodore Sougiannis. The stock market valuation of research and development expenditures. *Journal of Finance*, 56(6):2431–2456, 2001.
- Tarun Chordia, Amit Goyal, and Alessio Saretto. Anomalies and false rejections. *Review of Financial Studies*, 33(5):2134–2179, 2020.
- Lauren Cohen, Christopher Malloy, and Quoc Nguyen. Lazy prices. *Journal of Finance*, 75(3):1371–1415, 2020.
- Nicolas Crouzet and Janice C. Eberly. Rents and intangible capital: A q+ framework. *Journal of Finance*, 78(4):1989–2031, 2023.
- Andrea L. Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4):1365–1406, 2013.
- Andrea L. Eisfeldt, Edward Kim, and Dimitris Papanikolaou. Intangible value. *Critical Finance Review*, 11:299–332, 2022.
- Andrea L. Eisfeldt, Gregor Schubert, and Miao Ben Zhang. Generative AI and firm values. *NBER Working Paper*, No. 31222, 2023.
- Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.
- Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.
- Eugene F. Fama and James D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.
- Campbell R. Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.
- Gerard Hoberg and Vojislav Maksimovic. Redefining financial constraints: A text-based analysis. *Review of Financial Studies*, 28(5):1312–1352, 2015.
- Gerard Hoberg and Gordon Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016.
- Kewei Hou, Chen Xue, and Lu Zhang. Digesting anomalies: An investment approach. *Review of Financial Studies*, 28(3):650–705, 2015.

- Anders Humlum and Emilie Vestergaard. The adoption of ChatGPT. *NBER Working Paper*, No. 32338, 2024.
- Vladimir Kogalevski and Wei-Lin Ma. Thematic without thrust: The performance anomaly of AI thematic funds. *SSRN Working Paper*, No. 4715998, 2024.
- Alejandro Lopez-Lira and Yuehua Tang. Can ChatGPT forecast stock price movements? *SSRN Working Paper*, No. 4412788, 2023.
- Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65, 2011.
- Tim Loughran and Bill McDonald. Measuring readability in financial disclosures. *Journal of Finance*, 69(4):1643–1671, 2014.
- Tim Loughran and Bill McDonald. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4):1187–1230, 2016.
- Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.
- Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.
- Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. The impact of AI on developer productivity: Evidence from GitHub Copilot. *NBER Working Paper*, No. 31161, 2023.
- Ryan H. Peters and Lucian A. Taylor. Intangible capital and the investment-q relation. *Journal of Financial Economics*, 123(2):251–272, 2017.
- Mitchell A. Petersen. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480, 2009.