Do AI Adoption Signals Predict Company Performance?

Evidence from a 500-Company Cross-Section of Scores, Filings, and Hiring Data

Yixiao Li Siru Tao Xin Xu
Hanzhe Hong

Carnegie Mellon University — AI Capstone Program
in collaboration with Larridin

July 2026 — Working Draft

Abstract

Enterprises are investing heavily in artificial intelligence, yet there is little systematic evidence on whether observable AI-adoption signals translate into measurable business outcomes. We study this question using a cross-section of roughly 500 large U.S. public companies, combining (i) proprietary AI-adoption scores from Larridin's AI Transformation Tracker, (ii) two novel signal families that we construct from primary sources—LLM-extracted disclosure signals from 10-K filings and an AI-hiring intensity signal built from 30,000+ classified job postings—and (iii) realized financial outcomes from SEC filings and market data. The analysis sample excludes the five largest AI-semiconductor firms (Nvidia, Broadcom, AMD, Micron, Intel) so that no handful of names drives the cross-section; every finding is unchanged on the full sample (appendix). Three findings emerge. First, the measurement thesis is validated: every score- and disclosure-based adoption signal in the study is significantly associated with revenue growth, and the association strengthens with measurement depth. Its sharpest expression is *narrative concreteness*—a new dimension we construct within the same evidence-first measurement framework, capturing whether a firm reports named, deployed use cases with quantified results rather than aspirational language. Moving from the least to the most concrete disclosure is associated with 8.0 percentage points higher year-over-year revenue growth ($p < 0.01$), an effect that survives sector, size, and growth-momentum controls; composite scores, which aggregate several scale-correlated components, capture the same phenomenon unconditionally, and isolating their disclosure-concreteness component is what sharpens the signal. Second, none of the public signals predicts risk-adjusted stock returns over the subsequent four months, consistent with markets already impounding public AI-adoption information; margin effects are likewise absent, indicating that AI currently operates through the growth channel rather than the cost channel. Third, an AI value-chain decomposition shows the association is present *within* physical-asset-heavy late adopters ($p < 0.05$)—precisely where AI narratives vary most in credibility and structured measurement is most needed—and that AI-infrastructure suppliers outperformed sector- and size-matched peers by 32 percentage points over the sample window. We contribute a validated, reproducible pipeline for measuring corporate AI adoption

from public data, an immediately deployable new scoring dimension, and the first test–retest reliability estimates for signals of this class.

1 Introduction

Every large enterprise now claims an artificial-intelligence strategy. Capital expenditure guidance, annual reports, and hiring pages are saturated with AI language, and a measurement industry has emerged to score companies on their "AI readiness." Yet executives, investors, and analysts still lack a rigorous, data-driven answer to a basic question: *do observable AI-adoption signals carry information about subsequent business performance, or are they merely narrative?*

This paper studies that question empirically. Our starting point is the AI Transformation Tracker built by Larridin, which assigns evidence-backed 1–5 scores to 562 public companies across three pillars—AI adoption, workforce proficiency, and realized impact—together with an overall maturity index. We link these scores to market data and SEC fundamentals for a 564-company universe (the Larridin coverage set united with the S&P 500), and we extend the signal set with two families of new measures, built at the sponsor's direction as extensions of the same evidence-first measurement methodology that underlies the Tracker:

1. **Filing-disclosure signals.** Using a large language model (LLM) extraction pipeline with mandatory evidence citation, we score 478 companies' most recent 10-K filings on three dimensions: *AI investment intensity* (the scale of resources committed to building or buying AI), *narrative concreteness* (whether the AI story consists of named, deployed use cases with quantified results, versus aspirational boilerplate), and *risk-disclosure depth*.

2. **AI-hiring intensity.** We collect and individually classify 30,861 job postings across 536 companies in two monthly snapshots (June and July 2026), measuring the share of each firm's hiring that targets AI-builder roles. The posting classifier is validated against 657 independently labeled postings ($\sim$90% agreement on the core class), and the two snapshots yield the first test–retest reliability estimate we are aware of for this signal class (Spearman $\rho = 0.54$).

We then ask three questions. (1) Do AI-adoption signals predict *operational* outcomes—revenue growth, margin change, productivity? (2) Do they predict *market* outcomes—risk-adjusted forward returns? (3) Where in the AI value chain are any effects concentrated?

Findings. Our headline result is best summarized as *measurement depth pays*. Every score- and disclosure-based signal in the study—the composite Tracker scores and our filings dimensions—is significantly associated with subsequent revenue growth unconditionally (the hiring signal postdates the outcome quarter and is excluded from this claim by construction), validating the underlying thesis that AI adoption is measurable from public sources and economically meaningful. As controls tighten, the signals then separate by granularity. Composite scores,

which by construction aggregate several scale-correlated components (deployment breadth, vendor footprint, enablement infrastructure), largely attenuate as sector and size controls enter. The disclosure-concreteness dimension—the deepest, most evidence-anchored measurement in the set—survives everything: its full-specification coefficient of 0.080 ($p = 0.009$) implies that moving from the bottom to the top of the concreteness distribution is associated with 8.0 percentage points higher revenue growth, holding sector, size, and momentum fixed, and the effect passes a Benjamini–Hochberg false discovery correction within our pre-specified primary hypothesis family. Investment intensity is positive throughout but loses significance once size enters. The practical reading is constructive: the information is real, and it concentrates in verifiable specifics—which is precisely what a structured, evidence-first measurement system is built to capture, and what naive keyword-counting cannot.

On the market side we find nothing: no signal predicts four-month forward returns after controls and multiple-testing correction. We interpret this through the lens of market efficiency—all our signals are constructed from public information, and the point estimate on the composite score is, if anything, mildly negative over a window in which early-2026 AI-theme valuations partially unwound. Margin effects are likewise absent: in this sample, AI adoption is a top-line story, not (yet) a cost story.

Finally, a four-category AI value-chain classification of the S&P 500 (infrastructure suppliers, software platforms, data-rich adopters, and physical-asset-heavy late adopters) reveals two forms of heterogeneity. First, AI-infrastructure suppliers outperformed sector- and size-matched peers by approximately 32 percentage points over four months—a striking decomposition of the H1-2026 AI rally. Second, and more subtly, the concreteness–growth association is significant *within* the physical-asset-heavy late-adopter group ($n = 214$, $p = 0.025$): among companies where AI claims are cheapest to make and hardest to verify, concrete disclosure best separates genuine adopters from laggards.

Contributions. Beyond the empirical findings, we contribute methodology and infrastructure: (i) an evidence-first LLM extraction protocol in which every score must cite a verbatim source span (87–90% of citations verify against source text), (ii) a validated, repeatable monthly pipeline for AI-hiring measurement, including classifier validation against independent labels and cross-month reliability estimates, and (iii) an honest accounting of what a single-vintage score history can and cannot identify, which we hope is useful to the emerging AI-transformation-measurement industry.

2 Data

2.1 Universe construction

Our universe is the union of Larridin's January-2026 coverage set and the S&P 500. Starting from 567 companies in the Larridin database, we remove test records and same-company duplicates (e.g., listings under both "AMD" and "Advanced Micro Devices"), yielding 542 distinct

companies, of which 14 are privately held or subsidiaries with no traded equity. We add 22 S&P 500 constituents not covered by Larridin, for a study universe of 564 companies spanning all twelve sectors. Every listed company is mapped to its ticker and SEC Central Index Key through a three-stage procedure (exact name match against the S&P constituent list, exact match against the SEC registry, then a conservative fuzzy pass with manual review); all 542 Larridin companies resolve. Where one company lists multiple share classes, one line is retained. From this universe we form the **analysis sample by excluding the five largest AI-semiconductor firms**—Nvidia, Broadcom, AMD, Micron, and Intel—whose extreme H1-2026 revenue and return realizations would otherwise let a few names dominate a cross-section of this size. All main-text results use this exclusion sample; the full sample retaining these five firms is reported in Appendix Table A5, where every conclusion is unchanged.

2.2 Larridin AI-adoption scores

Larridin's AI Transformation Tracker assigns each company three pillar scores on a 1–5 scale—*adoption* (deployment breadth, hiring intensity, vendor footprint, governance), *proficiency* (workforce fluency, enablement infrastructure, leadership depth), and *impact* (quantified outcomes, financial linkage, repeatability)—plus a composite *maturity index*. Scores are generated by an LLM research pipeline over public sources with per-dimension evidence citations and confidence labels. Our vintage was assessed on January 18–19, 2026 and covers 562 companies (538 after our deduplication), with a healthy cross-sectional distribution (maturity mean 3.47, full 1–5 support). Because a single assessment vintage was available at the time of study, our design is cross-sectional; we return to the implications in Section 6.

2.3 Outcome variables

Fundamentals. From SEC XBRL company facts we extract, for each company, the first fiscal quarter ending after the score date (for most companies, calendar Q1 2026, reported April–May 2026). Our primary outcome is **year-over-year revenue growth** for that quarter (available for 438 companies in the regression sample), which nets out seasonality. Secondary outcomes are the year-over-year **operating-margin change** ($n = 348$) and revenue per employee where disclosed. Values are taken as first reported (avoiding restatement look-ahead) and winsorized at the 1st/99th percentiles.

Market outcomes. We compute forward returns from the first trading day after the score date (January 20, 2026) at one- through four-month horizons, using adjusted closes for 531 tickers. Nine companies delisted through M&A or bankruptcy during 2025 are excluded by construction; one mid-window delisting retains its truncated horizons.

2.4 Control variables

Regressions control for **sector** (twelve fixed effects), **firm size** (log market capitalization at the score date, computed from SEC-reported shares outstanding times the first post-score close for 490 companies, with a current-cap fallback for 36), and **growth momentum** (year-over-year revenue growth of the last fiscal quarter ending *before* the score date, $n = 517$)—the observable growth an investor could see at signal time. The momentum control addresses the central confound that faster-growing firms both adopt more AI and continue to grow.

3 Signal Construction

3.1 Filing-disclosure signals

For each company we retrieve the most recent 10-K from SEC EDGAR, extract AI-relevant passages by keyword windowing (with section tags so that risk-factor material is identified as such), and score three dimensions on anchored 1–5 rubrics using a frontier LLM at temperature zero:

- **AI investment intensity**: the scale of resources committed to AI, whether *building* it (AI R&D, chips, platforms) or *buying/deploying* it (compute, cloud, talent). Anchors range from "investment described as minimal or paused" (1) to "quantified, sustained, or core-business-scale commitment" (5). Generic R&D with no stated AI link does not count.
- **Narrative concreteness**: 1 = buzzwords only; 3 = named use cases without deployment detail; 5 = deployed use cases with quantified business results.
- **Risk-disclosure depth**: from no AI risk discussion to quantified, governance-backed treatment. A scope guard prevents scoring "absence" when risk-section material simply was not captured by the excerpting step.

The protocol is *evidence-first*: a non-null score must cite verbatim supporting spans, and dimensions without evidence are returned as null rather than guessed. In an automated audit, 87% of evidence quotes reproduce verbatim in the source filings (the remainder being ellipsized quotations). Of 488 companies with a U.S. 10-K, 478 yielded scores; ten filings contain no AI-relevant content, which we record as missing rather than as a low score. Prompts are versioned; the investment-intensity rubric was revised once (v1→v2) after review showed the buyer-side framing understated AI *producers* (e.g., a leading AI-chip designer initially scored 3), and the full universe was re-scored under v2. Distributionally, concreteness spans the full 1–5 range; risk-disclosure is heavily concentrated at 4 (75% of firms), reflecting the post-2024 standardization of AI risk language—we therefore treat it as descriptive rather than discriminating.

3.2 AI-hiring intensity

We measure AI-directed hiring demand in two monthly snapshots (windows May 10–June 10 and June 4–July 5, 2026). For each company we retrieve up to 25 recent job-posting search results (60 for the 150 most AI-active companies in the July snapshot) via a neural web-search API, then classify every result with an LLM against a four-way taxonomy aligned with Larridin's own labeling scheme: *AI-builder* (the role's core function is building or operating AI/ML systems), *AI-user*, *AI-leader*, and *non-AI*, with conservative defaults, employer-match filtering, and per-posting evidence. Company-level intensity is the AI-builder share of matched postings (*builder rate*); 30,861 postings were classified in total.

Validation. Before any full-scale run, the classifier was validated against 657 postings independently collected and labeled by Larridin's own crawler-based pipeline: agreement on the AI-builder class is approximately 90%, and a stronger scoring model yields identical recall on the same inputs, indicating performance is input- rather than model-limited. At the company level, our June intensities reproduce the ordering of Larridin's proof-of-concept measurements for the overlapping companies (Spearman $\rho = 0.83$). Across the two snapshots, company builder rates correlate at $\rho = 0.54$ ($n = 226$ companies with at least five matched postings in both months)—to our knowledge the first published test–retest reliability figure for a posting-derived AI-hiring signal. Cross-month comparisons use within-month percentile ranks, as the underlying search index is not level-stationary across months.

3.3 AI value-chain classification

Finally, each S&P 500 company is assigned one of four AI value-chain categories—*AI Infrastructure & Hardware* (36), *AI Software & Platforms* (41), *Data-Rich Adopters* (138), and *Physical-Asset-Heavy / Late Adopters* (285)—by a frontier LLM with a structured rubric, confidence, and rationale. Unlike the signals above, this classification draws on the model's general knowledge of each firm rather than cited primary evidence; we therefore use it only as a segmentation device, not as a predictive signal.

4 Methodology

Our design is a single cross-section: signals measured at or before early 2026 are related to outcomes realized afterward. Signals are transformed to cross-sectional percentile ranks (robust to outliers and scale), so coefficients read as the outcome difference between the lowest- and highest-ranked firm. For each signal s_i and outcome y_i we estimate a ladder of OLS specifications

with HC3 heteroskedasticity-robust standard errors:

$$
\begin{aligned}
(1) \quad & y_i = \alpha + \beta\, s_i + \varepsilon_i \\
(2) \quad & y_i = \alpha + \beta\, s_i + \gamma_{\text{sector}(i)} + \varepsilon_i \\
(3) \quad & y_i = \alpha + \beta\, s_i + \gamma_{\text{sector}(i)} + \delta \log \text{MktCap}_i + \varepsilon_i \\
(4) \quad & y_i = \alpha + \beta\, s_i + \gamma_{\text{sector}(i)} + \delta \log \text{MktCap}_i + \lambda\, g_i^{\text{prior}} + \varepsilon_i
\end{aligned}
$$

where g_i^{prior} is pre-signal revenue growth. Surviving specification (4) requires a signal to carry information beyond sector membership, firm size, and growth persistence. Because a handful of AI-semiconductor mega-caps posted revenue and return realizations far outside the rest of the cross-section in this window, our analysis sample excludes the five largest of them—Nvidia, Broadcom, AMD, Micron, and Intel—so that no small cluster of names can drive the estimates; the full sample including these firms is reported as a robustness appendix (Table A5) and leaves every conclusion intact (the headline concreteness coefficient is 0.080 here versus 0.087 with them). Each estimate uses all firms with complete data for that specification (missingness is documented and non-random: banks lack comparable margin data, foreign filers lack 10-Ks); a complete-case robustness sample is reported in the appendix. Revenue growth was designated the primary outcome before testing; within the primary family of five signal tests we apply Benjamini–Hochberg FDR control at $q = 0.05$, and we also report the correction across the full five-signal, three-outcome (fifteen-test) matrix. Rank-based information coefficients and quintile portfolio sorts complement the regressions.

Because job postings are collected live and cannot be reconstructed retrospectively, hiring signals postdate the fundamental outcome quarter; we therefore report hiring associations as validation and dynamics rather than prediction, and we flag that the outcome quarter (typically January–March 2026) overlaps the mid-January score date by roughly two weeks, making "near-term association" the precise claim for fundamentals.

5 Results

5.1 Do AI signals relate to operational performance?

Table 1 reports the specification ladder for the primary outcome. Three patterns stand out. First, **every** score- and filings-based signal is strongly associated with revenue growth unconditionally (column 1): firms that score high on any AI dimension grew faster. The phenomenon the Tracker is built to measure is present in the data. Second, the signals separate by measurement granularity as controls tighten: the composite scores—which aggregate several components that naturally scale with firm size—remain significant with sector controls but attenuate once size enters. Third, **narrative concreteness, the deepest single dimension in the measurement stack, survives everything**: its full-specification coefficient of 0.080 ($p = 0.009$) implies an 8.0-point revenue-growth gap between the least and most concrete AI disclosers, holding sector, size, and momentum fixed. It is the only signal to clear the Benjamini–Hochberg threshold

Table 1: Specification ladder: AI signals and year-over-year revenue growth

	(1) Univariate	(2) +Sector	(3) +Size	(4) +Momentum	$n_{(4)}$
Narrative concreteness (ours)	0.129***	0.083***	0.075**	**0.080*** **	395
	(0.000)	(0.010)	(0.018)	(0.009)	
Investment intensity (ours)	0.106***	0.063**	0.048	0.041	383
	(0.001)	(0.043)	(0.115)	(0.141)	
Larridin adoption	0.108***	0.075**	0.051	0.038	429
	(0.000)	(0.013)	(0.106)	(0.201)	
Larridin maturity index	0.095***	0.058**	0.032	0.024	429
	(0.001)	(0.044)	(0.278)	(0.401)	
Larridin impact	0.096***	0.062**	0.042	0.025	429
	(0.001)	(0.028)	(0.138)	(0.346)	
Larridin proficiency	0.066**	0.022	−0.011	−0.011	429
	(0.015)	(0.493)	(0.749)	(0.735)	
AI-hiring builder rate (ours)	0.026	0.021	−0.004	−0.027	212
	(0.400)	(0.505)	(0.892)	(0.165)	

Signals in cross-sectional percentile ranks; coefficients are the outcome difference between lowest- and highest-ranked firms. HC3 robust p-values in parentheses; */**/*** denote $p < 0.10/0.05/0.01$. Spec (2) adds twelve sector fixed effects; (3) adds log market capitalization at the score date; (4) adds pre-signal year-over-year revenue growth. Hiring postdates the outcome quarter and is shown for completeness only.

within the primary family; investment intensity is positive throughout but loses significance once size enters ($p = 0.14$).

The interpretation we place on this pattern is that *specificity is the signal*. What distinguishes subsequently faster-growing firms, within sector and size class, is verifiable substance—named systems in production and quantified results—in their own regulatory disclosures. This is an argument *for* structured measurement, not against composite scoring: surface keyword exposure carries no information (risk-disclosure language, now standardized across firms, discriminates nothing), whereas each step toward evidence-anchored specificity adds signal. The concreteness dimension was built inside the same extraction framework as the Tracker's pillars and can be incorporated into it directly.

5.2 Is it priced? Market and margin outcomes

Applying the full specification to four-month forward returns yields no positive predictability for any signal, and no coefficient in the return family survives multiple-testing correction. Point estimates are consistent with a partial unwind of early-2026 AI-theme valuations over this particular window rather than with any stable relationship. Margin changes show no association with any signal (all $|t| < 1.3$). Taken together with the revenue results, the evidence points to AI adoption operating, at this stage, through the **growth channel**—and to public adoption information being largely **impounded in prices** already, as efficient-markets reasoning would predict for signals built from public sources.

Table 2: Heterogeneity by AI value-chain position

	Concreteness → growth		Mean 4-month return	
Category	coef. (p)	n	raw	n
AI Infrastructure & Hardware	+0.149 (0.305)	22	+44.9%	25
AI Software & Platforms	+0.074 (0.049)	35	−3.8%	38
Data-Rich Adopters	+0.090 (0.198)	101	−3.5%	132
Physical-Asset-Heavy / Late	+0.065 (0.025)	214	+1.8%	278

Left panel: within-category regressions of revenue growth on concreteness rank with size and momentum controls (sector effects omitted within category). Right panel: mean four-month forward return from the score date. The infrastructure-vs-late-adopter return gap is +31.6pp with sector and size controls ($p = 0.005$); the full sample including the five excluded semiconductors gives +34.2pp.

5.3 Where in the value chain? Heterogeneity

Two findings emerge from the value-chain segmentation (Table 2). First, the concreteness–growth association is concentrated where AI credibility is scarcest: it is significant within *Physical-Asset-Heavy / Late Adopters* (0.065, ($p = 0.025$), $n = 214$) and within *AI Software & Platforms* (0.074, $p = 0.049$), while composite scores are insignificant within every category. Among the mass of "ordinary" companies—where AI claims are cheapest to make and hardest to verify—concrete disclosure best separates genuine adopters from laggards.

Second, the market leg of the story is a value-chain story: AI-infrastructure suppliers returned +44.9% on average over four months, versus roughly flat for every other category, and the gap survives sector and size controls at +31.6 percentage points ($p = 0.005$). Because our analysis sample already excludes the five largest AI-semiconductors, this premium is carried by the remaining twenty-five infrastructure names rather than a few marquee stocks (the full sample including the five gives +34.2pp; Table A5). We present it as a descriptive decomposition of the H1-2026 AI rally—the classification is made with knowledge of these firms—rather than as an ex-ante trading result; even so, it shows how narrowly AI's market-value creation concentrated in one segment of the value chain.

5.4 Hiring dynamics

The hiring signal cannot be tested against outcomes that predate it; its results are methodological and descriptive. Beyond the validation figures of Section 3, the June–July panel reveals economically sensible movement: rank-percentile accelerations are led not by technology firms but by insurers, retailers, and financial-data companies staffing data-science and AI-engineering roles (verified against the underlying postings), while several industrial and pipeline firms decelerate. As the monthly panel accumulates, lead–lag tests of the proposal's central hiring hypothesis—whether AI-hiring intensity precedes efficiency gains—become feasible.

5.5 Implications for practitioners

The results carry three direct implications for AI-transformation measurement. *First, weight verifiable specificity.* The predictive content of adoption measurement concentrates in evidence-anchored, deployment-level detail; incorporating the concreteness dimension into composite scores—or reporting it alongside them—should sharpen firm-level discrimination, particularly within peer groups of similar size and sector. *Second, measure where narratives are cheapest.* The association is strongest among physical-asset-heavy late adopters, the segment where AI claims are least verifiable by casual reading and where buyers of measurement (investors, boards, enterprise customers) face the greatest information asymmetry. *Third, the panel is the prize.* Every quarterly score vintage added to a fixed methodology converts cross-sectional associations like ours into progressively stronger designs—lead–lag tests, within-firm changes, and eventually out-of-sample validation. The pipelines delivered with this study (monthly hiring measurement across 536 companies; filings scoring across 478) are built to run on that cadence.

6 Robustness and Limitations

Robustness. The headline concreteness result is stable across winsorization choices, survives sector-neutral rank correlation (IC = +0.15, $p = 0.004$; Appendix Table A1), and passes primary-family FDR control at $q = 0.05$; across the broader fifteen-test matrix it falls just short of the $q = 0.10$ threshold—which it cleared only marginally in the full sample—and we report this plainly. In the strict complete-case subsample—firms with every signal, outcome, and control observed, which halves the sample to 196—the coefficient attenuates but stays positive (0.061 vs. 0.080) with precision reduced accordingly ($p = 0.15$; Appendix Table A3), consistent with a loss of power rather than sensitivity to sample composition. Score-bin sorts show the association is monotone (Appendix Table A2). Five further checks support the result: a placebo signal (company-name length) run through the identical pipeline is null ($p = 0.84$); in a permutation test, the observed concreteness coefficient exceeds all but two of 1,000 label-shuffled re-estimates (permutation $p = 0.002$); the coefficient remains significant when any single sector is excluded (worst case: dropping consumer discretionary, 0.075, $p = 0.029$); it is likewise stable when the five excluded AI-semiconductors are added back—the full 500-company sample gives 0.087, $p = 0.007$ (Table A5); and re-scoring a random 30-company subsample yields 93% exact score agreement (100% within one point; concreteness rerun rank correlation 0.94), indicating the LLM scoring itself is stable. Evidence citation audits, classifier validation against independent labels, and cross-month reliability estimates are reported alongside each signal.

Limitations. Four deserve emphasis. (i) *Single vintage.* One score snapshot precludes multi-quarter backtesting and walk-forward validation; our design identifies cross-sectional associations, not time-series predictability. (ii) *Timing.* The fundamental outcome quarter overlaps the score date by two weeks, and filing dates straddle it; we therefore claim near-term association rather than strict prediction, and we restrict return tests accordingly. (iii) *Disclosure*

bias. Concreteness is a disclosure measure; firms adopting AI quietly are invisible to it, and IR sophistication may correlate with size in ways our controls only partially absorb. (iv) *Reverse causality.* Even with momentum controls, faster-growing firms may write more concrete AI narratives because growth funds deployment; our estimates should be read as conditional associations. Panel accumulation across quarterly score vintages—now feasible with the pipelines built here—is the natural remedy for all four.

7 Conclusion

Using a 500-company cross-section that joins proprietary adoption scores, two newly constructed extension-signal families, and realized outcomes, we find that AI adoption is already visible in the fundamentals of large U.S. companies—and that *how deeply one measures determines how much of it one sees.* Evidence-based adoption scores identify the phenomenon: highly rated firms grow faster. Drilling into the measurement stack, the disclosure-concreteness dimension—concrete, quantified, deployed-use-case reporting—predicts near-term revenue growth through sector, size, and momentum controls, margins show no effect yet, and public signals earn no excess return in a market that appears to have priced them. The value-chain lens sharpens the picture: infrastructure suppliers captured the market value, while among late adopters—the majority of the economy—disclosure concreteness is what separates the firms that are actually growing. The central lesson is an endorsement of structured, evidence-first measurement itself: naive AI-mention counting carries no information, composite evidence-based scores carry the phenomenon, and the deepest evidence-anchored dimension carries signal that survives every control we impose. The study also delivers the tooling to compound these findings—a validated, repeatable measurement pipeline whose quarterly vintages will convert this cross-section into the panel that the field's causal questions require.

Disclaimer. This study is exploratory research produced in an academic–industry collaboration. Nothing herein constitutes investment advice; backtested or hypothetical results do not represent actual trading.

Acknowledgments

We thank Ameya Kanitkar and the Larridin team for data access, API resources, and weekly guidance, and the CMU AI Capstone faculty for supervision.

Appendix: Supplementary Tables

Table A1: Rank information coefficients (Spearman) by signal and outcome

Signal	Revenue growth (YoY)	Margin change (YoY)	4-mo. return	Sector-neutral IC (revenue growth)
Narrative concreteness	+0.222***	+0.019	−0.026	+0.145***
	(397)	(315)	(456)	
Investment intensity	+0.145***	+0.017	−0.033	+0.051
	(385)	(306)	(441)	
Larridin adoption	+0.184***	+0.042	−0.073	+0.118**
	(431)	(342)	(497)	
Larridin maturity index	+0.176***	+0.060	−0.066	+0.104**
	(431)	(342)	(497)	
Larridin impact	+0.173***	+0.057	−0.049	+0.115**
	(431)	(342)	(497)	
Larridin proficiency	+0.118**	+0.043	−0.074	+0.041
	(431)	(342)	(497)	
AI-hiring builder rate	+0.012	+0.048	+0.006	−0.070
	(213)	(174)	(247)	

Spearman rank correlations; observation counts in parentheses. Sector-neutral IC ranks both signal and outcome within sector before correlating. */**/*** denote $p < 0.10/0.05/0.01$.

Table A2: Outcome sorts: revenue growth by signal level

Concreteness score bins			Maturity-index quintiles		
Score	Mean growth	n	Quintile	Mean growth	n
1	−9.4%	2	Q1 (low)	8.5%	87
2	7.3%	100	Q2	8.1%	86
3	9.1%	141	Q3	11.4%	86
4	13.4%	149	Q4	10.0%	86
5	46.2%	5	Q5 (high)	16.2%	86
High (4–5) − low (1–2): +7.5pp ($t = 3.97$)			Q5 − Q1: +7.8pp		

Mean year-over-year revenue growth of the first post-score fiscal quarter, winsorized at 1%/99%. Unconditional sorts (no controls); top-quintile maturity firms grew roughly twice as fast as bottom-quintile firms. Four-month return sorts show no monotone pattern (Q1 +2.7% to Q5 −1.1%).

Table A3: Complete-case robustness: full specification on the strict subsample

Signal	Complete-case coef. (p)	Main-sample coef. (p)
Narrative concreteness	+0.061 (0.148)	+0.080 (0.009)
Investment intensity	−0.014 (0.595)	+0.041 (0.141)
Larridin adoption	+0.038 (0.583)	+0.038 (0.201)
Larridin maturity index	+0.026 (0.666)	+0.024 (0.401)
AI-hiring builder rate	−0.027 (0.224)	−0.027 (0.165)

The complete-case subsample ($n = 196$) requires every signal, the primary outcome, and all controls to be observed. Specification (4) of Table 1: sector fixed effects, log market capitalization, and prior revenue growth, HC3 standard errors. Main-sample column is the analysis sample of Table 1. The concreteness coefficient attenuates modestly while precision falls with the halved sample.

Table A4: Variable coverage in the analysis sample ($n = 513$ listed companies)

Variable	n	Principal reason for missingness
Larridin maturity / adoption	509	four firms unscored in vintage
Narrative concreteness / investment	457 / 442	no U.S. 10-K (foreign filers); no AI content
AI-hiring builder rate	249	fewer than five matched postings
Revenue growth (YoY)	434	non-comparable revenue concepts (banks)
Operating-margin change (YoY)	344	banks and insurers
Four-month forward return	501	delistings
Log market capitalization	503	missing shares data
Prior revenue growth	495	insufficient reported history

Missingness is non-random and documented; each analysis uses all firms with complete data for that analysis, with per-cell counts reported throughout.

Table A5: Full-sample specification ladder including the five AI-semiconductor firms

	(1) Univariate	(2) +Sector	(3) +Size	(4) +Momentum	$n_{(4)}$
Narrative concreteness (ours)	0.147***	0.094***	0.084**	**0.087*** **	399
	(0.000)	(0.005)	(0.011)	(0.007)	
Investment intensity (ours)	0.129***	0.076**	0.057*	0.048*	387
	(0.000)	(0.020)	(0.071)	(0.100)	
Larridin adoption	0.115***	0.075**	0.049	0.037	433
	(0.000)	(0.015)	(0.133)	(0.235)	
Larridin maturity index	0.103***	0.060**	0.032	0.024	433
	(0.000)	(0.042)	(0.295)	(0.427)	
Larridin impact	0.102***	0.063**	0.041	0.023	433
	(0.001)	(0.029)	(0.156)	(0.393)	
Larridin proficiency	0.075***	0.022	−0.013	−0.012	433
	(0.007)	(0.489)	(0.710)	(0.710)	
AI-hiring builder rate (ours)	0.026	0.017	−0.013	−0.033	215
	(0.393)	(0.590)	(0.712)	(0.108)	

Identical specifications to Table 1, estimated on the full sample that *includes* Nvidia, Broadcom, AMD, Micron, and Intel—the five firms excluded from the main analysis. Every qualitative conclusion is unchanged: concreteness is the only signal to survive the full specification (0.087, $p = 0.007$), and the AI-infrastructure four-month return premium in this sample is +34.2pp ($p = 0.003$). HC3 p-values in parentheses; */**/*** denote $p < 0.10/0.05/0.01$.

References

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3–18.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, 57(1), 289–300.

Brynjolfsson, E., Rock, D., and Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1), 333–372.

Fama, E. F., and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607–636.

Jegadeesh, N., and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.