STOCKHOLM SCHOOL OF ECONOMICS

Department of Accounting

BE351 Degree Project in Accounting and Financial Management

Spring 2026

The Gross Margin Illusion

Evidence That the Market Has Yet To Adjust Valuation Multiples for AI Firms Despite Visible Cost-Shifting into R&D

Lukas Karlberg, 25874 & Jarik Foth, 26014

Stockholm School of Economics

Abstract

The valuation of technology firms has traditionally relied on the idea that high gross margins signal scalable business models with low marginal costs. The rise of generative AI challenges this logic, as AI-driven products can involve substantial variable inference costs tied to compute usage. If these costs are classified as R&D rather than COGS, reported gross margins may overstate underlying scalability, creating what this thesis refers to as the “Gross Margin Illusion.”

Using data on U.S.-listed firms from 2018 to 2024, this thesis examines whether markets adjust valuation multiples for AI-exposed firms, defined as constituents of the BOTZ or ROBO thematic ETFs. The analysis first establishes the general relationship between gross margins and valuation in technology firms, then tests whether AI-exposed firms exhibit higher R&D intensity, and finally examines whether investors price their gross margins differently from traditional technology peers.

The findings suggest that markets largely do not distinguish AI-exposed firms from traditional peers when pricing gross margins. While gross margins remain positively associated with valuation multiples and AI-exposed firms report materially higher R&D-to-sales ratios, their gross margins are priced at the same, or slightly higher, intensity as those of other technology firms. This indicates that investors may still be applying a SaaS-era valuation logic to firms whose cost economics have begun to shift, although the data cannot conclusively determine whether this reflects delayed repricing or rational uncertainty.

Keywords: Generative AI, Gross Margin, Valuation, R&D, Accounting Discretion, Value-Relevance, SaaS.

Supervisor: Zeping Pan

Acknowledgements

We would like to thank our supervisor, Zeping Pan, for his guidance and support throughout this process. His feedback has consistently challenged us to think more critically and helped sharpen both our ideas and the overall direction of the thesis.

We are also grateful to our peers for the many discussions, inputs, and perspectives shared along the way. Those conversations have played an important role in developing this work and made the process more engaging.

Table of contents

1. Introduction

Over the past two decades, the valuation of technology firms has been anchored in a simple and powerful intuition: high gross margins signal scalable business models. In traditional Software-as-a-Service (SaaS) companies, the marginal cost of serving an additional customer is close to zero, meaning that incremental revenue largely translates into profit. As a result, gross margin has become one of the most important metrics used by investors to assess operating leverage and justify high valuation multiples.

The rapid emergence of generative artificial intelligence challenges this foundational assumption. Unlike traditional software, AI-driven products incur substantial variable costs through "inference", with each new user consuming linearly more computational resources. With non-zero marginal costs, growth is not free for AI companies, so current valuation standards developed based on SaaS business mechanics might misrepresent the valuation of AI companies and misinform investor decisions.

As an underlying mechanism behind these potential misled valuations we investigate current accounting standards which allow firms considerable discretion in how their costs are classified. In practice, big parts of inference expenses can be recorded as Research and Development (R&D) rather than Cost of Goods Sold (COGS), thereby preserving high reported gross margins even when underlying unit economics have changed industry-wide.

This mechanism can create a potential disconnect between reported accounting metrics and economic reality, a phenomenon this thesis refers to as the *Gross Margin Illusion*. If gross margins no longer accurately reflect scalability, the central question becomes whether financial markets recognize this distortion. Two opposing outcomes are plausible. On the one hand, investors may see through the accounting treatment and adjust valuation multiples accordingly, discounting firms whose margins are supported by discretionary classification. On the other hand, markets may continue to rely on reported gross margins as a key signal, leading to inflated valuations and potential misallocation of capital.

This thesis investigates which of these scenarios better describes current market behavior. Specifically, we first examine whether AI-exposed firms actually exhibit systematically different cost structures, and whether the market prices their gross margins differently compared to traditional technology firms. To do so, we employ a value-relevance framework and analyze a panel of U.S. publicly listed firms over the period 2018–2024. The empirical strategy proceeds in three steps: first, establishing the importance of gross margins for valuation in the broader technology sector; second, testing whether AI firms display higher R&D intensity consistent with cost shifting or structurally different cost bases; and third, evaluating whether investors apply a discount or premium to the gross margins of AI firms.

By linking accounting classification, firm-level cost structures, and market valuation, this study contributes to three strands of literature. First, it extends the value-relevance literature by examining whether the relationship between a key accounting metric and firm value shifts in response to technological change. Second, it adds to research on accounting discretion by highlighting how classification choices can materially affect widely used performance indicators. Third, it contributes to the emerging literature on the economics of AI by providing empirical evidence on how markets interpret and potentially misinterpret the financial signals of inference-driven business models.

2. Research Problem and Question

This thesis answers two related empirical questions. The first concerns the cost side and serves as an empirical foundation; the second concerns the market side and serves as core north star for this thesis:

1. Do AI-exposed firms differ from traditional technology firms in how they classify expenses, in particular in the share recorded as research and development rather than as cost of goods sold?

2. Do investors price the gross margins of AI firms differently from those of traditional technology firms to account for the different cost structure?

Together, these questions ask whether the accounting signals investors have used to value technology firms for decades, still track the changing economic reality.

Primary Research Question: Do AI-exposed firms exhibit systematically different cost structures from traditional technology firms, and does the market price their gross margins differently?

3. Institutional Background

3.1 GAAP Discretion on R&D vs Cost of Revenue

The premise that AI firms can plausibly classify variable inference costs as research and development rather than as COGS rests on the structure of US GAAP itself. ASC 730–10 ("Research and Development") defines research as "planned search or critical investigation aimed at discovery of new knowledge with the hope that such knowledge will be useful in developing a new product or service ... or a new process or technique" (FASB, ASC 730–10–20). The standard mandates that research and development costs be charged to expense when incurred (730–10–25–1) and provides an enumerated list of qualifying activities at 730–10–55–1. Critically, however, the boundary between R&D and other expense lines is subject to managerial interpretation: the standard does not enumerate which inputs (compute, cloud, data acquisition) flow through R&D rather than through cost of goods sold, leaving classification to management discretion subject to audit review.

Empirical evidence confirms that this discretion is exercised. Koh and Reeb (2015) document that 10.5% of firms reporting no R&D nonetheless hold patents, approximately fourteen times the rate of true zero-R&D firms and that these firms begin reporting R&D following exogenous auditor changes. The R&D line, in their analysis, is therefore a reporting choice rather than a mechanical readout of innovation activity. Laplante, Skaife, Swenson and Wangerin (2019) extend this finding into the strategic-classification setting: a one-standard-deviation increase in strategic R&D classification reduces the GAAP effective tax rate by approximately 1.7%. Together these findings show that classification shifts are economically material and that firms have clear incentive to relabel costs as R&D.

The historical analogue is ASC 985–20 ("Costs of Software to Be Sold, Leased, or Otherwise Marketed"), originally issued as Statement of Financial Accounting Standards No. 86 in August 1985 in response to the personal-computer software boom. ASC 985–20–25–1 specifies that "all costs incurred to establish the technological feasibility of a computer software product to be sold, leased, or otherwise marketed are research and development costs". A dedicated standard created precisely because the existing R&D guidance was ambiguous about how to treat the costs of an emerging software industry. No analogous standard yet exists for AI compute. The 2025 SEC Investor Advisory Committee AI Disclosure Recommendations and the 2025 FASB Accounting Standards Update on internal-use software costs (amending ASC 350–40) confirm that regulators are active in adjacent areas; neither directly addresses the COGS-versus-R&D classification of AI inference. The absence is itself part of the institutional setting this thesis examines.

3.2 SaaS Operating Leverage and the Gross Margin Premium

Public-market data on software firms make tangible the magnitude of what this thesis is contesting. Meritech Capital's tracking of approximately 100 listed SaaS companies reports a current median forward-revenue multiple of roughly 6.5×, down from a 2021 peak of 20.8×, with 90th-percentile firms still trading near 12× after compression from a 51.9× peak (Meritech Capital, 2024). The constituents of these indices typically report gross margins in the 70–85% range, well above the broader-market norm.

The size of this premium is not anchored in current earnings but in the operating leverage that high gross margins are read to imply. Amir and Lev (1996), examining the wireless-communications industry, showed that for high-growth, intangible-intensive firms, traditional earnings and book-value measures had limited explanatory power for stock prices, while non-financial operating indicators captured the scalability premium investors actually paid for. Rajgopal, Venkatachalam and Kotha (2003) extended the result to internet firms, demonstrating that scalability-related non-financial indicators provide substantial incremental explanatory power beyond accounting fundamentals. That is the empirical template the market would later apply to public SaaS firms via revenue multiples conditional on gross margin.

The McKinsey Rule-of-40 study sharpens the same pattern at the firm level: across forty public B2B SaaS firms, top-quartile retention cohorts trade at a median 21× forward revenue versus 9× for non-compliant peers, evidence that the market does not reward growth or margin in isolation but explicitly prices the combination of the two (Di Mattia et al., 2021). Gross margin operates not merely as a profitability ratio in this universe but as the value-relevant signal of operating-leverage capacity itself.

3.3 The Economics of Generative AI Inference

The most direct evidence that AI inference behaves as a variable cost comes from how rapidly its price moves. SemiAnalysis reports that the cost-per-token of GPT-3-equivalent quality has fallen approximately 1,200-fold since 2020, with providers competing along a frontier on which workloads are served near cost to gain market share (Patel et al., 2025). A line item that moves at compute-frontier speed cannot be a fixed input: each query against a large language model consumes compute capacity, GPU time, electricity, and datacenter cooling that scales directly with usage. Inference is, in that respect, a genuine variable cost of serving customers, and the practitioner literature documents both the magnitude and the volatility of this cost in detail.

Firm-specific evidence corroborates the institutional pattern. The Information reported in January 2026 that Anthropic cut its 2025 gross margin guidance to roughly 38%–40% including free-tier inference, a reduction of approximately 10 percentage points, attributing the cut to inference costs running 23% above projection on Google and Amazon servers (Efrati and Woo, 2026). SaaStr, citing The Information, reports that OpenAI's compute margin rose from approximately 35% in early 2024 to 70% by October 2025, but headline 2025 gross margin including free-tier inference is approximately 46%, implying that compute is on the order of half of revenue (Lemkin, 2025). Cross-sectional benchmarks show the same pattern: Jaipuria (2025) reports model-provider gross margins of 50–60% and an early-AI-application cohort averaging 25%, against the 70–80% SaaS norm. Andreessen Horowitz documents that AI firms "spend more than 80% of their total capital raised on compute resources" and that compute demand outstrips supply by a factor of ten (Appenzeller, Bornstein and Casado, 2023).

Microsoft's fiscal-year 2025 Form 10–K provides a natural experiment in classification. Microsoft, uniquely positioned to allocate compute either to internal R&D or to cost of revenue, reports that "gross margin percentage decreased, driven by the impact of scaling our AI infrastructure, booking AI compute through cost of revenue and accepting the margin compression that follows" (Microsoft Corporation, 2025). The empirical fact that a firm with the operational discretion to do otherwise classifies AI compute as cost of revenue demonstrates that the alternative, classification as R&D, is a reporting choice rather than an economic necessity. This sets up the central question: do firms whose products are wholly AI-driven exercise the same classification, and if not, what does the divergence imply for their reported gross margin signals?

4. Literature Review

4.1 The Value-Relevance Framework

The empirical strategy of this thesis sits within the value-relevance tradition in financial accounting research. Ohlson (1995) provides the theoretical scaffold: under clean-surplus accounting, equity value can be expressed as the sum of book value and the present value of expected abnormal earnings, formalizing accounting numbers as sufficient statistics for equity value. The Ohlson model establishes the warrant for regressing market-based valuation metrics on accounting inputs, an approach that has generated a large empirical literature.

Barth, Beaver and Landsman (2001) codify the methodology of this literature, framing the value-relevance research design as a legitimate empirical test of "the association between accounting amounts and equity market values," with a significant association interpreted as evidence that the accounting amount captures information used by investors in pricing shares. Their survey remains the canonical methodological reference.

Two refinements to the value-relevance literature are particularly relevant to the current setting. Collins, Maydew and Weiss (1997) show that the combined value-relevance of earnings and book values has not declined over time but that the relative weight has shifted from earnings toward book values, with intangible-intensity emerging as a key driver of the shift. The implication is that value-relevance coefficients are not structural constants, they vary as the underlying business-model composition of the economy changes. Lev and Zarowin (1999) push the same point further: the informativeness of reported financial statements has deteriorated in part because of the mismatching of intangible investments against revenues, and financial statements increasingly fail to capture the economics of innovative firms. Both papers establish precedent for the present thesis's central concern: that a specific value-relevance mapping (gross margin to enterprise value in tech firms) may be deteriorating because a technological shift has altered the relationship between the accounting line item and the underlying economics.

4.2 Gross Margin as a Scalability Signal in Technology Firms

The closest line-item evidence that gross margin specifically, rather than earnings or book value, carries the valuation signal in technology firms comes from the dot-com period. Trueman, Wong and Zhang (2000), examining a sample of business-to-consumer Internet firms, document that conventional bottom-line earnings were negatively associated with stock prices while gross profits were positively and significantly associated. The line-item asymmetry is precisely what a scalability-driven valuation regime would predict: at the top of the income statement, the cost-structure signal dominates; below it, expensed investment in growth depresses earnings without depressing the multiple. Novy-Marx (2013) extends the case from accounting to asset pricing, showing that gross profitability is a robust cross-sectional predictor of returns with explanatory power comparable to book-to-market, and characterising it as the cleanest accounting measure of true economic profitability precisely because it sits above the discretionary expense lines (R&D, SG&A) where classification choice operates.

That protective property is precisely what generative AI puts at risk: when inference costs flow through R&D rather than cost of revenue, the gross margin signal that value-relevance research identifies as the cleanest in technology firms becomes the one most exposed to classification distortion, the central concern this thesis interrogates.

The earlier wave of value-relevance work in technology established the same pattern at the level of operating proxies. Amir and Lev (1996), examining the wireless-communications industry, showed that for high-growth, intangible-intensive firms, traditional earnings and book-value measures had limited explanatory power for stock prices, while non-financial operating indicators captured the scalability premium investors actually paid for. Rajgopal, Venkatachalam and Kotha (2003) extended the result to internet firms, demonstrating that scalability-related non-financial indicators provide substantial incremental explanatory power beyond accounting fundamentals. Together these papers warrant the empirical framing under which Hypothesis 1 is formulated: in technology firms, where intangible intensity and operating leverage are at their highest, gross margin is the line item that should carry the valuation signal and the line item this thesis tests for whether that mapping continues to hold once the underlying cost structure has shifted.

4.3 Accounting Choice and R&D Discretion

The hypothesis that AI firms exhibit elevated R&D intensity (H2) draws on a long literature documenting that R&D classification under US GAAP is discretionary and that discretion has measurable consequences for both reported financials and equity prices. Lev and Sougiannis (1996) provide the foundational evidence that R&D is value-relevant: their estimated firm-specific R&D capital is significantly associated with both contemporaneous stock prices and subsequent returns, establishing that investors reward the R&D label and that managers therefore have economic incentive to route costs through it.

Koh and Reeb (2015) document that the recognition margin itself is porous: 10.5% of firms reporting no R&D in a given year nonetheless hold patents, approximately fourteen times the rate of genuinely non-innovative firms, and these "missing R&D" firms begin disclosing R&D following exogenous auditor changes. The R&D line item, in their analysis, is a reporting choice rather than a mechanical readout of innovation activity.

Most directly relevant to this thesis, Poonawala and Nagar (2019) document that managers actively engage in classification shifting between cost of goods sold and research and development. Examining the specific channel by which gross margin can be manipulated, they find that managers "are more likely to shift costs of goods sold to research and development expenses in order to meet or beat prior period's gross margin," attributing the R&D preference to "the relatively vague nature" of the line item and to investors' favorable view of R&D relative to selling, general and administrative expenses. Their evidence is the closest published precedent for the exact mechanism inferred from the present thesis's H2 finding: an 8.75-percentage-point gap in R&D-to-sales between AI firms and traditional tech-adjacent peers. Whether the divergence reflects substantive R&D investment, classification shifting, or some combination cannot be resolved without firm-level cost-allocation data not available in public filings, but the existence of a documented mechanism for COGS-to-R&D shifting is established literature.

4.4 AI-Specific Firm Economics

The economics of firms exposed to artificial intelligence is the subject of an emerging literature whose principal empirical contributions postdate the November 2022 release of ChatGPT. Eisfeldt, Schubert and Zhang (2023) construct firm-level workforce-exposure measures to generative AI and document a discrete repricing event: an Artificial-Minus-Human portfolio earned approximately 5% in the two weeks following ChatGPT's release, with the effect strongest among firms holding substantial data assets. The paper does two things this thesis builds on directly: it establishes the pre/post-ChatGPT empirical design that the present robustness analysis (§9.4) adopts, and it documents that the market can reprice AI exposure quickly when the channel is salient, sharpening the puzzle the present thesis investigates of whether the market has equally repriced the gross margin signal, where the channel is less visible.

Babina, Fedyk, He and Hodson (2024) examine the channels through which AI investment translates to firm value. Their published Journal of Financial Economics analysis finds that AI-investing firms exhibit higher growth in sales, employment, and market valuations, but that the gains flow primarily through product innovation rather than through unit-cost reduction. A companion working paper (Babina, Fedyk, He and Hodson, 2024b) shows that AI investment raises firms' systematic risk, consistent with the market capitalizing AI as a growth option embedded in equity values. The two findings together leave the cost-structure leg of the AI valuation question explicitly open: if the market capitalizes AI firms as growth options without pricing the unit economics of their product delivery, the gross margin signal, which the market ordinarily uses to assess scalability, could be carrying disproportionate weight precisely when its informational content is weakest.

Cao, Liu, Qiu and Zhao (2024) examine the informativeness of voluntary AI disclosure in corporate filings. AI disclosures in annual reports rose sharply following ChatGPT's release; the disclosures correlate positively with AI-related hiring and patents and predict future operational efficiency. Critically for the present thesis, however, firms whose high disclosures do not align with substantial AI employment underperform peers in long-run stock returns, evidence that markets only partially separate substantive AI engagement from narrative. This finding directly supports the interpretation developed in §9.3 that the market has incompletely priced the accounting-economics divergence in AI firms and offers the academic precedent for reading H3's positive interaction as evidence of incomplete repricing rather than as market failure.

Ante and Saggu (2025) provide methodological validation for the AI-firm identification strategy used in this thesis. Applying natural language processing to 10-K filings of 3,395 NASDAQ-listed firms from 2010 to 2022, they construct AI-engagement scores and demonstrate that the resulting indices match or exceed fourteen commercial AI-themed exchange-traded funds in risk-return profile and event-study responsiveness. Their finding licenses both ETF-membership identification (the present thesis's main specification, drawing on BOTZ and ROBO) and 10-K keyword identification (a robustness check available for future work) as substantively equivalent approaches to demarcating the AI-firm population.

4.5 Inference Cost as a Variable Cost of Software

The peer-reviewed literature on inference cost as a variable cost of software is nascent. Section 3.3 documents the practitioner's evidence on the scale and volatility of this cost. The research gap that section identifies, the absence of academic work directly testing the value-relevance implications of variable AI inference cost, is the contribution this thesis addresses. Where Eisfeldt, Schubert and Zhang (2023) and Babina, Fedyk, He and Hodson (2024) establish that AI exposure is repriced and that AI investment drives growth, neither examines whether the market has adjusted the accounting-to-economics mapping that traditionally underwrote technology-firm valuation. That question is the subject of the empirical analysis below.

4.6 Market Efficiency and the Incorporation of Accounting Information

The interpretation of the present thesis's H3 finding, that the market has not fully repriced the accounting-economics divergence in AI firms, is consistent with a long literature documenting that public accounting information is incorporated into prices with a lag rather than instantly. Bernard and Thomas (1989) provide the canonical demonstration in their analysis of post-earnings-announcement drift: stock prices adjust to earnings surprises over months rather than days, reflecting investors' failure "to fully recognize the implications of current earnings for future earnings. Sloan (1996) extends the pattern to the structure of accounting numbers themselves, showing that investors appear to fixate on reported earnings while failing to discount the lower persistence of the accrual component, a multi-year delay in pricing a specific accounting distortion.

Two papers specialize this result to the R&D and intangibles channel directly relevant to the present thesis. Chan, Lakonishok and Sougiannis (2001) document that high-R&D-to-market-value firms earn substantial excess returns, evidence that the market is overly pessimistic about R&D-intensive firms because expensed intangible investment is not transparent on the financial statements. Eberhart, Maxwell and Siddique (2004) show that firms with unexpectedly large R&D increases earn significantly positive long-horizon abnormal returns, a pattern they characterize as "consistent with investor underreaction" to R&D-expensed economics. Together, these papers establish that delayed market adjustment to accounting information is documented, that the delay is particularly pronounced in R&D-intensive settings, and that the present thesis's finding fits an established empirical pattern rather than violating market-efficiency norms.

5. Hypotheses

Drawing on Value-Relevance Theory (Ohlson, 1995), we hypothesize that investors use gross margin as a primary signal for future scalability and operating leverage. The three hypotheses are sequential: H1 establishes the baseline premium that H2 and H3 then probe for AI firms specifically.

H1 (The Tech Margin Premium): Gross Margin is positively associated with Enterprise Value-to-Revenue (EV/Sales) multiples, and this effect is significantly stronger for Technology firms than for non-Technology firms.

H2 (The R&D Smoking Gun): AI-exposed technology firms report significantly higher R&D-to-Sales ratios than traditional technology firms.

H3 (The Repricing Hypothesis, two-sided null): The market valuation of gross margins does not differ significantly between AI firms and traditional technology firms.

H3 is operationalised as the GM x I_AI interaction in Model 3. A negative coefficient would indicate that the market discounts gross margins for AI firms, while a coefficient close to zero would suggest that the market does not differentiate between AI and non-AI firms in pricing gross margins. A positive coefficient would suggest that the market applies a higher valuation premium to AI firms' gross margins.

6. Methodology and Research Design

6.1 Operationalization of Key Variables

To ensure exact replicability, all financial variables are constructed using standard Compustat North America Fundamentals Annual data items, merged with CRSP market data via the WRDS CCM linking table.

Dependent Variable (EV/Sales): Enterprise Value is calculated as Market Capitalization (CRSP month-end price × shares outstanding, evaluated at the fiscal year-end) plus Total Long-Term Debt (DLTT) plus Debt in Current Liabilities (DLC) less Cash and Short-Term Investments (CHE), scaled by Total Revenue (SALE).

Independent Variable 1 (Gross Margin): $(SALE-COGS)/SALE$

Independent Variable 2: (R&D Intensity), *calculated as R&D expense divided by sales:* $XRD/SALES$

Independent Variable 3 (I_{AI}): a binary variable set to 1 if the firm's ticker is a constituent of the Global X Robotics & Artificial Intelligence ETF (BOTZ) or the Robo Global Robotics & Automation Index ETF (ROBO), and 0 otherwise. A third thematic fund, the Amplify AI Powered Equity ETF (AIEQ), was explicitly excluded from the treatment group: AIEQ constituents are selected by an algorithmic stock-picking model rather than on the basis of the firm's exposure to AI as a business. Including AIEQ would contaminate the treatment group with broad-market firms (e.g., Apple, AbbVie, AIG) that are not AI-exposed in an economic sense.

Controls: Firm Size is the natural log of Total Assets (AT); Leverage is (DLTT + DLC) / AT; Sales Growth is the one-year percentage change in SALE.

6.2 Empirical Strategy

Firm fixed effects are not used in the main specifications. The treatment indicator I_AI is constructed from current ETF membership applied statically across all sample years and is therefore time-invariant within each firm. Firm fixed effects would be perfectly collinear with I_AI, leaving the coefficient unidentifiable rather than merely attenuated; the same constraint binds Model 2, where I_AI is the regressor of interest. We instead use industry (GICS sector) and year fixed effects, which absorb systematic differences across sectors and over time while preserving the cross-firm variation on which identification rests. Standard errors are clustered at the firm (gvkey) level throughout to allow for arbitrary serial correlation within firms.

Our sequential design utilizes three models:

Model 1 (Testing H1 – Tech vs. Rest, full sample):

$\text{EV/Sales}_{i\square} = \alpha + \beta_1 \cdot \text{GM}_{i\square} + \beta_2 \cdot \text{Tech\_Dummy} + \beta_3 \cdot (\text{GM}_{i\square} \times \text{Tech\_Dummy}) + \text{Controls} + \varepsilon_{i\square}$

Model 2 (Testing H2 – the R&D Shift; restricted to the Tech-adjacent sample):

$\text{R\&D/Sales}_{i\square} = \alpha + \beta_1 \cdot \text{I\_AI}_i + \text{Controls} + \varepsilon_{i\square}$

Model 3 (Testing H3 – the Repricing Hypothesis; restricted to the Tech-adjacent sample):

$\text{EV/Sales}_{i\square} = \alpha + \beta_1 \cdot \text{GM}_{i\square} + \beta_2 \cdot \text{I\_AI}_i + \beta_3 \cdot (\text{GM}_{i\square} \times \text{I\_AI}_i) + \text{Controls} + \varepsilon_{i\square}$

6.3 Broadened "Tech-adjacent" Universe for Models 2 and 3

Of the 47 US-listed firms in our AI treatment group (BOTZ ∪ ROBO), only 22 are classified in GICS sector 45 (Information Technology). The remaining 25, including Tesla, Alphabet, Intuitive Surgical, Deere, Emerson, and Illumina are classified in Consumer Discretionary, Healthcare, Industrials, or Communication Services. GICS classifies firms by end-product sector, not by underlying cost structure. The economic construct relevant to this thesis is AI-business exposure, which is not captured by sector classification; Tesla's compute cost for autonomous-driving inference is not less real because S&P labels Tesla "Consumer Discretionary".

We therefore define the Model 2 and Model 3 universe as the Tech-adjacent universe, the union of GICS 45 firms and firms identified as AI-exposed via thematic ETF membership. Within this universe, the control group is GICS 45 firms not in our AI treatment set. A GICS-45-strict robustness specification is reported alongside to show the main findings do not hinge on the broader universe.

7. Data Collection, Sample Construction, and Attrition

Because the thesis tests whether AI exposure changes the valuation relevance of gross margins, sample construction is central to the empirical design. The main challenge is that AI exposure is not directly observable in Compustat or CRSP. We therefore construct an AI treatment group externally, using thematic ETF membership, and compare these firms to a broader technology-adjacent universe. This approach allows us to identify firms whose business models are plausibly exposed to AI-related cost structures, while maintaining a sufficiently large control group for regression analysis.

The use of ETF membership is not intended to perfectly measure the intensity of AI usage within each firm. Rather, it provides a transparent and replicable proxy for market-recognized AI exposure. This is important because the thesis is concerned not only with underlying cost structures, but also with whether publicly visible AI exposure is priced differently by investors.

The sample period spans from 2018 to 2024, chosen to capture the mature cloud-computing era and the post–2022 Generative AI boom. We obtain financials from Compustat North America Fundamentals Annual and match to fiscal-year-end market capitalization via the CRSP Monthly Stock File, linked by the WRDS CCM ccmxpf_lnkhist linking table (LC and LU link types; primary links prioritized).

7.1 AI Treatment Group Identification

We cross-referenced current constituent lists of the BOTZ and ROBO ETFs. After removing non-US primary listings, cash/currency/index rows, and non-equity placeholders, the procedure isolates 47 unique US-listed firms across the 2018–2024 window, a static union of current holdings applied to all sample years. This produces 212 firm-year observations for the I_AI = 1 treatment group after all filters. AIEQ is retained on disk for audit but excluded from analysis (rationale above). Ante and Saggu (2025) demonstrate that ETF-based AI identification of this kind produces samples comparable in risk-return profile to 10-K-keyword identification, supporting the robustness of the treatment-group construction.

7.2 Sample Cleaning and Attrition

Starting from 92,566 raw firm-year observations, filters are applied sequentially and each drop is recorded in the pipeline's attrition log:

• Drop firm-years missing core fields (SALE, COGS, AT): –36,553 → 56,013
• Drop firm-years with SALE < $50M: –21,325 → 34,688
• Drop firm-years with missing/invalid EV/Sales: –8,103 → 26,585
• Drop firm-years missing lagged Sales Growth (first-year observations): −7,152 → 19,433
• Drop firm-years missing any control: 0 → 19,433 final analysis sample

All continuous variables are winsorized at the 1st and 99th percentiles, computed separately within each regression sample. Missing R&D (XRD) is replaced with zero; this assumption is defensible at the $50M revenue floor because R&D non-reporting at this scale typically indicates immaterial R&D expenditure.

7.3 Sample Sizes

Model 1 (full sample): 19,432 firm-year observations across 4,817 unique firms and seven fiscal years.

Models 2 and 3 (Tech-adjacent universe: GICS 45 ∪ AI firms): 2,958 firm-years across 670 unique firms. Of these, 212 firm-years (47 unique firms) are in the I_AI = 1 treatment group.

GICS-45-strict robustness universe: 2,851 firm-years across 649 firms.

8. Descriptive Statistics

Table 1 reports summary statistics across three subsamples: non-Technology firms, traditional Tech firms (GICS 45, non-AI), and AI firms (I_AI = 1). All variables are computed post-winsorization.

Table 1. Summary Statistics by Sample

Variable	Non-Tech	Traditional Tech (GICS 45, non-AI)	AI firms (I_AI = 1)
n (firm-years)	16,475	2,746	212
EV / Sales	–	4.44	6.21
Gross Margin	–	0.535	0.560
R&D / Sales	–	0.154	0.185
Size (ln AT)	–	7.16	8.65
Leverage	–	0.241	0.220
Sales Growth	–	0.156	0.170

This table reports mean values by sample group. Variables are winsorized at the 1st and 99th percentiles within each regression sample.

Two patterns are visible before any regression. First, AI firms report R&D-to-Sales 3.1 percentage points higher than traditional Tech firms (18.5% vs. 15.4%), consistent with H2. Second, AI firms trade at an EV/Sales multiple roughly 40% higher than traditional Tech (6.21 vs. 4.44), suggesting that the market places substantial premium on AI exposure. Whether this premium is mediated by the gross margin signal or independent of it, is the question Model 3 addresses.

These descriptive differences are important because they already point to the central tension of the thesis. AI firms report slightly higher gross margins than traditional technology firms, but also materially higher R&D intensity. If R&D partly contains costs that are economically closer to variable production costs, then reported gross margin may overstate the scalability of AI-exposed firms. At the same time, AI firms trade at higher EV/Sales multiples, suggesting that the market does not mechanically penalize them for this higher R&D burden. The regressions below test whether this pattern remains after controlling for firm size, leverage, sales growth, industry, and year effects.

9. Results

This section presents the empirical results in the same sequential order as the hypotheses. Model 1 first establishes whether gross margin is value-relevant in technology firms relative to the broader market. Model 2 then tests whether AI-exposed firms differ from traditional technology firms in their reported R&D intensity. Finally, Model 3 examines whether the market prices gross margins differently for AI-exposed firms. This structure is important because the interpretation of the final interaction term depends on the first two results: gross margin must first be shown to matter for valuation, and AI firms must then be shown to differ in cost structure, before it becomes meaningful to ask whether investors adjust how they price AI firms' gross margins.

Table 2 reports the main regression results for all three models. All specifications include industry (GICS sector) and year fixed effects and report standard errors clustered at the firm level.

Table 2. Main Regression Results

Variable	**(1) EV/Sales – Full sample**	**(2) R&D/Sales – Tech-adjacent**	**(3) EV/Sales – Tech-adjacent**
Gross Margin	1.487*** (0.312)	–	10.108*** (0.609)
Tech_Dummy	–1.347*** (0.173)	–	–
GM × Tech_Dummy	9.034*** (0.729)	–	–
I_AI	–	0.088*** (0.030)	–1.530 (1.757)
GM × I_AI	–	–	4.521* (2.706)
Size	0.175*** (0.025)	–0.005 (0.003)	0.376*** (0.076)
Leverage	2.026*** (0.254)	–0.041 (0.022)	0.539 (0.714)
Sales Growth	1.307*** (0.106)	0.036** (0.014)	3.903*** (0.529)
Constant	–7.872*** (0.482)	0.036 (0.040)	–12.898*** (1.351)
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes
N	19,432	2,958	2,958
R^2	0.232	0.043	0.352

*significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$*

9.1 H1 – The Tech Margin Premium (Model 1)

The interaction term GM × Tech_Dummy is positive and highly significant ($\beta = 9.034$, $p < 0.001$). This means that the association between gross margin and EV/Sales is substantially stronger for technology firms than for non-technology firms. For non-technology firms, a one-unit increase in gross margin is associated with a 1.487-unit increase in EV/Sales. For technology firms, the corresponding association is 10.521, calculated as the sum of the gross-margin coefficient and the interaction term, 1.487 + 9.034. The market therefore appears to pay approximately seven times more for a given increase in gross margin in technology firms than in the broader economy.

This finding supports H1 and confirms that gross margin is not treated as a neutral accounting ratio in technology valuation. Instead, it appears to function as a scalability signal. In technology firms, a higher gross margin suggests that incremental revenue can be converted into future operating profit at a relatively high rate, which helps explain why investors attach higher valuation multiples to such firms. Establishing this baseline is important because it confirms that there is a gross-margin premium worth testing in the AI context.

9.2 H2 – The R&D Smoking Gun (Model 2)

Within the tech-adjacent sample, the coefficient on I_AI is positive and statistically significant ($\beta = 0.088$, $p = 0.004$). This indicates that AI-exposed firms report R&D-to-Sales ratios approximately 8.8 percentage points higher than non-AI tech-adjacent firms, after controlling for firm size, leverage, sales growth, industry, and year effects. H2 is therefore supported.

Economically, the magnitude is meaningful. The result suggests that AI-exposed firms differ from traditional technology firms not only in how the market describes them, but also in how their cost structures appear in reported accounting data. Since R&D is located below gross profit in the income statement, higher R&D intensity can preserve reported gross margins even if some of the underlying costs are economically linked to product delivery.

Importantly, this result should not be interpreted as direct proof that AI firms misclassify costs. Higher R&D intensity may reflect genuine innovation investment, classification discretion, or a combination of both. However, in the context of the accounting literature on COGS-to-R&D

shifting, the result is consistent with the mechanism proposed in this thesis: AI-exposed firms carry a materially different cost structure, and part of that difference may be reflected below gross profit rather than in COGS.

9.3 H3 – The Repricing Hypothesis (Model 3)

The central interaction term, GM × I_AI, is positive and marginally significant ($\beta = 4.521$, $p = 0.095$). Since H3 is formulated as a two-sided null hypothesis, this result provides weak evidence against the null at the 10% level, but not at the 5% or 1% levels. Economically, the coefficient suggests that the association between gross margin and EV/Sales is higher for AI-exposed firms than for non-AI tech-adjacent firms. Specifically, the estimated gross-margin coefficient is 10.108 for non-AI firms and 14.629 for AI-exposed firms, calculated as 10.108 + 4.521.

This result does not support the valuation-penalty interpretation. If investors were discounting AI firms' gross margins because they suspected that reported margins overstate scalability, the expected sign of the interaction would be negative. Instead, the coefficient is positive. The point estimate therefore suggests that investors continue to reward reported gross margins in AI-exposed firms, despite the higher R&D intensity documented in Model 2.

However, the statistical weakness of the result is important. The coefficient is only marginally significant, meaning that the evidence should not be interpreted as strong proof that AI firms receive an additional gross-margin premium. A more cautious interpretation is that the results provide no evidence of a market penalty and weak evidence of continued reliance on gross margin as a valuation signal for AI-exposed firms. Taken together with the H2 result, this supports the idea that the market has not fully adjusted how it interprets gross margins in firms with AI-related cost structures.

9.4 Robustness

Table 3 reports the coefficient on GM × I_AI (or GM × I_AI_BOTZ in specification (b)) across four alternative specifications.

Table 3. Robustness Checks on the Interaction Coefficient

Specification	β on GM × I_AI	Std. Err.	p-value	N
Main spec (Tech-adjacent, BOTZ ∪ ROBO)	+4.521*	(2.706)	0.095	2,958
(a) BOTZ-only (strict AI definition)	+6.666	(4.089)	0.103	2,958
(b) Pre-ChatGPT subsample (2018–2021)	+8.399**	(3.960)	0.034	1,446
(c) Post-ChatGPT subsample (2022–2024)	+3.794	(2.705)	0.161	1,512
(d) GICS 45 strict universe	+0.218	(3.784)	0.954	2,851

*significance levels: *** p <0.01, ** p <0.05, * p <0.10*

The interaction is positive across all specifications, indicating that none of the alternative specifications produce the negative interaction predicted by a valuation-penalty interpretation. The BOTZ-only specification, which applies a stricter AI definition, produces an even larger positive coefficient ($\beta = 6.67$), although it narrowly misses conventional significance ($p = 0.103$). This suggests that the positive direction is not solely an artefact of the broader AI classification.

The most informative result is the pre/post-ChatGPT split. In the pre-ChatGPT period (2018–2021), the interaction is large and statistically significant ($\beta = 8.40$, $p = 0.034$), implying a substantial additional gross margin valuation premium for AI firms. In the post-ChatGPT period (2022–2024), the coefficient remains positive but falls to 3.79 and is no longer statistically significant at conventional levels. The point estimates therefore suggest attenuation rather than a reversal into a valuation penalty. This pattern is consistent with gradual, still-incomplete market repricing rather than a full correction, in line with the literature on delayed market adjustment to accounting information (Bernard and Thomas, 1989; Sloan, 1996; Chan, Lakonishok and Sougiannis, 2001; Eberhart, Maxwell and Siddique, 2004).

In the GICS-45-strict universe, the interaction is statistically indistinguishable from zero ($\beta = 0.218$, $p = 0.954$). The sharp decline relative to the main specification suggests that the positive interaction is partly related to the inclusion of AI firms outside the narrow GICS 45 classification. This supports the relevance of the broadened tech-adjacent universe, as the AI-related gross margin premium appears most visible among firms not captured by the strict technology-sector definition.

Overall, the robustness tests support a cautious interpretation of the main result. None of the alternative specifications produces the negative interaction that would be expected if investors clearly penalized AI firms' gross margins. The BOTZ-only specification suggests that the positive direction is not simply driven by an overly broad AI definition. The pre/post-ChatGPT split suggests that the AI gross margin premium has weakened after generative AI and its cost implications became more visible, but the coefficient does not reverse into a penalty. Finally, the GICS-45-strict specification shows that the result depends partly on defining AI exposure more broadly than the traditional technology sector. Taken together, the robustness tests do not prove mispricing, but they consistently fail to support the alternative view that the market has already imposed a clear valuation discount on AI firms' reported gross margins.

10. Discussion

10.1 Interpreting the Three Results

The three empirical results should be interpreted jointly rather than as separate findings. Model 1 shows that gross margin remains highly value-relevant in technology firms, confirming that investors continue to treat it as a key signal of scalability and operating leverage. Model 2 shows that AI-exposed firms report significantly higher R&D intensity than traditional tech-adjacent firms, suggesting that their reported cost structures differ materially from the firms for which the traditional SaaS valuation logic was originally developed. Model 3 then shows that, despite this difference, the market does not appear to discount AI firms' gross margins. Instead, the interaction between gross margin and AI exposure is positive, although only marginally significant.

The central contribution of the thesis therefore lies in the combination of these findings. The market strongly rewards gross margin in technology firms, AI-exposed firms display signs of a different cost structure, yet investors do not appear to reduce the valuation weight placed on AI firms' reported gross margins. This pattern is consistent with the idea that the conventional gross margin pricing rule remains in use even when the underlying economics of product delivery may have changed.

10.2 Market Mispricing or Rational Uncertainty?

A key question is whether this pattern should be interpreted as evidence of market mispricing. The results are consistent with incomplete repricing, but they do not prove it. One interpretation is that investors continue to apply a SaaS-era valuation framework to AI-exposed firms, even though inference costs make the economics of AI products less scalable than traditional software. Under this interpretation, reported gross margins overstate the true operating leverage of AI firms, and the market has not yet fully incorporated this accounting-economic divergence.

However, an alternative interpretation is rational uncertainty. AI cost curves are changing rapidly, inference costs may decline over time, and firms may develop pricing models that allow them to pass compute costs on to customers. Investors may therefore be unwilling to penalize current gross margins if they believe today's cost structure is temporary or if they expect AI firms to achieve stronger future pricing power. From this perspective, the absence of a valuation penalty does not necessarily reflect investor error; it may reflect uncertainty about the long-run economics of AI.

The robustness results provide some support for the incomplete-repricing interpretation, but only cautiously. The pre/post-ChatGPT split shows that the gross margin interaction falls after generative AI becomes more visible, suggesting that the market may have started to incorporate the cost implications of AI. However, the coefficient remains positive and does not reverse into a penalty. This suggests attenuation rather than full correction.

10.3 Implications for Investors

For investors, the findings suggest that reported gross margin should be interpreted more cautiously for AI-exposed firms than for traditional software firms. In the SaaS context, high gross margins are often interpreted as evidence of near-zero marginal costs and strong future operating leverage. In AI-driven business models, this interpretation may be less reliable. A high reported gross margin may still reflect product quality, pricing power, or scale advantages, but it may also partly reflect the accounting location of compute-related costs.

This implies that investors should avoid treating gross margins as mechanically comparable across traditional SaaS firms and AI-exposed firms. Instead, gross margin analysis should be supplemented with information on R&D intensity, cloud infrastructure costs, compute dependence, cost-of-revenue disclosures, and the extent to which AI usage scales with customer activity. The thesis therefore does not suggest that gross margin has become irrelevant, but rather that its interpretation has become more conditional.

10.4 Implications for Accounting Standard-Setters

For accounting standard-setters, the findings point to a potential disclosure issue rather than necessarily a recognition issue. The results do not imply that current accounting treatment is incorrect or that firms are violating existing standards. Rather, they suggest that existing line-item definitions may not fully capture the economics of AI-driven business models. If recurring inference costs can plausibly be classified below gross profit, then reported gross margin may become less comparable across firms with different cost-allocation practices.

The historical precedent of software accounting is relevant here. ASC 985–20 was introduced to clarify software cost treatment during an earlier technological shift. Generative AI may create a similar need for more targeted guidance or disclosure around compute-related costs. Even if recognition rules remain unchanged, additional disclosure about how firms classify AI infrastructure, cloud, model-training, and inference-related costs would improve comparability and help investors better interpret reported margins.

10.5 Implications for Value-Relevance Research

For value-relevance research, the thesis illustrates how the relationship between accounting numbers and economic constructs can change when business models evolve. Gross margin remains statistically value-relevant in the technology sector, but the economic meaning of gross margin may not be stable across traditional software firms and AI-exposed firms. In traditional software, gross margin is closely linked to scalability. In AI-driven firms, the same accounting measure may also reflect classification choices around compute and infrastructure costs.

This finding is consistent with the broader value-relevance literature showing that accounting numbers are not mechanically informative in all settings. Their usefulness depends on whether the accounting measure continues to capture the economic construct investors believe it represents. The AI setting therefore provides a useful case for studying how technological change can weaken, or at least complicate, the interpretation of familiar accounting signals.

11. Limitations and Future Research

11.1 Treatment Group Identification

The first limitation concerns the identification of AI-exposed firms. The treatment group is based on current BOTZ and ROBO ETF membership, applied statically across the 2018–2024 sample period. This means that a firm included in one of these ETFs today is treated as AI-exposed throughout the full sample period, even if its AI exposure may have become more economically meaningful only later. This creates a potential timing issue in the treatment definition.

However, the purpose of the ETF-based approach is not to perfectly measure the exact year in which each firm became AI-exposed. Rather, it provides a transparent and replicable proxy for firms that the market plausibly recognizes as exposed to AI, robotics, or automation. This is appropriate for the thesis because the research question concerns not only underlying cost structures, but also whether publicly visible AI exposure is priced differently by investors. Future research could improve this identification by using point-in-time ETF holdings or constructing annual AI-exposure scores from 10-K filings, AI-related patents, or AI-related hiring data.

11.2 Binary AI Exposure Measure

A related limitation is that AI exposure is measured as a binary indicator rather than as a continuous intensity measure. In reality, firms differ substantially in the extent to which AI is central to their business model. For some firms, AI is core to product delivery and may directly affect variable costs through inference. For others, AI may be one component of a broader automation, robotics, or software strategy. Treating all AI-exposed firms as equally exposed may therefore introduce measurement noise.

This limitation likely makes it harder to detect a precise market response, since firms with very different degrees of AI dependence are grouped together. Future research could address this by measuring AI exposure continuously, for example through the frequency and context of AI-related terms in annual reports, AI-related patent activity, AI job postings, or segment-level revenue exposure. Such an approach would allow researchers to test whether the gross margin pricing effect is stronger among firms whose business models are more directly dependent on AI inference.

11.3 Intensity and Cost Classification

The third limitation concerns the interpretation of R&D intensity. Model 2 shows that AI-exposed firms report significantly higher R&D-to-Sales ratios than traditional tech-adjacent firms. This is consistent with the proposed mechanism that some AI-related costs may appear below gross profit rather than in COGS. However, the result should not be interpreted as direct proof of cost misclassification. Higher R&D intensity may reflect genuine innovation investment, model development, product experimentation, or long-term capability building.

The thesis therefore interprets the R&D result as evidence of a cost-structure divergence, not as definitive evidence of opportunistic accounting behavior. Public Compustat data do not allow us to observe whether specific compute, cloud, data-processing, or model-serving costs are classified as R&D or cost of revenue at the firm level. Future research could address this limitation through hand-collected 10-K disclosures, where researchers examine how firms describe cloud infrastructure, AI compute, model training, and inference-related costs in their accounting notes and management discussion sections.

11.4 Accounting Data and Gross Margin Measurement

A further limitation relates to the construction of gross margin. The thesis uses standardized Compustat data to ensure comparability and replicability across the full sample. However, Compustat's COGS measure may not always match firms' own reported "cost of revenue" figures in their 10-K filings. In particular, firms may differ in how they allocate depreciation, stock-based compensation, cloud infrastructure costs, or amortization across operating expense categories.

This means that the gross margin measure used in the regressions may differ from the exact gross margin figures investors observe in company filings or investor presentations. Because the methodology is applied consistently across firms, this issue should not mechanically bias the relative comparison between AI and non-AI firms. However, it does mean that the absolute magnitude of gross margins should be interpreted with caution. Future work could compare Compustat-based margins with hand-collected as-reported gross margins for a smaller sample of AI-exposed firms.

11.5 Further Research

Taken together, these limitations point toward several promising directions for future research. First, future studies could construct a point-in-time AI-exposure measure rather than relying on static ETF membership. Second, researchers could use textual analysis of 10-K filings to distinguish between firms that merely mention AI and firms whose revenues and costs are materially shaped by AI-related products. Third, hand-collected accounting disclosures could help identify whether AI compute costs are classified above or below gross margin. Finally, future work could extend the sample period as more post-ChatGPT observations become available, allowing a stronger test of whether the market eventually reprices AI firms' gross margins.

These extensions would allow future research to move from the broad question addressed in this thesis, whether the market currently prices AI firms' gross margins differently, toward a more precise understanding of when, where, and why the gross margin signal becomes less informative in AI-driven business models.

12. Conclusion

This thesis examined whether the established relationship between gross margin and valuation multiples in technology firms continues to hold in the context of generative AI. The central concern was that AI-driven products may involve materially different cost economics from traditional SaaS models, since inference costs scale with usage and may partly be classified below gross profit as R&D. If this is the case, reported gross margins may no longer capture scalability as cleanly as they did in the traditional software model.

The empirical analysis proceeded in three steps. First, the thesis established that gross margin is strongly value-relevant in the technology sector. The market attaches a substantially larger valuation premium to gross margins in technology firms than in the broader market, confirming that gross margin functions as an important scalability signal. Second, the analysis showed that AI-exposed firms report significantly higher R&D-to-Sales ratios than traditional tech-adjacent firms. This finding is consistent with the view that AI firms have a different cost structure and that some economically relevant costs may appear below gross profit. Third, the thesis tested whether investors adjust for this difference when valuing AI firms' gross margins. The results do not show evidence of a valuation penalty. Instead, the interaction between gross margin and AI exposure is positive across the main specification and robustness tests, although the strength of the evidence varies.

Taken together, the findings suggest that the market has not clearly adjusted the valuation weight placed on AI firms' reported gross margins, despite evidence that these firms differ in their reported cost structures. This does not necessarily prove market mispricing. Investors may rationally expect inference costs to decline, pricing models to improve, or AI firms to develop stronger scale economies over time. However, the results do suggest that the conventional SaaS-era interpretation of gross margin remains influential even in a setting where the underlying economics may have changed.

The contribution of the thesis is therefore twofold. Empirically, it provides early evidence that AI exposure does not currently lead investors to discount reported gross margins, even though AI-exposed firms report higher R&D intensity. Conceptually, it shows how technological change can weaken the link between an accounting measure and the economic construct it is assumed to represent. Gross margin remains value-relevant, but its interpretation has become more conditional.

The broader implication is that gross margin may no longer mean the same thing for all technology firms. In the SaaS era, it was often a clean signal of scalability and near-zero marginal cost. In the AI era, it may also reflect where compute-related costs are placed in the income statement. Until disclosure standards and investor models catch up with the economics of inference, reported gross margins should be interpreted with greater caution when valuing AI-exposed firms.

13. Bibliography

Academic and primary-source references appear in alphabetical order. Practitioner sources are listed separately at the end.

Academic and Primary Sources

Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. Journal of Accounting and Economics, 22(1–3), 3–30.

Ante, L., & Saggu, A. (2025). Quantifying a firm's AI engagement: Constructing objective, data-driven AI stock indices using 10-K filings. Technological Forecasting and Social Change, 210, Article 123841.

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics, 151, Article 103745.

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2024b). Artificial intelligence and firms' systematic risk (SSRN Working Paper No. 4868770).

Barth, M. E., Beaver, W. H., & Landsman, W. R. (2001). The relevance of the value relevance literature for financial accounting standard setting: Another view. Journal of Accounting and Economics, 31(1–3), 77–104.

Bernard, V. L., & Thomas, J. K. (1989). Post-earnings-announcement drift: Delayed price response or risk premium? Journal of Accounting Research, 27, 1–36.

Cao, Y., Liu, M., Qiu, J., & Zhao, R. (2024). Information in disclosing emerging technologies: Evidence from AI disclosure (SSRN Working Paper No. 4987085).

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. Journal of Finance, 56(6), 2431–2456.

Collins, D. W., Maydew, E. L., & Weiss, I. S. (1997). Changes in the value-relevance of earnings and book values over time. Journal of Accounting and Economics, 24(1), 39–67.

Eberhart, A. C., Maxwell, W. F., & Siddique, A. R. (2004). An examination of long-term abnormal stock returns and operating performance following R&D increases. Journal of Finance, 59(2), 623–650.

Eisfeldt, A. L., Schubert, G., & Zhang, M. B. (2023). Generative AI and firm values (NBER Working Paper No. 31222).

Financial Accounting Standards Board. (2009, codified). Accounting Standards Codification Topic 730: Research and Development. (Originally SFAS No. 2, 1974.)

Financial Accounting Standards Board. (2009, codified). Accounting Standards Codification Topic 985, Subtopic 20: Costs of Software to Be Sold, Leased, or Otherwise Marketed. (Originally SFAS No. 86, 1985.)

Financial Accounting Standards Board. (2025, September 18). Accounting Standards Update 2025–06: Internal-Use Software Costs (amendment to ASC 350–40).

Koh, P.-S., & Reeb, D. M. (2015). Missing R&D. Journal of Accounting and Economics, 60(1), 73-94.

Laplante, S. K., Skaife, H. A., Swenson, L. A., & Wangerin, D. D. (2019). Limits of tax regulation: Evidence from strategic R&D classification and the R&D tax credit. Journal of Accounting and Public Policy, 38(2), 89–105.

Lev, B., & Sougiannis, T. (1996). The capitalization, amortization, and value-relevance of R&D. Journal of Accounting and Economics, 21(1), 107–138.

Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. Journal of Accounting Research, 37(2), 353–385.

Microsoft Corporation. (2025). Form 10-K for the fiscal year ended June 30, 2025. U.S. Securities and Exchange Commission.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. Journal of Financial Economics, 108(1), 1–28.

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. Contemporary Accounting Research, 11(2), 661–687.

Poonawala, S. H., & Nagar, N. (2019). Gross profit manipulation through classification shifting. Journal of Business Research, 94, 81–88.

Rajgopal, S., Venkatachalam, M., & Kotha, S. (2003). The value relevance of network advantages: The case of e-commerce firms. Journal of Accounting Research, 41(1), 135–162.

SEC Investor Advisory Committee. (2025, December 4). AI Disclosure Recommendations.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review, 71(3), 289–315.

Trueman, B., Wong, M. H. F., & Zhang, X.-J. (2000). The eyeballs have it: Searching for the value in Internet stocks. Journal of Accounting Research, 38(Supplement), 137–162.

Practitioner Sources (industry data and firm-specific reporting)

Appenzeller, G., Bornstein, M., & Casado, M. (2023, April 27). Navigating the high cost of AI compute. Andreessen Horowitz.

Di Mattia, D., Gupta, K., Hernandez, F., Kasztelaniec, K., Khurana, T., & Shah, J. (2021, August 3). SaaS and the Rule of 40: Keys to the critical value creation metric. McKinsey & Company.

Efrati, A., & Woo, E. (2026, January). Anthropic lowers gross margin projection as revenue skyrockets. The Information.

Jaipuria, T. (2025, September 2). The state of AI gross margins in 2025. Tanay's Newsletter.

Lemkin, J. (2025, December 22). Have AI gross margins really turned the corner? The real math behind OpenAI's 70% compute margin. SaaStr.

Meritech Capital. (2024, March 7). Meritech Software Puls–Public SaaS Comparables.

Patel, D., Kourabi, A., Knuhtsen, R., & Doug. (2025, January 31). DeepSeek debates: Chinese leadership on cost, true training cost, closed model margin impacts. SemiAnalysis.

14. Appendix

14.1 AI Disclosure section

The authors of this thesis have used AI-based tools, primarily ChatGPT and Claude, as supportive instruments during the research and writing process. These tools have mainly been used to assist with programming related tasks in Python, including debugging code, identifying potential errors, and generating and restructuring code for descriptive statistics, data handling, and regression analysis. AI tools were also used as a supporting aid when structuring and implementing parts of the empirical regression framework and assessing the feasibility of certain methodological approaches.

In addition, AI tools have been used to a limited extent in the writing process. Their primary function has been to support language refinement, such as improving sentence clarity, suggesting alternative wording, and identifying grammatical inconsistencies. The analytical reasoning, interpretation of findings, theoretical framing, variable selection, and conclusions of the thesis remain entirely the responsibility of the authors.

The use of AI tools has contributed to the efficiency and overall quality of the thesis by reducing time spent on technical troubleshooting and by improving readability and linguistic precision. In the empirical analysis, AI assistance also helped streamline the coding and implementation process for regression models and data preparation. However, the authors recognize several limitations and risks associated with the use of AI generated assistance. First, AI generated outputs may contain inaccurate or misleading information, particularly in technical or methodological contexts. Second, excessive reliance on AI assistance while writing may result in repetitive language or reduced originality in expression.

To mitigate these risks, all AI-assisted outputs have been critically reviewed and verified by the authors. Programming suggestions and methodological guidance were cross-checked using academic literature, software documentation, course materials, and other external sources. Similarly, all written content and empirical specifications were reviewed and adapted to ensure consistency with the authors' own analysis, reasoning, and understanding of the methodology.

The authors conclude that AI tools can serve as valuable support instruments for limited technical, methodological, and linguistic tasks when used critically and responsibly. However, such tools should not be treated as authoritative sources of knowledge, nor as substitutes for independent academic judgment, methodological understanding, or original analysis.