Semantic Compensation: Mapping the Structural Displacement of Unrecognised Technological Intangible Assets in Corporate Reporting

Abstract

The rapid accumulation of Artificial Intelligence capital has exacerbated the measurement gap between economic reality and financial statements, as accounting standards (IAS 38) fail to capitalize key digital investments. This paper leverages Large Language Models (LLMs) to decode the "Shadow Balance Sheet" of AI capital from corporate narratives. Constructing a novel Technological Intangible Assets Disclosure Intensity (TIA-DI) index for the STOXX Europe 50 (2017–2024), we find that while recognized intangible assets remain static, narrative AI intensity has surged by 11.1%, signalling a hidden accumulation of productive capacity. We document a "Decoupling Effect": narrative disclosure is largely orthogonal to traditional R&D expenditures, except in digital sectors where it acts as a necessary signal. Our findings offer a scalable proxy for tracking AI diffusion and support ongoing EFRAG/IASB efforts to modernize intangible asset accounting."

1. Introduction

The rapid diffusion of Artificial Intelligence (AI) and data-driven technologies is reshaping the production function of modern firms, yet its impact on aggregate productivity statistics remains elusive, a phenomenon described as the "Productivity J-Curve" (Brynjolfsson, Rock, and Syverson, 2021). A primary explanation for this paradox is a measurement crisis: the massive investments required to deploy AI (ranging from data cleaning to organizational restructuring) are largely intangible and, under current accounting conventions, remain invisible in standard capital stock metrics (Corrado et al., 2021). The dislocation between a firm's economic reality and its financial statements has become a critical friction for economists and investors attempting to gauge the true extent of AI adoption.

This friction is rooted in the conservative design of accounting standards. As noted by Barker et al. (2022), regulations such as IAS 38 prioritize reliability over relevance, prohibiting the capitalization of most internally generated intangible assets. This creates a disconnect that regulators are currently rushing to address; the EFRAG (2024) and the IASB are actively reviewing IAS 38 to modernize the recognition of software and intellectual property. However, until these reforms materialize, the balance sheet, historically the primary source of truth for economic measurement, has lost its value relevance for digital-intensive firms (Lev, 2018). Relying solely on reported R&D expenditures fails to capture the specific accumulation of AI, software, and data assets, leading to systematic errors in estimating the firm's true productive capacity.

To bridge this measurement gap, recent literature has turned to alternative data sources. Babina et al. (2024), for instance, utilize employee resumes to measure AI investment through human capital. While valuable, labor-based metrics focus on *inputs* (hiring) and may overlook the proprietary data assets, algorithms, and strategic focus that constitute the firm's *organizational* capital. In this paper, we propose a complementary approach: decoding the "Shadow Balance Sheet" of AI capital directly from corporate narratives. We posit that managers, aware of the information asymmetry caused by accounting standards, use voluntary disclosure as a "compensatory mechanism" (Zambon et al.,

2020), signaling the existence of these invisible assets through the unstructured text of annual reports.

We introduce a novel measure, the Technological Intangible Assets Disclosure Intensity (TIA-DI) index, constructed using a hybrid Natural Language Processing (NLP) framework applied to the STOXX Europe 50 (2017–2024). Unlike traditional approaches that simply count keywords (Lewis and Young, 2019), our methodology leverages Large Language Models (specifically SBERT embeddings) to capture the *semantic centrality* of AI concepts within the firm's strategy. We posit that this disclosure intensity serves as an observable proxy for unobserved intangible capital. While we cannot directly impute the monetary value of these assets due to accounting constraints, our index captures the strategic commitment and resource allocation towards AI that R&D figures alone fail to reflect. Thus, we treat narrative disclosure not just as corporate communication, but as a rich signal of the firm's latent productive capacity.

This study advances the literature on economic measurement and corporate disclosure by refining the understanding of the "Decoupling Hypothesis" in the digital economy. Our core empirical contribution reveals a sectoral dichotomy. While we observe that narrative disclosure on intangibles operates independently of traditional R&D expenditures in the general sample, this relationship recouples strongly within the software and technology sectors. This finding challenges the assumption that standard innovation metrics uniformly capture digital adoption. It demonstrates that where the accounting blind spot is largest, in digital-native firms whose value drivers are almost entirely unrecognized intangibles, narrative disclosure acts as a necessary substitute for asset recognition. This suggests that economists relying solely on aggregate R&D data may systematically underestimate the capital stock of AI-intensive firms, while ignoring the compensatory signal managers are already sending through text.

By applying Large Language Models to decode these signals, this paper validates Semantic Compensation as a scalable econometric framework. We show that the market is already seeking, and finding, information in unstructured text that standard metrics miss. Our evidence underscores the urgent need for standard setters (EFRAG, IASB) to formalize the recognition of digital intangibles, as the current reliance on voluntary narrative substitution highlights the structural inadequacy of the balance sheet in the AI era.

2. Theoretical Framework and Hypotheses Development

2.1. Innovation Investment and Narrative Compensation

The relationship between innovation inputs (R&D) and voluntary disclosure constitutes a central tension in the accounting and economics literature. Under the lens of Signaling Theory and Voluntary Disclosure Theory, managers possess private information about the future cash flows generated by their intangible investments. Since R&D creates information asymmetry between insiders and outsiders (Aboody & Lev, 2000), managers have strong incentives to disclose private information to signal firm quality, reduce valuation uncertainty, and lower the cost of capital (Diamond & Verrecchia, 1991; Healy & Palepu, 2001).

Empirically, prior research supports a complementary relationship. Merkley (2014) documents that firms with high R&D intensity provide more informative narrative disclosures to contextualize their earnings performance, arguing that "hard" accounting numbers require "soft" narrative explanations to be interpretable. Similarly, Gu and Li (2007) find that managers of high-tech firms supplement financial statements with voluntary disclosures about innovation pipelines to guide analyst expectations and avoid undervaluation and Barth et al. (2001) demonstrated that in industries with high intangible intensity, analyst forecast errors are significantly higher, forcing managers to compensate with increased voluntary disclosure to close the valuation gap.

However, the specific nature of AI investments introduces unique frictions that complicate this relationship. First, the Proprietary Cost Hypothesis (Verrecchia, 1983) suggests a counterbalancing force: firms may withhold details about strategic AI algorithms to prevent knowledge spillovers to competitors (Li et al., 2018; Glaeser & Landsman, 2023). Unlike patentable hardware, AI models and data moats are often protected by secrecy rather than legal rights, increasing the cost of transparency. Consequently, highly innovative firms may rationally choose opacity to preserve their competitive advantage, creating a negative association between investment intensity and disclosure.

Second, is the "Invisible Balance Sheet" phenomenon. As argued by Lev (2018) and Barker et al. (2022), accounting standards like IAS 38 are designed for the industrial era, mandating the immediate expensing of internally generated intangibles due to uncertainty over future benefits. This creates a disconnect: a firm may be aggressively investing in AI (high economic capital) while reporting low or stagnant intangible assets on the balance sheet. Koh and Reeb (2015) identify this as missing R&D, showing that many innovative firms report zero R&D because their investments are classified as routine operating expenses.

Nevertheless, recent macro-productivity research offers a competing perspective: the "Decoupling" or "Orthogonality" hypothesis. Corrado et al. (2021) and Brynjolfsson et al. (2021) argue that a substantial portion of AI-driven value stems from intangible dark matter, such as organizational redesign, proprietary databases, and human capital training, that falls outside the traditional definition of R&D expenses. Under this view, narrative disclosure does not merely mirror R&D intensity but rather captures a distinct, orthogonal dimension of the firm's capital that input metrics fail to register. If the narrative captures this unmeasured dark matter rather than recognized inputs, the correlation between standard R&D intensity and AI disclosure may be weaker or non-existent, implying that they represent two independent information channels rather than a single redundant signal.

Given these opposing theoretical mechanisms, where signaling incentives drive a positive association to reduce asymmetry, while proprietary costs and the strategic protection of "invisible" assets drive a negative association to prevent spillovers, it is ex-ante unclear which effect dominates for technological intangibles. Therefore, we state the following hypothesis in non-directional form:

H1: There is an association between innovation intensity and the narrative disclosure of Technological Intangibles (TIA_DI).

2.2. The Information Chasm and Sectoral Sensitivity

The reliance on narrative disclosure is not uniform across the economy but is fundamentally conditioned by the firm's production function. While proprietary cost theory (Verrecchia, 1983; Glaeser & Landsman, 2023) might suggest that technology firms would hide their unique algorithms to prevent competitive mimicry, the information gap hypothesis suggests that for these firms, the link between investment and explanation must be structurally stronger.

In traditional sectors, capital is tangible and innovation often relates to process efficiency, which may not require extensive narrative justification. In contrast, for firms in the software and IT services sector, value creation is driven by scalability without mass (Haskel & Westlake, 2018). As Corrado et al. (2021) argue, the asset base of these firms, consisting primarily of code, data, and organizational designs, is immediately expensed under current accounting rules, creating a structural information chasm. Consequently, R&D expenditures in software firms are significantly more opaque to investors than R&D in manufacturing, where outcomes are often physical or patentable.

To bridge this gap, managers in digital sectors must rely on voluntary disclosure as a critical substitute for asset recognition. Barth et al. (2001) demonstrate that in high-intangible sectors, the market penalizes unexplained R&D with higher valuation uncertainty. Furthermore, Babina et al. (2024) find that for tech firms, AI investment primarily drives product innovation rather than cost efficiencies. Unlike process improvements, new AI-driven product value propositions require complex narrative explanations to be understood by customers and investors.

Therefore, we posit that the software effect operates as a magnifying glass on innovation efforts. It is not merely about a different baseline of disclosure, but about a stronger sensitivity: in digital sectors, every unit of innovation investment generates a greater imperative for narrative explanation than in traditional industries.

H2 *The association between innovation intensity and narrative AI disclosure (TIA-DI) is significantly stronger for firms in the Software and Technology services sectors than for firms in traditional industries.*

2.3. The Moderating Role of Institutional Protection

The relationship between innovation investment and corporate narrative is not structurally invariant but is deeply embedded in the institutional architecture of the country. The Law and Finance literature, pioneered by La Porta et al. (1998), establishes that the quality of legal enforcement and investor protection fundamentally shapes the incentives for financial reporting. However, how these institutional factors moderate the voluntary disclosure of unregulated intangibles remains a subject of theoretical debate, characterized by two countervailing forces: the demand for credibility versus the substitution of trust.

One stream of literature suggests a complementarity effect, arguing that strong legal institutions are a prerequisite for credible voluntary disclosure. As demonstrated by Leuz et al. (2003), in regimes with high investor protection, insiders face severe penalties for expropriation and misleading reporting. Consequently, the soft information contained in

narrative disclosures becomes a credible signal rather than mere "cheap talk." Under this view, firms in strong institutional environments are more likely to leverage narrative disclosure to reduce the information asymmetry inherent in R&D, as the market places a higher valuation on verified communication. Conversely, the substitution hypothesis suggests that strong institutions reduce the marginal necessity of voluntary disclosure. In environments with weak investor protection, financial statements lack credibility due to the high risk of earnings management or asset diversion. To compensate for this trust deficit, managers must rely heavily on narrative explanations as a bonding mechanism to justify their investment decisions to skeptical investors (Doidge et al., 2007).

Recent analytical and empirical work reinforces this substitution channel through the lens of proprietary costs Casella et al. (2025) argue that for firms with high intangible capital, the cost of public disclosure is elevated due to knowledge spillovers that benefit competitors. In jurisdictions with strong institutions, where audited financial figures already command a baseline of trust, managers may optimally choose opacity regarding their specific technological pipeline to protect their intellectual property. As Glaeser and Landsman (2025) demonstrate, the decision to disclose proprietary information is a strategic trade-off; when the institutional "shield" sufficiently assures investors against agency risks, the imperative to provide granular narrative details diminishes, allowing firms to shield their innovations from competitive mimicry.

Given these opposing theoretical mechanisms—where strong institutions may enhance the credibility of disclosure but simultaneously reduce the market's demand for it due to lower agency costs and higher proprietary concerns—it is ex-ante unclear which effect dominates for technological intangibles. Therefore, we state the following hypothesis in non-directional form:

H3: The country-level institutional protection of investors significantly moderates the association between firm-level innovation intensity and the narrative disclosure of technological intangibles.

3. Methodology

3.1 Sample and Data

This analysis is based on a corpus of annual reports and equivalent corporate disclosure documents in text format. The dataset comprises the Annual Reports of companies listed in the STOXX Europe 50 index for the period 2017–2024. All documents are initially processed to ensure coding consistency. The text is normalised to lowercase and special characters are replaced with blank spaces to ensure compatibility with subsequent tokenisation routines. Word count statistics are calculated before and after pre-processing to maintain transparency regarding data reduction.

3.2. Index TIA DI

We introduce the Technological Intangible Assets Disclosure Intensity (TIA-DI) index, a novel composite metric designed to capture the "Shadow Balance Sheet" of the modern firm. Unlike traditional disclosure proxies that rely on unidimensional keyword frequency (e.g., Loughran & McDonald, 2011), the TIA-DI framework operationalizes the multidimensionality of AI capital through a triadic architecture: (i) Scope (lexical

density), (ii) Thematic Structure (latent topics), and (iii) Strategic Centrality (semantic embedding distance).

This methodology represents a methodological leap from "Bag-of-Words" approaches to Semantic Compensation. By integrating probabilistic topic modeling (LDA) with state-of-the-art transformer language models (SBERT), the index transcends simple term-counting to quantify the semantic proximity between a firm's narrative and the frontier of AI technology. This allows us to distinguish between firms that merely mention buzzwords ("AI-washing") and those where AI concepts are structurally embedded in the value creation narrative.

The construction of the index follows the OECD/Eurostat Handbook on Constructing Composite Indicators (Joint Research Centre, 2008) but adapts it to the domain of unstructured economic data. By triangulating statistical precision (TF-IDF) with neural contextual understanding (Embeddings), the TIA-DI offers a robust, scalable instrument for economists to "impute" the missing stock of intangible capital directly from corporate text.

3.2.1. Component 1: Volume intensity analysis

The volume intensity component measures the lexical density of technology-related dissemination relative to the total word population of the document, weighted by categories of conceptual importance. The analysis uses a curated dictionary of domain-specific terms organised into five thematic categories that reflect different aspects of the technology, AI, and data ecosystem based on the work of (Cao et al., 2024; Lewis & Young, 2019; Wang, 2024):

- Artificial Intelligence: ‘artificial intelligence,’ ‘machine learning,’ ‘deep learning,’ ‘neural network,’ ‘generative AI,’ ‘large language model’.
- Data Capital: ‘big data,’ ‘data analysis,’ ‘data science,’ ‘data mining,’ ‘data infrastructure,’ ‘data assets’.
- Algorithms: ‘algorithm,’ ‘algorithmic,’ ‘optimisation,’ ‘recommendation system,’ ‘prediction model’.
- Software Engineering: ‘software development,’ ‘cloud computing,’ ‘microservices,’ ‘platform architecture,’ ‘DevOps’.
- Technical Talent: ‘data scientist,’ ‘software engineer,’ ‘ML engineer,’ ‘engineering team,’ ‘R&D team’.

Keyword occurrence is identified by case-insensitive pattern matching and delimited by word boundaries to avoid partial word matches. For multi-word phrases (e.g., ‘machine learning,’ ‘data infrastructure’), the entire phrase is matched as a unit to preserve semantic integrity. These frequencies were then normalised by dividing them by the total number of words in the document, thus obtaining a category proportion or thematic intensity ratio (frequency/total_words) for each category and document. This procedure generates the component indicators x_{ki}, which are comparable between documents of different lengths and form the basis for constructing the composite index, in line with the general formulation (Chen et al., 2022):

$$C_i = \sum_k w_k \, x_{ki}$$

Subsequently, these category proportions were aggregated using a weighted linear combination based on expert judgement and theoretical grounding in digital strategy literature. Specifically, weights were assigned to each category according to its conceptual importance in the literature on digital capabilities. The strategy digital contemporánea requires moving beyond technological investment to focus on the capacity to exploit information and massive datasets in competitive decisions (Bharadwaj et al., 2013). In this context, artificial intelligence has evolved from an efficiency tool to a central strategic dynamic capability: firms utilizing AI to explore new value creation forms are 2.7 times more likely to capture competitive opportunities (Ransbotham et al., 2021). Consequently, weights prioritize value *exploitation* categories over *enabling input* categories: Artificial Intelligence (0.30), Data Capital (0.25), Algorithms (0.20), Software Engineering (0.15) and Technical Talent (0.10). Thus, for each document i, the composite technological emphasis index was calculated as:

$$C_i = 0.30 \cdot x_{\text{AI},i} + 0.25 \cdot x_{\text{Data},i} + 0.20 \cdot x_{\text{Algo},i} + 0.15 \cdot x_{\text{SoftEng},i} + 0.10 \cdot x_{\text{Talent},i}$$

where $x_{\text{cat},i}$ denotes the normalised proportion of terms from each category in document i. This formulation follows the mathematical framework for constructing composite indices through linear combination of component indicators, ensuring that the resulting index reflects both the relative presence of each category and its theoretical weight in the construct of interest.

3.2.2. Component 2: Thematic Approach Analysis (LDA Modelling)

The thematic focus component identifies the extent to which a company's disclosure is organised around topics semantically related to technology and artificial intelligence (TIA). All documents underwent a standardised NLP pre-processing workflow using the NLTK library. The text was fragmented into individual tokens, subsequently removing punctuation, converting the content to lowercase, and filtering out stop words that do not contribute semantic value. Tokens of less than three characters were also filtered out, and lemmatisation was applied to consolidate morphological variations into their linguistic roots.

To capture the disclosure structure beyond mere keyword density, a document-term matrix (DTM) was constructed. The matrix was configured with a vocabulary of 1,500 features, filtering terms with a document frequency greater than 80% or less than 2 occurrences, and including n-grams (1, 2) to preserve compound concepts such as ‘artificial intelligence’. The Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003), a Bayesian hierarchical model that allows each document to be represented as a mixture of latent topics, was applied to this matrix. The model was run in batch mode with 10 topics and a random seed to ensure the reproducibility of the results (Pedregosa et al., 2011).

The relevance of the extracted topics was evaluated by examining the 20 main terms of each topic. A topic was classified as ‘relevant to TIA’ if at least three of its main terms had semantic overlap with the defined TIA keyword dictionary. For each document i, the

Topic Focus Score (SET_i) was calculated as the sum of the document-topic probabilities of all topics identified as relevant:

$$SET_i = \sum_{k \in K_{TIA}} P(\text{topic}_k \mid \text{document}_i)$$

where K_{TIA} represents the set of topics related to technology and AI. This score, ranging from 0 to 1, quantifies the proportion of the document devoted to technological narratives, allowing for a robust measurement of the firm's strategic focus.

3.2.3. Component 3: Strategic Centrality Analysis (Dual Validation)

The strategic centrality component assesses the degree of alignment between a company's corporate document (DC_i) and its narrative commitment to technology and value creation. This dimension distinguishes between peripheral mentions and a central strategic positioning by comparing the document with two synthetic reference anchors: a synthetic reference document composed of all the terms in the TIA dictionary concatenated (TIA_{REF}) and a synthetic reference document composed of strategic keywords (e.g., 'strategy,' 'strategic,' 'competitive advantage,' 'value creation,' 'innovation,' 'growth,' 'digital transformation,' 'business model') ($STRAT_{REF}$).

For each document, we calculate the similarity in two different vector spaces. Where v_i is the vector representing document DC_i and r_k is the reference anchor vector k. Similarity is defined by the cosine of the angle between several vectors (Sun, 2024):

$$S(DC_i, REF_k) = \frac{v_i \cdot r_k}{\| v_i \| \| r_k \|} = \frac{\sum_{j=1}^{n} v_{i,j} \cdot r_{k,j}}{\sqrt{\sum_{j=1}^{n} v_{i,j}^2} \cdot \sqrt{\sum_{j=1}^{n} r_{k,j}^2}}$$

where:

- In TF-IDF space, the component $v_{i,j}$ epresents the statistical weight j in document i, weighted by its relative importance in the corpus.
- In the Embeddings (SBERT (Reimers & Gurevych, 2019)) space, the vector $v_i \in R^{384}$ captures latent semantic features using the pre-trained *all-MiniLM-L6-v2*.

To ensure that technological content is evaluated within a business management framework, the centrality score C_i will be defined as a linear combination of similarities with both anchors:

$$C_i = \alpha \cdot S(DC_i, TIA_{REF}) + (1 - \alpha) \cdot S(DC_i, STRAT_{REF})$$

A priority weight has been established for technological relevance, with $\alpha = 0.6$ and a 40% weight given to the strategic context of value creation.

The final centrality component is obtained by averaging the results of the statistical (TF-IDF) and neural (SBERT) models, thus minimising the specific biases of each technique:

$$C_{i,Final} = \frac{C_{i,TF\text{-}IDF} + C_{i,Emb}}{2}$$

The use of ensemble models makes the index more robust by combining multiple independent predictors, thereby minimising errors. By incorporating advanced text representations such as document embeddings, the model captures qualitative information that is essential for identifying profitable signals in corporate disclosures, surpassing analysis based solely on standard terminology (Schmitz et al., 2023).

3.2.4. Composite index integration.

Finally, we aggregate the analysed dimensions into a single indicator, the TIA-DI. To mitigate inherent biases, we adopt an ensemble averaging approach that combines the statistical precision of lexical frequencies with the semantic depth of transformer language models.

$$TIA\text{-}DI_{i,\text{met}} = 0.40(V_i) + 0.30(F_i) + 0.30(C_{i,\text{met}})$$

where:

- The volume score V_i provides the primary signal of material intensity.
- The focus score F_i ensures that the content is not scattered terminology, but rather organised into coherent themes identified by LDA.
- Strategic centrality C_i validates that technological concepts are positioned as pillars of the business model and value creation.

Este enfoque de ensemble mitiga sesgos inherentes a cada método individual: evita sobreponderación de extensión textual (balance de Volume con Focus), reduce incidencia de menciones incidentales (Strategic Centrality), y asegura que el índice refleja tanto magnitud como profundidad de disclosure tecnológico (Blei et al., 2003; Reimers & Gurevych, 2019; Gentzow et al., 2019). The final value of the index for each company $TIA\text{-}DI_{i,\text{Final}}$ is obtained by averaging both implementations:

$$TIA\text{-}DI_{i,\text{Final}} = \frac{TIA\text{-}DI_{i,\text{TF-IDF}} + TIA\text{-}DI_{i,\text{Emb}}}{2}$$

This ‘double layer’ architecture reinforces construct validity. While the TF-IDF component captures the technical accuracy of the terms used (what the company says), the Embeddings component captures the underlying semantic context (what the company means). The convergence of both results provides a robust metric against the use of boilerplate text and technological ‘AI-washing’.

3.2.5. Validation of the Dissemination Measure (TIA_DI)

To ensure that the proposed TIA-DI index reliably captures the strategic disclosure intensity of Technological Intangible Assets (TIA) and is not merely a statistical artefact, we subjected the variable to a multidimensional validation protocol.

First, we assessed discriminant validity by controlling for verbosity bias. Correlation analysis between TIA_DI and total document length (word count) revealed near-total

independence ($r = -0.065$, $p > 0.10$). This result confirms that the index captures the qualitative density of AI information and does not simply reward the most extensive reports, overcoming a common limitation in textual analysis studies.

Second, we verified internal consistency and semantic convergence. A positive correlation ($r = 0.136$) was observed between lexical volume (term frequency) and embedding-based centrality (SBERT). This indicates structural coherence: companies that devote more textual space to AI tend to integrate it more deeply into their core strategic narrative, validating the hybrid architecture of the index.

Finally, the index demonstrated high discriminatory power. The t-test between the upper and lower quartiles of the sample yielded a t-statistic of 34.27 ($p < 0.001$), showing that the TIA_DI is highly sensitive in distinguishing between companies with a clear strategic orientation towards AI and those with superficial mentions.

3.3 Variables

To test the hypotheses data downloaded from the LSEG platform (https://www.lseg.com/en) was used. Given the presence of missing values in key accounting variables, and to avoid the selection bias that would result from eliminating these observations, the values were imputed using the MICE (Multivariate Imputation and Chained Equations) technique with the Random Forest Regressor base estimator (Van Buuren & Groothuis-Oudshoorn, 2011).

The primary independent variable is R&D Intensity, calculated as the ratio of annual research and development expenditure to total sales, which serves as a proxy for the firm's tangible investment in innovation. To capture sectoral heterogeneity, we employ Sector Technology Intensity, a binary indicator set to one if the firm operates in Software and IT services or Fintech sectors, and zero otherwise. The institutional environment is measured using the International Property Rights Index (IPRI) - Intellectual Property Subindex, developed by the Property Rights Alliance (https://internationalpropertyrightsindex.org/countries) . Specifically, we construct a binary indicator, Strong_Inst_Prot, set to one if the firm is listed in countries belonging to the first quartile (highest protection) of the IPRI Intellectual Property Subindex, and zero otherwise.

We also include a set of firm-level control variables to mitigate omitted variable bias. Firm Size is measured as the natural logarithm of total assets to control for visibility and resource availability. Profitability is proxied by Return on Assets (ROA), calculated as the ratio of net income to total assets. Analyst Coverage, defined as the natural logarithm of the number of analysts plus one, controls for the external demand for information. Finally, Leverage, measured as the ratio of total debt to total assets, accounts for the influence of capital structure on disclosure incentives.

3.4. Panel data analysis

To address potential endogeneity arising from omitted variable bias, our baseline specification relies on Pooled Ordinary Least Squares (OLS) models that include an extensive set of time-varying controls (Firm Size, ROA, Leverage), year fixed effects to capture macroeconomic shocks common to all firms, and standard errors clustered at the

firm level to account for serial correlation and heteroscedasticity within firms (Petersen, 2009; Wooldridge, 2010). Although firm fixed effects are common in panel data analyses, their use would prevent the estimation of coefficients for key time-invariant variables central to this study, such as Sector (H2) and Country (H3). In addition, fixed time dummies control for temporal heterogeneity, while the inclusion of a comprehensive set of firm-level controls mitigates the risk of omitted variable bias. Given that TIA-DI reflects a firm's disclosure behaviour rather than a direct investment metric, the risk of reverse causality between technological disclosure and R&D Intensity, though possible, is expected to be relatively small.

The model specification is as follows:

$$TIA\ DI_{it} = \beta_0 + \beta_1 RD\ Intensity_{it} + \beta_2 Controls_{it} + \gamma Year_t + \varepsilon_{it}$$

To analyse the moderating effects (hypotheses 2-5), interaction terms were introduced:

$$\begin{aligned} TIA\ DI_{it} = {} & \beta_0 + \beta_1 RD\ Intensity_{it} + \beta_2 Controls_{it} \\ & + \beta_3 (RD\ Intensity_{it} \times Moderator_{it}) + \beta_4 Log\ Assets_{it} + \beta_5 ROA_{it} \\ & + \sum_{t=1}^{T-1} \gamma_t\, Year_t + \varepsilon_{it} \end{aligned}$$

4. Results

4.1 Temporal Evolution and Descriptive Analysis

The longitudinal analysis reveals a distinct upward trend in TIA reporting. The mean TIA_DI score increased from 0.0279 in 2017 to a peak of 0.0310 in 2023. This 11.1% growth suggests that firms are increasingly using narrative sections to signal value. Furthermore, the standard deviation has contracted (from 0.0203 to 0.0159), indicating institutional isomorphism: what was once a differentiator has become a standardised market expectation.

Table 1: Yearly Statistics of TIA_DI (Embedding-Based)

Year	Mean	Median	Std. Dev.	N
2017	0.0279	0.0271	0.0203	42
2018	0.0258	0.0287	0.0192	45
2019	0.0249	0.0266	0.0185	45
2020	0.0275	0.0279	0.0193	44
2021	0.0257	0.0275	0.0174	46
2022	0.0286	0.0282	0.0171	45
2023	0.0310	0.0314	0.0169	46
2024	0.0291	0.0298	0.0159	47

As illustrated in Figure 1 (Mean TIA_DI vs. Std Dev), the mean disclosure intensity (solid line) exhibits structural growth from 2017, interrupted briefly in 2021—possibly due to a temporary narrative shift towards supply chain resilience post-COVID—before rebounding strongly to reach a historical peak in 2023 (0.0310). This peak chronologically coincides with the widespread emergence of Generative AI, compelling firms to explicitly position themselves regarding this technology in their annual reports.
Even more revealing is the shaded area (or error bars) representing the standard deviation. Visually, the dispersion progressively contracts towards 2024. This indicates a phenomenon of institutional isomorphism: the gap between "leading" and "lagging" firms is closing. What was a voluntary differentiation strategy for a few firms in 2017 had become a standardized market expectation by 2024, homogenizing the technological discourse across the STOXX 50 index.

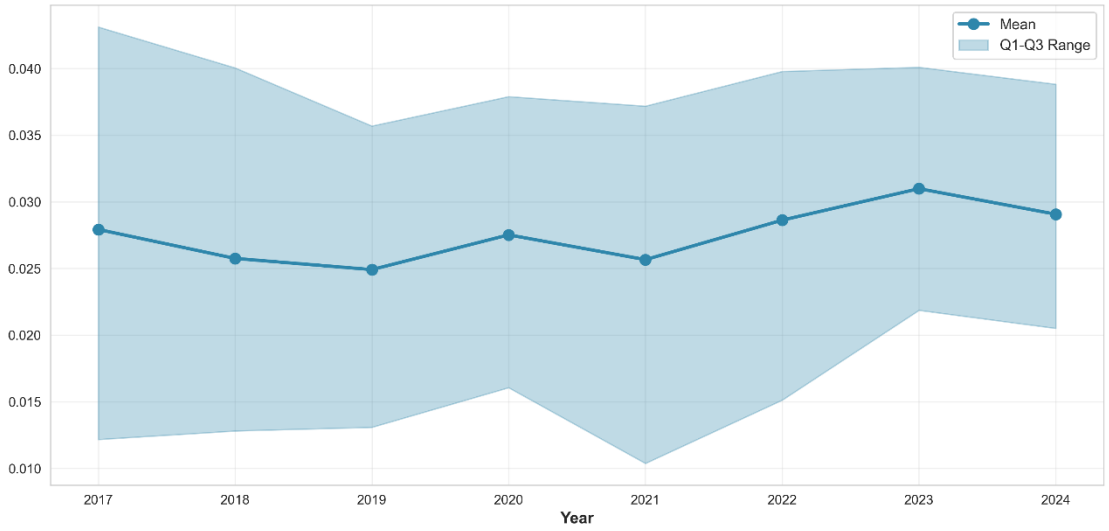


Figure 1. TIA_DI index

Table 2 presents the ranking of firms based on their average Technological Intangible Assets Disclosure Intensity (TIA-DI). The descriptive statistics reveal a counter-intuitive pattern: the unconditional volume of AI narrative is highest among sectors undergoing digital transformation rather than among digital natives.

The top rankings are dominated by large incumbents in the Financial and Industrial sectors. BNP Paribas (0.0466), Airbus (0.0451), and BASF (0.0427) exhibit the highest intensity, reflecting the significant communication efforts required to signal the integration of "Industry 4.0" and Fintech paradigms into traditional business models. In contrast, the Software and Technology sector exhibits a lower average volume of disclosure. Notably, SAP, the leading software firm in the sample, appears in the bottom quartile (Rank 47). This ranking highlights a key structural difference: while traditional sectors maintain high baseline levels of disclosure—potentially to manage transformation expectations—technology firms appear to be more reserved in their unconditional narrative intensity. As we demonstrate in the subsequent multivariate analysis, this lower volume does not imply less innovation, but rather a more direct "recoupling" between disclosure and R&D investment for digital natives.

Table 2: Top Disclosing Firms by Average TIA_DI

Rank	Empresa	Mean	Vol	Max	Rank	Empresa	Mean	Vol	Max
1	BNP PARIBAS	.0466	.0137	.0624	**25**	INDITEX	.0273	.0105	.0441
2	AIRBUS	.0451	.0133	.0697	**26**	ENI	.0262	.0189	.0417
3	BASF	.0427	.0199	.0734	**27**	ADIDAS	.0262	.0171	.0403
4	NOKIA	.0425	.0322	.0774	**28**	SCHNEIDER	.0257	.0085	.0431
5	DANONE	.0398	.0070	.0489	**29**	ASML	.0248	.0319	.0713
6	VOLKSWAGEN	.0397	.0033	.0433	**30**	INFINEON	.0243	.0042	.0300
7	SIEMENS	.0397	.0105	.0512	**31**	BBVA	.0235	.0245	.0633
8	INTESA SANP.	.0389	.0041	.0430	**32**	AXA	.0223	.0042	.0277
9	PROSUS	.0387	.0143	.0638	**33**	AHOLD DELHAIZE	.0219	.0077	.0317
10	ING GROUP	.0384	.0100	.0469	**34**	SANTANDER	.0218	.0209	.0496
11	ENEL	.0374	.0164	.0676	**35**	IBERDROLA	.0207	.0167	.0517
12	DEUTSCHE BÖRSE	.0372	.0234	.0888	**36**	ALLIANZ	.0197	.0172	.0563
13	BKW	.0371	.0122	.0583	**37**	ADYEN	.0185	.0241	.0602
14	MERCEDES-BENZ	.0368	.0124	.0560	**38**	STELLANTIS	.0179	-	.0179
15	LVMH	.0361	.0160	.0517	**39**	VONOVIA	.0175	.0112	.0317
16	AIR LIQUIDE	.0361	.0157	.0512	**40**	VINCI	.0174	.0151	.0405
17	SAINT-GOBAIN	.0342	.0251	.0641	**41**	HERMÈS	.0156	.0072	.0266
18	KONE	.0340	.0125	.0620	**42**	SANOFI	.0135	.0115	.0385
19	DEUTSCHE TELEK.	.0336	.0040	.0389	**43**	BAYER	.0128	.0072	.0240
20	NORDEA	.0325	.0154	.0472	**44**	ESSILORLUXOT.	.0128	.0099	.0270
21	DHL GROUP	.0302	.0152	.0601	**45**	FERRARI	.0100	.0085	.0200
22	L'ORÉAL	.0293	.0130	.0444	**46**	FCA (FIAT)	.0077	.0091	.0234
23	TOTALENERGIES	.0287	.0089	.0384	**47**	SAP	.0031	.0150	.0218
24	MUNICH RE	.0286	.0088	.0432	**48**	AB INBEV	.0004	.0156	.0382

4.3 Correlation Analysis

Table 3 presents the pairwise Pearson correlation coefficients between the main variables. The preliminary bivariate evidence indicates that narrative disclosure intensity (TIA-DI) is negatively associated with profitability (ROA, r = -0.27***), suggesting that firms with weaker financial performance tend to provide more extensive technological narratives. Conversely, we observe a positive correlation with firm size (Log Assets, r = 0.26***), consistent with larger firms having greater reporting resources. Notably, the correlation between R&D Intensity and TIA-DI is statistically significant (r = 0.18**), indicating a positive relationship between research and development investments and the intensity of technological narrative disclosure—a finding that contrasts with preliminary expectations and which we examine in the subsequent multivariate analysis.

Table 3: Correlations with statistic significance (*** p<0.01, ** p<0.05, *p<0.1)

	TIA_DI	**R&D Intensity**	**ROA**	**Log Assets**	**Analysts**	**Strong_Inst Prot**
TIA_DI	1.00	0.18**	-0.27***	0.17***	-0.01	0.15*
R&D Intensity	0.18**	1.00	-0.15*	0.11	-0.28***	0.37***
ROA	-0.27**	-0.15*	1.00	-0.70***	0.03	-0.09
Log Assets	0.26***	0.11	-0.70***	1.00	-0.21**	0.15*
Analysts	-0.01	-0.28***	0.03	-0.21**	1.00	-0.28***
Strong_Inst Prot	0.15*	0.37***	-0.09	0.15*	-0.28***	1.00

4.4 Empirical Results and Discussion

Our econometric analysis provides evidence on how European firms strategically use narrative disclosure to signal their technological capital. The baseline regression (H1) reveals that R&D intensity alone does not statistically predict higher levels of narrative AI disclosure. The coefficient is positive (0.008) but lacks statistical significance ($p > 0.10$), with a standard error of (0.009). This null finding contradicts classical signaling theory, which would posit a direct link between innovation inputs and disclosure, and instead supports a decoupling hypothesis. In the digital economy, firms often expense critical AI investments—such as cloud infrastructure and data acquisition—as SG&A rather than capitalizing them as R&D. Consequently, our TIA-DI index captures a component of organizational capital that is distinct from traditional accounting metrics. This finding aligns with Babina et al. (2024), who argue that AI-driven growth stems from product-innovation investments that do not correspond linearly to reported R&D figures. Thus, our index provides incremental information to the market rather than merely mirroring reported expenditures.

This general lack of association shifts dramatically when we account for sectoral heterogeneity. We find strong evidence for the Software Effect (H2), as the interaction term R&D × Software is positive and statistically significant (0.038*, $p < 0.10$, SE = 0.010). The Software sector dummy itself is negative and highly significant (-0.0271***, $p < 0.01$), indicating that software firms have lower baseline disclosure intensity, but this deficit is systematically compensated by a stronger R&D-disclosure link. This result resolves the apparent "decoupling" observed in the full sample: while traditional firms do not systematically translate R&D spending into narrative AI disclosure, digital-native firms exhibit a strong recoupling between investment and signal. This finding validates the theoretical framework of Haskel and Westlake (2018) regarding scalability without mass. In the Software and IT Services sector, where primary value drivers (code, algorithms, data moats) are immediately expensed under current accounting rules (Corrado et al., 2021), R&D expenditures are inherently more opaque than in manufacturing. Consequently, for these firms, the financial statements are structurally less informative, rendering narrative disclosure a necessary substitute rather than a discretionary choice. The data suggest that software firms are compelled to use narrative channels to bridge the valuation gap generated by their invisible capital, a pattern that aligns with the concentration of intangible intensity noted by Bloom et al. (2021). Thus, the innovation signal is not absent; it is simply conditional on the intangible nature of the firm's production function.

Figure 2 illustrates the relationship between R&D intensity (X-axis) and TIA-DI (Y-axis), differentiating observations from the software and IT sector (blue dots) from other sectors (red dots). The regression lines with 95% confidence intervals reveal a positive and significant slope only in Software (solid red), while in other sectors (blue) the association is flat or slightly negative. This visualises the ‘Software Effect’: digital firms compensate for their lower disclosure baseline with greater R&D-disclosure sensitivity, resolving the apparent ‘decoupling’ of the aggregate sample.

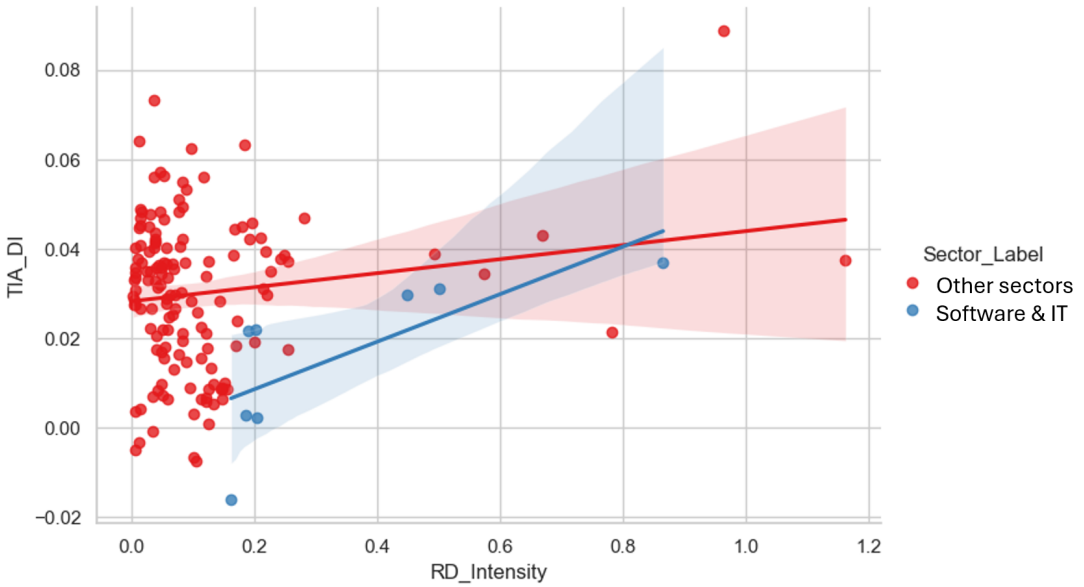


Figure 2. Scatter plot of TIA-DI versus R&D intensity by sector.

Regarding the institutional environment (H3), we fail to find statistically significant evidence for a moderating effect. Although the interaction term between R&D intensity and institutional protection exhibits a negative coefficient (-0.004, SE = 0.018), consistent with the directionality of the substitution hypothesis (where stronger legal shields might reduce the need for narrative signaling), the effect is statistically indistinguishable from zero ($p > 0.10$). Similarly, the direct association between strong institutions and disclosure levels is positive but insignificant (0.0048, SE = 0.0068). These null results suggest that, at least within our European sample, the pressure to disclose technological intangibles is driven more by firm-specific strategic incentives, such as capital intensity and sector (H2), than by country-level legal variations.

The Robustness specification reveals a negative and highly significant coefficient for R&D Intensity scaled by firm assets (-0.213**). This suggests that when innovation is measured relative to the asset base rather than sales, the relationship turns negative, indicating that firms with high capital-adjusted R&D intensity—potentially those with concentrated proprietary tech—prefer opacity over disclosure to protect competitive advantages.

Table 3. Regression Analysis Results (All Hypotheses)

Independent Variable	**H1Baseline**	**H2Software**	**H3Inst. Prot.**	**Robustness(Alt. IV)**
RD Intensity	0.008 (0.009)	0.011 (0.009)	0.007 (0.012)	
RD Intensity (Assets)				-0.213*** (0.076)
ROA	-0.104* 0.056)	-0.088* (0.051)	-0.104* (0.057)	-0.048 (0.030)

Log(Assets)	0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	
Software Dummy		-0.0271*** (0.034)		
RD Int. X Software (H2)		0.038* (0.010)		
Analysts				
Strong Inst Prot			0.0048 (0.0068)	
RD Int. X Inst. Prot. (H3)			-0.004 (0.018)	
Year Fixed Effects	Yes	Yes	Yes	Yes

Note: The dependent variable is the Technological Intangible Assets Disclosure Intensity (TIA_DI). Robust standard errors clustered at the firm level are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

To further address potential endogeneity concerns arising from simultaneity between R&D investment and technological disclosure, we implemented several robustness checks. First, we estimated specifications including lagged values of the main regressors (Lagged_IV), which capture persistence in firms' behaviour and reduce the risk that contemporaneous shocks jointly affect R&D Intensity and TIA-DI. Second, we conducted a Hausman-type endogeneity test by augmenting the baseline model with the first-stage residuals from an auxiliary regression of R&D Intensity on its predictors; the residual term is positive and statistically significant (coefficient = 0.0564, p = 0.0332), indicating that R&D Intensity is endogenous and that naïve OLS estimates may be biased. Third, we re-estimated the model in first differences, which removes time-invariant unobserved heterogeneity at the firm level. Across these alternative specifications, the substantive pattern of results remains broadly consistent, suggesting that, conditional on controls and year effects, our main inferences are not driven by severe endogeneity problems.

5. Conclusion

This study provides quantitative evidence regarding the existence of a shadow balance sheet for technological capital, demonstrating that the visibility of artificial intelligence assets is determined by strategic choices rather than being a direct output of innovation expenditures. Our analysis reveals a decoupling between traditional research and development expenses and narrative signals in the general population of firms. This disconnect suggests that standard accounting metrics fail to capture the full extent of digital adoption in the broader economy. However, this relationship changes significantly within digital sectors. We find that for software and technology service firms, narrative disclosure acts as a necessary substitute for asset recognition, effectively bridging the information gap created by the expensing of intangible investments.

Finally, regarding the institutional environment, we observe a limited influence of legal frameworks on disclosure strategies. Although the data hints at a potential substitution

pattern, where strong legal regimes might slightly reduce the marginal pressure to signal innovation, the effect is negligible compared to firm-level drivers. This suggests that the strategic decision to reveal or withhold technological details is primarily dictated by sectoral imperatives and proprietary costs, rather than by the broader country-level legal architecture.

These findings are subject to limitations. First, our measure of narrative disclosure relies on the frequency of specific keywords, which may not fully capture the qualitative nuance or sentiment of the disclosures. Second, while we control for industry and country factors, our sample is limited to European firms, potentially restricting the generalizability of the institutional substitution effect to other regions with different legal traditions, such as the United States or Asia. Finally, our focus on annual reports may overlook other channels of communication, such as earnings calls or social media, where managers might also signal their technological capabilities. Despite these constraints, this research highlights that the reporting of technological intangibles is a complex equilibrium shaped by the interplay of asset tangibility, sector dynamics, and the legal architecture of the market.

Acknowledgement

The authors declare that this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Furthermore, the authors confirm that there are no financial relationships or potential conflicts of interest to disclose in relation to this work.

References

Aboody, D., & Lev, B. (2000). Information asymmetry, R&D, and insider gains. The Journal of Finance, 55(6), 2747–2766.

Ash, E., & Hansen, S. (2024). Large language models in economics. In *Handbook of Experimental Economics* (Vol. 3). Cambridge University Press.

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

Barker, R., Penman, S., Linsmeier, T. J., & Cooper, S. (2022). Accounting for intangible assets: Suggested solutions. Accounting and Business Research, 52(6), 601–627.

Barth, M. E., Kasznik, R., & McNichols, M. F. (2001). Analyst coverage and intangible assets. *Journal of Accounting Research*, *39*(1), 1–34.

Ball, R., Kothari, S. P., & Robin, A. (2000). The effect of international institutional factors on properties of accounting earnings. Journal of Accounting and Economics, 29(1), 1–51.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). DIGITAL BUSINESS STRATEGY: TOWARD A NEXT GENERATION OF INSIGHTS (Vol. 37, Number 2).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. In *Journal of Machine Learning Research* (Vol. 3).

Bloom, N., Hassan, T. A., Kalyani, A., Lerner, J., & Tahoun, A. (2021). The diffusion of disruptive technologies. *NBER Working Paper Series*, No. 28999.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies." *American Economic Journal: Macroeconomics* 13 (1): 333–72**.**

Cao, Y., Liu, M., Qiu, J., & Zhao, R. (2024). *Determinants and Information Content of Voluntary Artificial Intelligence Disclosure*.
https://www.thecorporatecounsel.net/blog/2024/09/sec-chair-addresses-ai-washing-by-public-companies.html.

Chen, R., Ji, Y., Jiang, G., Xiao, H., Xie, R., & Zhu, P. (2022). Composite Index Construction with Expert Opinion. *Journal of Business and Economic Statistics*, *41*(1), 67–79. https://doi.org/10.1080/07350015.2021.2000418

Christensen, H. B., Hail, L., & Leuz, C. (2013). Mandatory IFRS reporting and changes in enforcement. *Journal of Accounting and Economics*, *56*(2–3), 147–177.

Corrado, C., Haskel, J., Jona-Lasinio, C., & Iommi, M. (2021). Artificial intelligence and productivity: An intangible assets approach. *Oxford Review of Economic Policy*, *37*(3), 435–458.

De Villiers, C., & Marques, A. (2016). Corporate social responsibility, country-level predispositions, and the consequences of choosing a level of disclosure. *Accounting and Business Research*, *46*(2), 167–195.

Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, liquidity, and the cost of capital. The Journal of Finance, 46(4), 1325–1359.

Dye, R. A. (2001). An evaluation of "essays on disclosure" and the disclosure literature in accounting. Journal of Accounting and Economics, 32(1–3), 181–235.

Matthew Gentzkow & Jesse Shapiro & Matt Taddy, 2016. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech," *Working Papers id:11114, eSocialSciences.*

Gu, F., & Li, J. Q. (2007). The credibility of voluntary disclosure and insider stock transactions. Journal of Accounting Research, 45(4), 771–810.

Haskel, J., & Westlake, S. (2018). *Capitalism without capital: The rise of the intangible economy*. Princeton University Press.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of Accounting and Economics, 31(1–3), 405–440.

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies*, *43*(9), 834–864.

Joint Research Centre. (2008). Handbook on constructing composite indicators: methodology and user guide. *OECD publishing.*

Khanna, T., Palepu, K. G., & Srinivasan, S. (2004). Disclosure practices of foreign companies interacting with U.S. markets. *Journal of Accounting Research*, *42*(2), 475–508.

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

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1998). Law and finance. *Journal of Political Economy*, *106*(6), 1113–1155.

Lewis, C., & Young, S. (2019). Fad or future? Automated analysis of financial text and its implications for corporate reporting. *Accounting and Business Research*, *49*(5), 587–615. https://doi.org/10.1080/00014788.2019.1611730

Lev, B. (2018). *Ending the accounting-for-intangibles status quo*. *European Accounting Review*, *28*(4), 713–736.

Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. *Journal of Financial Economics*, *69*(3), 505–527.

Li, Yinghua, Yupeng Lin, and Liandong Zhang. (2018) "Trade secrets law and corporate disclosure: Causal evidence on the proprietary cost hypothesis." Journal of Accounting Research 56.1, 265-308.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65.

Lys, T., Naughton, J. P., & Wang, C. (2015). Signaling through corporate accountability reporting. Journal of Accounting and Economics, 60(1), 56–72.

Merkley, K. J. (2014). Narrative disclosure and earnings performance. *The Accounting Review*, *89*(2), 725–757.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Passos, A., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. In *Journal of Machine Learning Research* (Vol. 12). http://scikit-learn.sourceforge.net.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. The Review of Financial Studies, 22(1), 435–480.
Ransbotham, S., Candelon, F., Kiron, D., Lafountain, B., & Khodabandeh, S. (2021). In collaboration with The Cultural Benefits of Artificial Intelligence in the Enterprise.
Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *ArXiv Preprint ArXiv:1908.10084*.
Rogers, J. L., & Van Buskirk, A. (2009). Shareholder litigation and changes in disclosure behavior. *Journal of Accounting and Economics*, *47*(1–2), 136–156.
Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, *5*, 179–194.
Wang, S. (2024). Data Asset Information Disclosure and Capital Market Efficiency: Empirical Research Based on Big Data Text Mining. In *Membrane Technology ISSN* (Vol. 327).
Zambon, Stefano, et al. "The Accounting Treatment of Unaccounted Intangibles: Exploring the Trade Offs of Preparers, Users and Professionals." Accounting in Europe (2025): 1-28.
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.
Wysocki, P. D. (2004). Discussion of ultimate ownership, income management, and legal and extra-legal institutions. *Journal of accounting research*, *42*(2), 463-474.