AI's Double-Edged Sword: Investment, Data, and the Risk of Default

Ziwei Ren, Xu Feng, and Yajun Xiao*

Abstract This paper examines how AI investment and data assets affect corporate credit risk. Using Chinese listed firms, we construct four complementary measures of AI investment, asset-based, labor-based, LLM-based, and text-based, and link them to firms' distance-to-default. We find that benchmark-level AI investment reduces default risk, while excessive firm-specific investment increases it by eroding profitability and reflecting risk-taking and competitive pressure. The dominance of this adverse effect yields a negative overall relation between AI investment and credit risk. Cash flow risk is the transmission channel: benchmark-level AI improves cash flow quality, whereas excessive investment worsens it. High-quality data assets complement benchmark-level AI by stabilizing cash flow, but this benefit fades once investment becomes excessive. Overall, the impact of AI on credit risk depends on both investment intensity and data quality, operating primarily through cash flow dynamics.

Keywords: Artificial intelligence; AI Investment; Benchmark and Firm-specific AI Investment; Data; Default Risk

---

*Ziwei Ren, College of Management and Economics, Tianjin University, E-mail: renziwei@tju.edu.cn. Xu Feng, College of Management and Economics and China Center for Social Computing and Analytics, Tianjin University, E-mail: fengxu@tju.edu.cn; Yajun Xiao, Xi'an Jiaotong-Liverpool Unveristy, E-mail: yajun.xiao@xjtl.edu.cn. We are grateful for helpful comments and suggestions from Dan Luo, Iichiro Uesugi (Discussant), Ruichen Ma (Discussant), Nan Li (Discussant), Yuting Li (Discussant), and seminar participants at Harbin Institute of Technology, Nanjing University, Nanjing Audit University, ShanghaiTech University, Xi'an Jiaotong University, Southwestern University of Finance and Economics, and the 2024 Sydney Banking and Financial Stability Conference, 2024 19th International Conference of Financial System Engineering and Risk Management, 30th International Conference Computing in Economics and Finance, the 16th National Academic Conference on Youth Management Science and Systems Science, the 15th International Annual Conference of the Chinese Scholars Association for Management Science and Engineering, 2024 Greater China Area Finance Conference, 2024 3rd Annual International Finance Conference, the 21st Chinese Finance Annual Meeting, 2021 Future Finance and Economics Conference. We acknowledge financial support from the National Natural Science Foundation of China (No. 72342022 and No.72141304)

1 Introduction

Firms are investing in Artificial Intelligence (AI) technologies at an unprecedented pace, with global corporate investment surging from $14.6 billion in 2013 to $189.6 billion in 2022.[1] While AI offers significant promise, investment is uncertain, irreversible, and costly, making project failure likely.[2] Moreover, its success critically depends on access to data availability (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2021; Veldkamp and Chung, 2024). The net benefits of AI investment therefore remain unclear. Existing research largely highlights its advantages—improved decision-making, enhanced investment strategies, and productivity gains—yet little is known about its credit risk implications.[3] Does AI investment increase or reduce default likelihood? Is current investment close to its optimal benchmark? Do data assets mitigate these risks? This paper develops novel firm-level measures of AI investment to examine its effect on corporate credit risk and address these questions.

We investigate this question in the context of China for two main reasons. First, China has emerged as a leading competitor in the global AI innovation race, alongside the European Union and the United States (Beraja et al., 2023). Second, the richness and access of diverse data sources in China that enable us to construct four complementary measures of AI investment—some based on directly observed data and others derived through proxies—allowing for cross-validation across measures. As noted by Seamans and Raj (2018), a key challenge in AI research is the lack of firm-level data on AI investment, which is crucial for assessing its impact on firm productivity and informing policy decisions on AI adoption.

[1]See the 2022 AI index report at the Stanford Institute for Human-Centered Artificial Intelligence (HAI) available at https://aiindex.stanford.edu/ai-index-report-2022/.

[2]Westerman et al. (2014) document failed digital transformations at major U.S. corporations, and Accenture's 2019 Digital Transformation Index reports that only 9% of Chinese firms achieved significant positive outcomes.

[3]Recent research emphasizes AI's benefits. At the aggregate level, AI investment boosts economic growth (Acemoglu and Restrepo, 2018; Aghion et al., 2017; Graetz and Michaels, 2018; Seamans and Raj, 2018), with a 0.71% rise in total factor productivity over ten years (Acemoglu, 2025). At the firm level, AI adoption enhances growth and productivity (Alderucci et al., 2020), sales (Rock, 2019), product innovation (Cockburn et al., 2018; Babina et al., 2023, 2024), product quality (Fedyk et al., 2022), and task-specific efficiency (Eloundou et al., 2023; Brynjolfsson et al., 2025). Risks, however, remain underexplored, as AI increases systemic exposure (Babina et al., 2023)

Specifically, we develop four annual measures of AI investment at the firm level. The first measure, drawn from the China Stock Market & Accounting Research (CSMAR) database, is defined as the logarithm of one plus the ratio of AI investment to total sales. CSMAR identifies AI-related assets disclosed in firms' annual financial reports, which we treat as AI investment in that year. We refer to this as the **asset-based AI investment**. Because disclosure of AI-specific investments is not mandatory, this measure is incomplete and missing for certain firms and years.

The second measure, following Chen and Srinivasan (2023), applies textual analysis of annual reports by constructing a dictionary of AI-related keywords and recording their frequency, with higher frequencies indicating stronger AI engagement. This dictionary-based method is transparent and replicable but limited in its ability to capture semantic depth and contextual nuances in firms' narratives. To address this, we adopt a third measure, inspired by Jha et al. (2024), which leverages a Large Language Model (LLM) to analyze annual reports and generate scores that proxy the extent of firms' AI investment. We refer to these as the **text-based** and **LLM-based AI investment** measures, respectively

The fourth measure, following Babina et al. (2023, 2024), uses job posting data from 51job.com, a leading recruitment platform in China with comprehensive listings. It captures the intensity of AI-related human capital, defined as the ratio of AI-related hires to total hires in a given year, and is referred to as the **labor-based AI investment**. This measure has two limitations: it spans only a shorter sample period (2014–2022) and, lacking employee resume data, cannot incorporate information from the labor supply side.

We link firms' AI investment measures to their one-year-ahead distance-to-default (DD) (Merton, 1974), a standard proxy for credit risk. Illustratively, plotting AI investment against DD (Figure 1) reveals a downward-sloping relationship, indicating that higher AI investment is associated with greater default risk. This finding is reinforced by panel regressions of DD on AI investment measures, controlling for firm characteristics and fixed effects, where the negative relationship remains both statistically and economically significant. The evidence suggests that, at the firm level, the costs of AI investment can outweigh the benefits. This stands in contrast to prior research emphasizing AI's positive effects on aggregate economic growth (Aghion et al., 2017; Seamans and Raj, 2018; Acemoglu, 2025) and on firm growth and valuation (Alderucci et al., 2020; Babina et al., 2024).

We reconcile these differences through a conceptual framework that parameterizes the Goldstein et al. (2001) default risk model with AI investment and data affecting cash flow directly, which serves to guide our empirical analysis. Because firms hold an "option to wait" but incur substantial sunk costs once they commit (Dixit and Pindyck, 1994), they often calibrate AI spending against market or industry benchmarks that signal when expected productivity gains outweigh risks. Consequently, we decompose AI investment into two components: a benchmark level and a deviation from the benchmark. The deviation can be positive or negative, with larger positive values indicating greater investment beyond the benchmark. This deviation captures potential managerial risk-taking behavior, for which we provide empirical evidence. At the benchmark level, AI investment boosts productivity, enhances profitability, and reduces default risk. By contrast, excessive investment beyond the benchmark erodes profitability and raises default risk. AI technologies such as algorithms and robotics rely on abundant, high-quality data to be effective. Data can smooth productivity shocks and stabilize profitability, while poor data quality may instead amplify volatility (Veldkamp and Chung (2024)).

We treat AI investment and data as exogenous factors influencing a firm' s cash flow, expected growth, and volatility, adapting the framework of Goldstein et al. (2001). The signs of the first-order derivatives of these variables with respect to AI investment —evaluated at both the benchmark level and deviations from it—reveal how cash flows respond to changes in AI investment and whether cash flow risk improves or deteriorates. For example, a positive (negative) derivative of cash flow level with respect to benchmark investment suggests that AI investment enhances (diminishes) cash flows and thereby reduces (increases) credit risk. These derivative-based predictions form the basis of our empirical tests.

Data generate value only when AI investment is undertaken, and their joint effect can be either complementary or substitutive. Under complementarity, data reinforce the benefits of investment, whereas under substitution, they offset them. Consequently, the cross-derivatives of cash flow level, expected growth, and volatility with respect to benchmark investment and data, as well as deviations from the benchmark and data, determine how cash flows—and thus credit risk—respond to AI investment conditional on data. For instance, a positive (negative) cross-derivative of cash flow level with respect to benchmark investment and data indicates complementarity (substitution), implying that data increase (decrease) cash flow levels and thereby reduce (amplify) credit risk.

Our empirical analysis then examines how a firm's AI investment—both its alignment with the benchmark and its deviations from it—affects credit risk. For each of the four measures, we decompose firm-level AI investment into a market-driven component and a firm-specific component. To construct the market-driven component, we regress firm AI investment on the market average within a panel framework that includes fixed effects and firm controls. The fitted value, obtained as the coefficient times the market average, captures market-driven AI investment. The residual represents firm-specific AI investment, reflecting deviations from the benchmark.[4] Such deviations may capture unique innovation but also involve a greater risk of failure (Brynjolfsson et al., 2019; Chen and Srinivasan, 2023). Importantly, the market-driven (benchmark) and firm-specific (deviation) components are orthogonal by construction.

Our panel analysis, which regresses one-year-ahead DD on market-driven and firm-specific AI investment, shows that market-driven investment is positively associated with DD (lower credit risk), while firm-specific investment has a significant negative effect (higher credit risk). This suggests that AI's impact on credit risk depends on whether firms invest up to the market benchmark, where productivity gains outweigh risks. Large deviations from the benchmark can turn the net benefits of AI investment negative—a pattern evident in China, where we observe a downward-sloping relationship between

[4] As an alternative, we define the market-driven component as the industry mean of AI investment and the firm-specific component as the firm's AI investment minus the industry mean. This decomposition is used as a robustness check.

AI investment and credit risk in Figure 1.[5] This finding is robust across endogeneity tests, alternative measures of market-driven and firm-specific AI investment, and alternative credit risk measures. Moreover, using a five-year rolling window to estimate the coefficients, we find that the effects remain stable. For example, regressions of DD on firm-specific AI investment consistently yield small, negative coefficients that remain flat over time, suggesting that firms persistently fail to invest at the optimal level. Prior research emphasizing the benefits of AI technologies has largely overlooked the importance of investment intensity (Cockburn et al., 2018; Rock, 2019; Alderucci et al., 2020; Fedyk et al., 2022; Babina et al., 2024).

We now turn to the relationship between AI investment and data. While much attention has been devoted to the economics of data (Veldkamp and Chung, 2024), less is known about whether investments in AI technologies, such as machine-learning algorithms, align with the data required for their training and implementation. Following Beraja et al. (2023), we consider that firms may gain access to valuable government data by providing services to the state, and such collaboration could increase the success rate of their AI projects. Using textual analysis of all Chinese government procurement contracts, we construct a data dummy that equals one if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. We then examine the interaction between this data dummy and both market-driven and firm-specific AI investment. Our panel regressions show that the coefficient of DD on the interaction between market-driven AI investment and the data dummy is positive and significant, whereas the interaction between firm-specific AI investment and the data dummy is insignificant. These findings suggest that abundant, high-quality data complements AI investment when it is close to the optimal benchmark, but this complementarity disappears once investment exceeds the benchmark. The results remain robust when we use broader measures of data, such as knowledge capital (Ewens et al., 2024) and R&D stock (Aghion et al., 2013).

Our empirical analysis of credit risk builds on the assumption that AI investment and

[5]Using text- based and LLM-based measures derived from U.S. company 10-Ks, we document a similar negative relationship for U.S. firms. This does not necessarily contradict Babina et al. (2024), who study the cross-sectional impact of AI investment on firm growth, whereas our analysis captures within-firm variation.

data affect firm cash flows mechanically within the conceptual framework of Goldstein et al. (2001). To assess this assumption, we use two proxies for cash flows: cash flow scaled by assets and return on assets (ROA). We then estimate panel regressions of cash flow level, expected growth, and volatility on the variables of interest. The results support the assumption. Specifically, market-driven AI investment is positively associated with higher cash flow levels, stronger expected growth, and lower volatility. By contrast, firm-specific AI investment is negatively associated with both cash flow level and expected growth, and shows no significant effect on volatility. These findings are consistent with our assumption that market-driven AI investment promotes more stable and resilient cash flows, thereby lowering default risk, whereas firm-specific AI investment weakens cash flows and increases default risk.

Regarding the interaction between AI investment and data, we find that data complements market-driven AI investment by reducing cash flow volatility, while having no effect on cash flow level or growth. In contrast, its interaction with firm-specific AI investment shows no significant impact on cash flow risk. These results suggest that data primarily stabilizes volatility rather than enhancing cash flow performance, consistent with models in (Farboodi and Veldkamp, 2021; Veldkamp and Chung, 2024) that assume data smooths demand shocks. However, when firms invest beyond the market benchmark, additional high-quality data fails to complement such investment and does not mitigate risk.

Understanding why some firms invest beyond the benchmark is crucial. We identify two key characteristics: first, risk-tolerant firms are more likely to favor innovative, firm-specific AI technologies, expecting them to confer a competitive edge; second, firms in highly competitive industries tend to adopt more firm-specific AI to outperform rivals and manage competitive pressures. For policymakers, these findings underscore the importance of guiding AI investment toward optimal levels to support, rather than undermine, financial stability.

Our study contributes to the growing literature on AI in several important ways. At the firm level, prior research has documented a wide range of benefits from AI adoption, including enhanced growth and productivity (Alderucci et al., 2020), increased sales (Rock, 2019), improved product innovation (Cockburn et al., 2018; Babina et al., 2024), higher product quality (Fedyk et al., 2022), firm growth and market valuation (Chen and Srinivasan, 2023), and greater task-specific efficiency (Eloundou et al., 2023; Brynjolfsson et al., 2025). Despite these advances, the risks associated with AI adoption remain comparatively underexplored. A notable exception is Babina et al. (2023), who show that AI increases firms' systemic risk exposure in the cross-section, thereby raising questions about the trade-off between efficiency gains and heightened vulnerabilities at both the firm and market level.

Our paper differs from Babina et al. (2023) by focusing on the downside risk of corporate creditworthiness and examining within-firm variation rather than cross-sectional patterns. Specifically, we make two contributions. First, we show that the effect of AI on risk critically depends on the investment level. When firms invest at the optimal market benchmark, AI adoption enhances cash flow and reduces credit risk, consistent both with the positive cross-sectional relationship between AI and firm growth documented by Babina et al. (2024) and with the broader literature highlighting the benefits of AI adoption. Second, we demonstrate that excessive investment beyond the optimal benchmark reverses these gains: cash flow declines, credit risk increases, and the net effect of AI turns negative. This finding suggests that overinvestment in firm-specific AI amplifies firms' idiosyncratic risk exposure.

Our paper also contributes to the growing literature on the economics of data. Jones and Tonetti (2020) were among the first to highlight the role of data in fostering innovation, modeling it as an input into endogenous growth.[6] A related strand of research, pioneered by Farboodi et al. (2019) and further developed by Farboodi and Veldkamp (2021), views data as a valuable byproduct of firm activities that reduces uncertainty about optimal production techniques and thereby enhances productivity. Building on this framework, Mihet et al. (2025) propose a dynamic model in which firms transform raw data into knowledge through AI but face informational entropy: without adequate

[6] Veldkamp and Chung (2024) provides an excellent review of the macroeconomics of data.

AI investment, accumulating more data leads to information overload and diminishing returns. In line with this perspective, our study investigates whether data complement physical AI investment in shaping corporate credit risk. We show that such complementarity is conditional: data help stabilize cash flow and mitigate risk when AI investment is at the optimal benchmark, but these benefits disappear—and may even reverse—when investment deviates substantially from the benchmark.

Our paper also responds to the call by Seamans and Raj (2018) for systematic collection of firm-level AI investment measures to better understand its micro-level consequences. Existing work has made important progress along these lines. For example, Chen and Srinivasan (2023) construct a text-based measure from 10-K reports to capture the extent of digital activity among non-tech U.S. firms, while Babina et al. (2023, 2024) use job posting data to proxy for firms' demand for AI-related human capital. Building on these approaches, we construct three complementary measures of AI investment: a text-based measure following Chen and Srinivasan (2023), a labor-based measure following Babina et al. (2023, 2024), and an LLM-based measure following Jha et al. (2024). Employing multiple measures enables cross-validation and strengthens the robustness of our results.

Finally, our paper contributes to the literature on agency problems, where managers pursue growth or projects beyond what maximizes shareholder value (Jensen, 1986, 1993), sometimes resulting in empire-building behavior. In such cases, free cash flow may be allocated to low-return or high-risk projects, increasing the firm's exposure to operational and financial uncertainty. We show that firm-specific AI investment—i.e., AI investment exceeding the optimal benchmark—deteriorates cash flow and elevates credit risk. These findings align with evidence that idiosyncratic and firm-specific risks significantly contribute to credit risk in the bond market (Campbell and Taksler, 2003; Chen et al., 2009; Bao et al., 2011). In this context, our results provide insights for policymakers regarding the regulation and monitoring of AI investment.

The remainder of this paper is structured as follows: Section 2 reviews the conceptual framework and presents our hypotheses. Section 3 outlines the data and variables employed in our study. Section 4 presents baseline results, demonstrating a negative association between AI investment and firm default risk. Section 5 explores the pivotal role of data in the mechanism of influence. Section 6 characterizes firms that extensively utilize firm-specific AI. Section 7 concludes the paper.

2 Conceptual Framework and Hypothesis

In this section, we examine the cash flow channels through which AI investment and data affect default risk. The stylized framework is descriptive, grounded in plausible assumptions, and forms the basis for our hypotheses.

Firms are assumed to operate with an AK technology, where capital and labor are fixed and normalized to one. This eliminates scale effects, allowing us to focus exclusively on technology shocks. Accordingly, output is given by

$$Y_t = A_t.$$

Productivity evolves as a geometric Brownian motion under the risk-neutral measure:

$$dA_t = gA_t\,dt + \sigma A_t\,dW_t^Q,$$

which admits a closed-form solution $A_t = A_0 e^{(g-0.5\sigma^2)t+\sigma W_t^Q)}$, where A_0 denotes the initial productivity, g the risk-neutral growth rate, and σ the volatility, all assumed deterministic and independent from time. We further assume the existence of a stochastic discount factor such that the risk-adjusted discount rate is constant at k. Upon installing new investment, cash flow is adjusted upward proportionally to its current level. We assume that AI investment affects not only the cash flow level but also its expected growth rate and volatility. These adjustments are modeled as deterministic functions of AI investment and data, which we specify in detail below.

Our treatment proceeds in three steps. First, suppose the firm possesses and exercises a growth option at time $t = 0$ by investing in AI—an investment that is costly, irreversible, and uncertain, yet potentially productive. Building on established micro-foundations, we identify three channels through which AI investment and data affect cash flows. Immediately upon exercise, the initial cash flow is scaled by a factor $\Pi_1 > 0$, becoming $\Pi_1 A_0$; it increases if $\Pi_1 \geq 1$ and decreases if $0 < \Pi_1 < 1$. Theoretical dynamic capital structure models commonly apply such scaling to preserve homogeneity of the cash flow process and maintain tractability (Hackbarth and Mauer, 2012). Second, the expected growth rate adjusts to $\Pi_2 g$, where $\Pi_2 \geq 1$ raises growth and $0 < \Pi_2 < 1$ lowers it. Third, cash flow volatility changes to $\Pi_3 \sigma$, with $\Pi_3 \geq 1$ amplifying volatility and $0 < \Pi_3 < 1$ dampening it.

Second, it remains an open question whether data complement or substitute AI technologies, although recent studies emphasize their complementarities. Farboodi and Veldkamp (2021) show that AI investment enables firms to collect data as a function of output Y_t, which can then be used to smooth technology shocks (i.e., cash flow shocks). However, if mismanaged, data may amplify shocks in the absence of sufficient AI investment, as illustrated by Mihet et al. (2025). Inspired by these insights, we assume that data, conditional on AI investment, can alter cash flows—either enhancing them (higher level and growth rate, lower volatility) or deteriorating them (the opposite effect). Formally, this implies that the adjustment factors Π_i, $i = 1, 2, 3$, depend jointly on AI investment and data, providing a mechanism through which the two shape the variability of firm cash flows.

Third, we assume that the adjustment factors Π_i $(i = 1, 2, 3)$ explicitly depend on AI investment, denoted by I, and data, denoted by D. Firms are assumed to anchor their AI investment around a benchmark $\bar{I}$—the level at which expected benefits optimally balance the associated risks. In the context of investment under uncertainty, it is well established that market/industry averages can serve as benchmarks, embedding aggregated information on technology adoption, cost structures, and expected returns, and thereby guiding individual firms' choices. Nonetheless, due to agency conflicts (Jensen, 1986, 1993), actual investment may deviate from this benchmark by an amount Δ, reflecting managerial preferences that diverge from the risk–return optimum. We thus specify the dependence of the adjustment factors as

$$\Pi_i := \Pi_i(\bar{I}, \Delta, D).$$

Putting these elements together, benchmark-level AI investment is efficient in the sense that the adjustment factor Π_1 or Π_2 increases with $\bar{I}$, or Π_3 decreases with $\bar{I}$, i.e.,

$$\frac{\partial \Pi_1}{\partial \bar{I}} > 0, \quad \frac{\partial \Pi_2}{\partial \bar{I}} > 0, \quad \text{or} \quad \frac{\partial \Pi_3}{\partial \bar{I}} < 0.$$

Economically, this implies that benchmark-level AI investment improves cash-flow quality: higher investment raises the level and growth rate of cash flows while dampening volatility.[7] Enhanced cash-flow quality, in turn, reduces firms' vulnerability to default risk.[8] By contrast, deviation from the benchmark, Δ, is inefficient in that Π_1 or Π_2 decreases with Δ, or Π_3 increases with Δ:

$$\frac{\partial \Pi_1}{\partial \Delta} < 0, \quad \frac{\partial \Pi_2}{\partial \Delta} < 0, \quad \text{or} \quad \frac{\partial \Pi_3}{\partial \Delta} > 0.$$

Thus, deviation investment deteriorates cash-flow quality and elevates default risk. The overall impact of AI investment I on firm credit risk is therefore determined by the interplay between these opposing forces.

These considerations guide our empirical strategy. For example, when $\frac{\partial \Pi_1}{\partial \bar{I}} > 0$, higher benchmark-level AI investment $\bar{I}$ is expected to raise cash flow levels and reduce the likelihood of default, holding the initial cash flow A_0 constant. This motivates the regression

$$\text{CashFlowLevel}_{t+1} = \beta_0 + \beta_1 \bar{I}_t + \textbf{Controls}_t + \textbf{FEs} + \varepsilon_{t+1}.$$

$\beta_1 > 0$ aligns with the proposed mechanism. Analogous regressions apply for deviation Δ and for the expected growth rate of cash flows. Similarly, the prediction that benchmark

[7] For each case, the adjustment factor must be restricted to take admissible values. For our purpose, it suffices to consider the signs of these derivatives.

[8] A deterioration in cash flow quality—reflected in lower levels and growth rates, and higher volatility—increases the likelihood of default, consistent with Goldstein et al. (2001).

investment reduces volatility ($\frac{\partial \Pi_3}{\partial \bar{I}} < 0$) motivates

$$\text{CashFlowVolatility}_{t+1} = \beta_0 + \beta_1 \bar{I}_t + \textbf{Controls}_t + \textbf{FEs} + \varepsilon_{t+1},$$

where a negative β_1 would confirm the volatility-dampening role of AI investment.

Conditional on AI investment, the role of data is captured by the signs of the second-order cross derivatives of Π_i with respect to $(\bar{I}, D)$ and (Δ, D). These cross derivatives determine whether data complement or substitute AI investment. For example, for the cash-flow level:

$$\frac{\partial^2 \Pi_1}{\partial \bar{I}\, \partial D} = \begin{cases} > 0, & \text{complementarity,} \\ = 0, & \text{no effect,} \\ < 0, & \text{substitution,} \end{cases} \qquad \frac{\partial^2 \Pi_1}{\partial \Delta\, \partial D} = \begin{cases} > 0, & \text{complementarity,} \\ = 0, & \text{no effect,} \\ < 0, & \text{substitution.} \end{cases}$$

The remaining cases for the impact of AI investment and data on expected growth rate and volatility follow analogously. These considerations yield the following empirical design. For benchmark investment:

$$\text{CashFlowLevel}_{t+1} = \beta_0 + \beta_1 \bar{I}_t + \beta_2 (\bar{I}_t \cdot D_t) + \textbf{Controls}_t + \textbf{FEs} + \varepsilon_{t+1}.$$

$\beta_1 > 0$ reflects $\frac{\partial \Pi_1}{\partial \bar{I}} > 0$ and $\beta_2 > 0$ indicates complementarity between AI investment and data ($\frac{\partial^2 \Pi_1}{\partial \bar{I} \partial D} > 0$). For deviation investment:

$$\text{CashFlowLevel}_{t+1} = \beta_0 + \beta_1 \Delta_t + \beta_2 (\Delta_t \cdot D_t) + \textbf{Controls}_t + \textbf{FEs} + \varepsilon_{t+1},$$

where $\beta_1 < 0$ is consistent with $\frac{\partial \Pi_1}{\partial \Delta} < 0$, and $\beta_2 < 0$ suggests that data substitute for deviation investment, amplifying cash-flow risk ($\frac{\partial^2 \Pi_1}{\partial \Delta \partial D} < 0$). It is plausible that β_2 equals zero, indicating that data neither complements nor substitutes AI investment.

Default occurs when a firm' s cash flow falls to a critical threshold Goldstein et al. (2001). Since cash flow follows a Geometric Brownian motion, default risk rises when the initial asset value A_0 or the expected growth rate g declines, or when volatility σ increases. This underscores that the effect of AI investment and data quality on default risk operates through these three determinants. Our approach explicitly describes the net effect of the trade-off between the costs of AI investment and the potential gains it generates, in conjunction with data, on cash flow. Intuitively, investing in AI and related data is beneficial when expected gains outweigh costs, and may be detrimental when costs dominate—a question we leave for future theoretical investigation.

3 Data and Variable Construction

3.1 Data

The sample used in our study comprises A-share listed firms in China during the 2006-2023 period. From the CSMAR database, we collect firm-level financial information, including firm size, listing age, Tobin's Q, debt-to-asset ratio, asset growth, earnings per share, return on assets (ROA), and cash flow. We also obtain current liabilities, non-current liabilities, market value, stock prices, and the one-year government bond rate to calculate distance to default (DD).

In addition, we collect data on AI-related intangible and fixed assets from CSMAR for 2010–2023 to construct an asset-based measure of firms' AI investment. We also obtain annual reports from the WIND database and apply textual analysis together with large language models (LLMs) to develop two complementary measures: a text-based measure and an LLM-based measure.

Finally, we exploit a proprietary dataset of corporate job postings from 51job.com, covering approximately 68 million postings between 2014 and 2022.[9] This unique dataset provides detailed information at the job posting level, such as company name, posting

[9]51job.com is one of China's leading online recruitment platforms, widely recognized for its comprehensive and representative job listings. With millions of postings from companies across various industries, it serves as a trusted resource for both job seekers and employers. According to the information published on their website, 51job.com has over 100 million registered users and an enormous database of 96 million resumes. The platform's extensive reach and credibility make it a reliable source of data on corporate hiring trends, and its job postings are highly regarded by professionals as reflecting almost real opportunities in the Chinese labor market (He et al., 2023).

date, job title, job function, job description, and number of recruits, which enables us to capture AI investment from a human capital perspective.

3.2 AI Investment Measures

3.2.1 Measurement of AI Investment

To capture firms' heterogeneous engagement with AI, we construct four complementary measures of AI investment at the firm-year level. The first measure relies on firms' disclosed AI-related assets in financial statements, thereby reflecting balance-sheet commitments. The second measure applies a dictionary-based textual analysis to annual reports, offering a transparent benchmark of AI-related disclosure. The third measure leverages a large language model (LLM) to evaluate the semantic content of annual reports and extract contextual signals of AI investment. Finally, the fourth measure exploits detailed job posting data to quantify firms' demand for AI-related human capital. Together, these four measures allow us to triangulate firms' AI investment from disclosure, assets, and labor perspectives, while mitigating the limitations of any single proxy. The details are as follows.

Asset-based AI investment: Our first measure of firm-level AI investment is derived from the CSMAR database, which applies textual analysis to firms' financial statement notes to identify AI-related intangible and fixed assets ranging from 2010 to 2023. Intangible assets are classified as AI-related when their descriptions include terms such as AI, software, intelligence, information platforms, systems, data, service platforms, internet, cloud computing, information technology, 5G, Internet of Things, and blockchain. Fixed assets are classified as AI-related when they include items such as computers, electronic equipment, data equipment, automated equipment, information equipment, servers, intelligent terminals, communication equipment, integrated equipment, storage equipment, computing power, CPUs, and networks. We then aggregate the value of these AI-related intangible and fixed assets, scale the sum by firm sales, and take the natural logarithm to construct an asset-based measure of AI investment, AI_{Asset}. This approach directly captures firms' balance sheet commitments to AI-related resources but also faces limitations: because disclosure of AI-specific asset items is not required under accounting standards, the measure is missing for many firm-year observations and may suffer from underreporting bias.[10]

Text-based AI investment: Following Chen and Srinivasan (2023), we then construct a text-based measure of AI investment by applying a dictionary-based approach to annual financial reports ranging from 2006 to 2023. We build an AI dictionary based on Chen and Srinivasan (2023) and the report of "WIPO Technology Trends –Artificial Intelligence".[11] Chen and Srinivasan (2023) extract digital-related words from firms' annual reports to proxy the level of digital transformation. World Intellectual Property Organization (WIPO) also provides selected AI categories and terms in their technology trends report. We manually select AI terms from these two sources to build the AI dictionary (Table A2).

For each firm-year observation, we count the total frequency of AI-related terms in the firm's annual report and define our measure, AI_{Text}, as the natural logarithm of this frequency.[12] While this dictionary-based approach offers transparency and replicability, it remains limited in its ability to capture the semantic depth and contextual nuances of firms' narratives. This limitation motivates our complementary LLM-based measure, which can detect more subtle forms of AI engagement beyond explicit terminology.

LLM-based AI investment: Our third measure adapts the approach of Jha et al. (2024) and employs an open-source Large Language Model (Qwen2.5) to process firms' annual financial reports.[13] Specifically, each firm's annual report is divided into text seg-

[10] Approximately 10% of firm-year observations lack relevant data. For these cases, we impute missing values with zeros. This approach is justified on two grounds. First, it is consistent with economic intuition: firms that do not disclose AI-related assets are unlikely to have incurred material AI investments during the reporting period. Second, it mitigates potential data biases and sample selection issues that would arise from dropping such observations, thereby preserving sample completeness and enhancing the robustness of our empirical results.

[11] The report can be found at: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

[12] To prevent the interference of firm and organization names or titles in the word frequency statistics (e.g., "China Academy of Automation", "China Artificial Intelligence Center", "Artificial Intelligence Expert"), we set the following deletion rule in the textual analysis. First, we use text stop codes, such as ".", "!", "…", and "?", to identify the beginning and ending of each sentence. Second, AI-related terms are not considered if name- or title-related words, such as "company", "group", "research center", "laboratory", "engineer", and "expert" appear within a 10-word range around the AI terms.

[13] We adapt Qwen2.5 because it is among the most advanced models for processing Chinese-language

ments of approximately 1,500 characters. The segmentation into 1,500-character chunks is motivated by the input and output token limitations of large language models: such a length is sufficiently detailed to capture meaningful information while ensuring reliable and efficient model output. For each segment, we use the following prompt:

The following text is excerpted from a listed company's annual report. You are a financial expert. Based solely on this text, please answer the following:

(1) AI investment score assessment: Evaluate the extent of the company's actual AI investment. There are five options: Very High, High, Medium, Low, and None. Please select one of the five.

(2) Confidence score: For your choice in (1), provide a confidence score from 1 to 5, where a higher number indicates greater certainty in your assessment. (3) Explanation: Provide a one-sentence justification for your answers to (1) and (2).

Your output must strictly follow this format: AI Investment Score: [result] - Confidence Score: [result] - Explanation: [sentence].

We then aggregate the LLM outputs across all text segments within a firm. Our final measure, AI_{LLM}, is defined as the natural logarithm of the sum of AI investment scores across segments, weighted by their corresponding confidence scores. Compared with traditional keyword-based or manual annotation approaches, the LLM-based approach enables a more accurate interpretation of textual semantics, effectively distinguishing genuine AI investment activities from generalized expressions. At the same time, this method preserves efficiency and scalability under large-scale corpora, thereby offering a more comprehensive, precise, and effective measure of firms' AI investment.

Labor-based AI investment: Our final measure follows Babina et al. (2023, 2024) and exploits job posting data from 51job.com, a leading online recruitment platform in China. We first identify AI-related jobs based on their description only if it contains AI-

texts, and using an open-source model also helps to control computational costs. When using LLM, we focus on the Management Discussion and Analysis (MD&A) sections of corporate annual reports from 2010 to 2023. MD&A provides management's detailed discussion of the firm's operating performance and investment activities for the year, thereby offering a more direct source of information on AI-related investments. We set 2010 as the starting point of the sample period because, prior to that, the MD&A sections lacked standardized formats and exhibited substantial heterogeneity in disclosure practices, which could undermine the accuracy of our measurement.

related keywords.[14] Following this, we calculate the total number of recruits of AI-related job and other jobs for each company. For each publicly listed company, we aggregate the number recruits of AI-related job and other jobs from both the parent company and its subsidiaries. Finally, we compute the ratio of the number recruits of AI-related jobs to the number recruits of total jobs at the firm-year level. We define this proportion as the variable AI_{Labor}. This measure, termed labor-based AI investment, reflects firms' demand for AI-related human capital. While this proxy provides valuable information on firms' labor allocation toward AI, it is restricted to a shorter sample period (2014-2022) and does not incorporate data from the labor supply side (e.g., resumes), potentially omitting important dimensions of AI human capital.

Table A3 presents the pairwise correlations among the four firm-level measures of AI investment. Panel A reports the pooled correlation coefficients across the full sample. All four measures are significantly positively correlated, indicating that each reflects firms' engagement in AI to some extent. However, the correlations are below one, suggesting that these measures capture distinct and complementary dimensions of AI investment.

Panel B of Table A3 shows the mean of firm-level correlation coefficients of four measures, while Panel C displays the median of firm-level correlation coefficients. All measures are positively and significantly correlated, suggesting that they capture a common underlying dimension of AI investment, yet the moderate magnitude of the coefficients (generally 0.3–0.6) indicates that they also provide complementary information. Among them, AI_{Text} and AI_{LLM} exhibit the strongest correlation, confirming the validity of LLM-based measures as they align closely with traditional text-based methods. AI_{Asset} is moderately correlated with AI_{Text} and AI_{LLM}, highlighting potential discrepancies between firms' reported AI investments in financial statements and their narrative disclosures, possibly due to differences in disclosure practices or AI washing. By contrast, AI_{Labor} shows the weakest correlations with other proxies, underscoring that recruitment captures a distinct human-capital dimension of AI investment. Comparing the median and mean values, we find that the medians were consistently higher than the means.

[14] AI-related keywords are from Chen and Srinivasan (2023) and the report of "WIPO Technology Trends –Artificial Intelligence".

This indicates a right-skewed distribution, revealing substantial heterogeneity in AI investment.

In conclusion, these results confirm that our four measures are both overlapping and complementary, justifying our use of a multidimensional approach to comprehensively capture firms' AI investment.

3.2.2 The Decomposition of AI Investment

Building on the concept of "peer effects" in corporate investment (Dougal et al., 2015; Fracassi, 2017), we decompose a firm's AI investment into two components: a market-driven component, aligned with aggregated peer behavior and reflecting industry trends, and a firm-specific component, capturing the innovation-driven portion that deviates from these trends. This allows us to distinguish reactive from proactive AI investment.

We proxy market-wide AI investment using the annual average across all firms and estimate the following panel regression:

$$AI_{i,t} = \alpha + \beta\, Market_t + \sum Controls_{i,t} + \xi_i + \varepsilon_{i,t}, \tag{1}$$

where $AI_{i,t}$ denotes total AI investment for firm i in year t, including $AI_{Asset,i,t}$, $AI_{Text,i,t}$, $AI_{LLM,i,t}$, and $AI_{Labor,i,t}$; $Market_t$ is the logarithm of market-wide AI investment; $Controls_{i,t}$ include size, ROA, age, Tobin's Q, debt-to-asset ratio, asset growth, and earnings per share; and ξ_i represents firm fixed effects. Year fixed effects are omitted since the market average is identical across firms.

Using the estimated coefficients, we decompose each firm's AI investment into the market-driven component, $\hat{\beta}_i\, Market_t$, and the firm-specific component, $\hat{\varepsilon}_{i,t}$, representing deviations from the market average. The orthogonality of these components simplifies subsequent analyses. For the main analysis, we use standardized values of all three types of investment: total, market-driven, and firm-specific AI investment.

In our robustness checks, we apply two alternative methods to decompose each firm's four AI investment measures. First, we replace the market-wide trend with the industry-average AI investment and apply the same regression-based decomposition. Second, we define the market-driven component as the industry mean AI investment, with the firm-specific component equal to the difference between total investment and this mean.

3.3 Credit Risk Measures

In our primary analysis, we measure default risk using the distance to default (DD), a metric derived from the Merton (1974) model. A comprehensive explanation of this measure and its estimation is included in the Appendix. To ensure the robustness of our findings, we also employ alternative proxies for credit risk: the Z-score from Altman (1968) and the negative of the O-score from Ohlson (1980).

3.4 Controls

Following the literature (Campbell et al., 2008; Eberhart et al., 2008; Brogaard et al., 2017; Chen and Srinivasan, 2023), we control for factors that may affect a firm's default risk, including size (*Size*), profitability (ROA), firm value ($TobinQ$), capital structure (*Leverage*), and total asset growth rate.

3.5 Summary Statistics

Table 1 presents the summary statistics for the main variables, with detailed definitions provided in Table A1 in the Appendix. The key dependent variable of interest is distance to default (DD). The logarithmic measure of one-year distance to default (DD), which assesses bankruptcy risk, shows less variation among firms, with a higher mean of 0.783 and a smaller standard deviation of 0.352. This is consistent with the overall stability of Chinese firms. Notably, the average one-year DD values in 2008 and 2015, years marked by significant turmoil in the Chinese stock market, are 0.30 and 0.31 (untabulated), respectively, highlighting the substantial risks these firms faced during the global financial crisis and the subsequent equity turmoil in China.

In measuring AI investment, the four indicators—asset-based, text-based, LLM-based, and labor-based—show distinct patterns. Asset-based AI investment (AI_{Asset}) is relatively low on average (0.011), suggesting that AI-related assets account for only a small fraction of firms' total assets. However, maximum values reach as high as 1.513, pointing to highly concentrated AI investment among a limited set of firms. The mean value of text-based AI investment (AI_{Text}) is 1.812, with a relatively large standard deviation of 1.563, indicating considerable heterogeneity in AI-related disclosures across firms' annual reports. The LLM-based AI investment (AI_{LLM}) has a mean of 0.417 and a standard deviation of 0.253, reflecting a more concentrated distribution. The labor-based AI investment (AI_{Labor}) averages 0.032, with a median of zero, suggesting that most firms do not engage in AI-specific hiring; nevertheless, the maximum value of 1 reveals that some firms make significant commitments to AI talent acquisition. Taken together, across all four measures, average AI investment remains low, but dispersion is pronounced, consistent with the notion of "cautious participation by most, aggressive commitment by a few."

More importantly, market-driven and firm-specific components depict clear differences. For the market-driven component, the standard deviation is much smaller than the mean across all four measures, indicating relative stability. By contrast, the firm-specific component exhibits substantially higher variability, reflecting a volatile pattern. This suggests that while firms broadly align with prevailing market trends, they simultaneously pursue markedly idiosyncratic strategies, thereby potentially exposing themselves to additional risks.

[Insert Table 1 Here]

4 AI Investment and Default Risk

4.1 AI Investment and DD

Figure (1) shows a negative relationship between firm-level AI investment and distance to default across the four AI investment measures, indicating that higher AI investment is associated with greater default risk. To test this relationship more rigorously, we estimate the following panel regression model:

$$DD_{i,t+1} = \alpha + \beta, AI_{i,t} + \sum Controls_{i,t} + \delta_t + \xi_i + \varepsilon_{i,t+1}, \tag{2}$$

Regression (2) is an unbalanced panel model with year and firm fixed effects.[15] $DD_{i,t+1}$ represents the one-year-ahead distance-to-default (logarithm) for firm i. $AI_{i,t}$ is AI investment for firm i in year t, including $AI_{Asset,i,t}$, $AI_{Text,i,t}$, $AI_{LLM,i,t}$, and $AI_{Labor,i,t}$. $Controls_{i,t}$ are firm-level control variables, including size (*Size*), profitability (*ROA*), firm age (*Age*), value (*TobinQ*), leverage (*Leverage*), total asset growth rate (*AssetGrowthRate*), and earnings per share (*EPS*). $\varepsilon_{i,t+1}$ is the error term. Standard errors are clustered at the firm level.

We report the regression results in Table 2. Across all four measures, AI investment is negatively associated with subsequent credit risk. The estimated coefficients are economically meaningful and statistically significant at the 1% or 5% levels after controlling for firm characteristics. Specifically, based on the most stringent specification, a one–standard deviation increase in AI_{Asset}, AI_{Text}, AI_{LLM}, and AI_{Labor} reduces DD—the logarithm of distance to default—by 4.0% (= 0.014/0.352), 2.8% (= 0.010/0.352), 2.0% (= 0.007/0.352), and 1.7% (= 0.006/0.352), respectively, on average.

We quantify the increase in default probability associated with a one standard deviation rise in AI investment, focusing on asset- and labor-based measures, which exhibit the largest and smallest coefficients. A one standard deviation increase in the asset-based (labor-based) measure raises the probability of default by approximately 8.4% (3.5%) relative to the average firm's default probability, indicating economically meaningful effects.[16] Hence, AI investment can pose considerable risk to firms, due to the uncertainty

[15]It's worth noting that our sample includes financial firms, whose AI use is also indispensable for our research. And when excluding these firms, our results are robust. We also exclude AI-related firms (whose core business is AI) from the sample and re-run the analysis. Our findings remain consistent, which provides our conclusions are not merely driven by the unique characteristics of the AI industry but instead possess broader applicability.

[16]The default probability is calculated as $P = \Phi(-DD)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Using the sample mean of $\ln(DD) = 0.783$ (corresponding to $DD_{\text{mean}} = e^{0.783} = 2.188$), the average default probability is $P_{\text{mean}} = \Phi(-2.188) = 1.43\%$. A one–standard-deviation increase in asset-based (labor-based) AI investment reduces $\ln(DD)$ by 0.014 (0.006), implying new default probabilities of $P_{\text{new}} = 1.55\%$ (1.48%). Thus, the default probability rises by approximately 8.4% (3.5%) relative to the sample mean.

and high costs associated with irreversible investments in disruptive technologies.

Regression analyses using a five-year rolling window show that the effects of AI investment on firms' DD are time-varying but largely flattened (Figure 2). Overall, AI investments are associated with higher default risk, reflected in predominantly negative coefficients. Asset- and labor-based measures consistently indicate risk-enhancing effects, while text- and LLM-based measures briefly converge toward zero or turn positive around 2019–2020 before reverting negative. The evidence suggests that AI investment generally reduces DD and increases corporate risk exposure, with little indication that it systematically promotes firm growth or creditworthiness in the progress, even as AI adoption becomes more mature.

[Insert Table 2 Here]

[Insert Figure 2 Here]

We next examine the impact of AI investment on firm default risk by re-estimating regression (2), replacing overall AI investment with either market-driven or firm-specific AI investment, and report the results in Table 3. Standard errors are clustered at the firm level. The coefficients of Asset-based, Text-based, LLM-based, and Labor-based $MarketDrivenAI$ measures regressed on DD are 0.010, 0.056, 0.085, and 0.111, respectively, all positive and significant at the 1% level. Therefore, when a firm aligns its AI investment with peer benchmarks, such investment enhances cash flow quality and lowers default risk, consistent with the positive effects of AI technologies documented in the literature (Cockburn et al., 2018; Rock, 2019; Alderucci et al., 2020; Babina et al., 2024).

In contrast, the coefficients of Asset-based, Text-based, LLM-based, and Labor-based $FirmSpecificAI$ measures in the regression on DD are -0.006, -0.004, -0.003, and -0.004, respectively, all negative and significant at the 1% or 5% levels. These results indicate that when a firm invests in AI beyond the optimal peer benchmark, profitability declines and default risk rises. Because firm-specific AI investment exhibits greater variability, such deviations can offset the gains from market-driven investment, resulting in net losses. Consequently, firm-specific AI investment has a statistically and economically stronger impact on default risk than market-driven investment.

It is useful to compare our approach with that of Babina et al. (2023, 2024). First, Babina et al. (2024) finds that firms with higher AI investments experience greater growth in sales, employment, and market valuations. Their results hold across sectors, reflecting the broad increase in AI investments, which aligns with our finding that market-driven AI investment lowers default risk. Second, Babina et al. (2023) shows that firms with higher AI investments are associated with higher market betas. We extend this by demonstrating that deviations from market trends—captured by firm-specific AI investment—are linked to poorer corporate creditworthiness, highlighting the risks of idiosyncratic AI initiatives. Finally, our panel regression approach with firm and year fixed effects allows us to capture within-firm variations over time, whereas Babina et al. (2023, 2024) focus on cross-sectional effects.

[Insert Table 3 Here]

4.2 Robustness

4.2.1 Endogeneity

The endogeneity issue arises as usual. Firms in distress could resort to high-risk, AI innovations in a desperate attempt to improve their circumstances, which in turn could further increase risk. To address this potential reverse causality relationship between a firm's investment of AI and its default risk, we use the total number of AI-related academic papers as the instrumental variable of AI investment. Breakthroughs in academic research stimulate the development of commercial AI tools that firms can leverage, establishing relevance to firm AI investment. Seamans and Raj (2018) suggest that this measure is unlikely to directly influence individual firm investment decisions, suggesting exogeneity.

We conduct a search for AI-related papers published between 2006 and 2023 in the China National Knowledge Infrastructure (CNKI) database using the terms "artificial intelligence" and its Chinese equivalent "人工智能".[17] Additionally, we search the Web

[17]CNKI is the largest digital library of Chinese academic papers. English academic journals that are published by Elsevier, Springer Nature, Wiley and Taylor & Francis are also included in CNKI.

of Science (WoS) database using "artificial intelligence". We derive three measures of academic research output: one from CNKI using the English term, one from CNKI using the Chinese term, and one from WoS. We aggregate the counts of AI-related papers from these searches to define the annual research output measure, denoted $ResearchOutcomeAI_t$ for year t.

We present our instrumental variable (IV) regression results in Table 4. The coefficients for $\widehat{AI_{Asset}}$, $\widehat{AI_{Text}}$, $\widehat{AI_{LLM}}$, and $\widehat{AI_{Labor}}$, all of which are the fitted values from the first stage, are statistically significant and negative, with estimated values of -0.312, -0.006, -0.041, and -0.086, respectively.[18] Therefore, the instrumented AI investment negatively affects *DD* and increases the risk, further confirming that the negative correlation between AI investment and performance is causal.

[Insert Table 4 Here]

4.2.2 Additional Tests

We conduct a battery of robustness checks and summarize the findings in this section. First, to further enhance the robustness of our market-driven and firm-specific AI investment measures, we employ several alternative approaches. Specifically, we replace the market average of AI investment with the industry average and reconstruct the corresponding measures in equation 1. As shown in Panel A and Panel B of Table 5, our results remain consistent and robust.

We then define the market-driven component as the industry mean AI investment, with the firm-specific component equal to the difference between firm total AI investment and this mean. The results in Table A4 reveal that higher industry-level AI investment is associated with lower firm risk. In contrast, firm-specific deviations from the industry mean are positively related to risk.

[18] In the first stage, we run time-series regressions separately for each firm, regressing the number of AI-related papers on the proxy for the firm's AI investment to generate firm-level predicted AI investment. In the second stage, we replace the original AI investment measure with these predicted values and regress the firm's default distance (DD) on them. We conduct a weak instrument test using the Kleibergen-Paap rk Wald F statistic. This statistic exceeds the critical value established by Stock and Yogo (2005), which allows us to reject the null hypothesis of weak correlation between our instrument and the endogenous variable. In simpler terms, the test confirms that our chosen instrument is sufficiently strong.

[Insert Table 5 Here]

Second, we consider an alternative measure of firm default risk. In addition to the distance-to-default (DD) of Merton (1974) model, we consider Altman (1968)'s Z-score and the negative of Ohlson (1980)'s O-score. Both Z-score (lower score implies higher bankruptcy risk) and O-score (lower score implies higher risk) are widely used measures based on traditional accounting data (Dichev, 1998). Panel A and B of Table 6 demonstrates that these alternative measures yield results consistent with our main analysis.

[Insert Table 6 Here]

4.3 Cash Flow Channel: AI Investment and Default Risk

Our conceptual framework in Section 2. posits that both market-driven and firm-specific AI investments influence firms' cash flows—specifically their level, expected growth rate, and volatility—and thereby shape default risk. This section empirically examines the validity of these assumptions.

It is important to note that cash flows can be negative. In our sample, roughly 20% of cash flow observations are negative. Gorbenko and Strebulaev (2010) report that for U.S. corporations, 17.5% of EBITDA and 25.4% of cash flow from operations are negative. Given this prevalence, defining growth directly from cash flow is implausible. We therefore use cash flow over assets, denoted as *CashFlow*, to capture the level of cash flows, while return on assets, denoted as *ROA*, is employed to capture growth. The volatilities of these two measures, *CashFlowVol* and *ROAVol*, serve as proxies for second-order risk.

Table 7 and Table 8, using a panel specification similar to equation (2), report the relationship between market-driven AI investment and firms' cash flow risk. Panel A of both tables shows that all four measures of market-driven AI investment are strongly and positively associated with cash flow level and growth rate, with statistical significance at the 1% and 5% levels. Panel B of these tables reveals a significantly negative association between market-driven AI investment and the volatility of cash flow. These findings suggest that market-aligned AI investment enhances cash flow quality and thereby reduces default risk, in line with assumptions Section 2.

[Insert Table 7 Here]

[Insert Table 8 Here]

Firm-specific AI investment exhibits a markedly different pattern. With respect to cash flow level and growth, Panel A of Tables 9 and 10 show that firm-specific AI investment is significantly negatively associated with both measures. Regarding volatility, Panel B of Table 9 shows negative but insignificant coefficients, whereas Panel B of Table 10 reports negative coefficients, two of which are statistically significant at the 10% level, indicating a modest adverse effect on cash flow volatility. Overall, these results support the assumptions outlined in Section 2 and suggest that when firms deviate from prevailing market trends in their AI investment, cash flow quality tends to deteriorate.

[Insert Table 9 Here]

[Insert Table 10 Here]

In conclusion, these findings support the channels through which AI investment influences default risk. Market-driven AI investment strengthens firms' financial health by improving cash flow quality, thereby mitigating default risk. In contrast, firm-specific AI investment increases cash flow risks, ultimately raising the likelihood of default.

5 AI Investment and Data Complementarity

Just as humans require food to function, AI systems rely on a continuous stream of data to train and improve performance. Yet, it remains unclear whether data complements or substitutes AI investment in influencing a firm's default risk. In the following sections, we empirically examine the joint effects of AI investment and data on default risk. We first assess their direct impact on default risk and then investigate the underlying cash flow channels

5.1 DD and Data Quality

For firms, access to high-quality data is essential for successful AI implementation. While private firms generate their own data, government data is typically broader in scope and more comprehensive. In China, however, private access to public-sector data is often restricted. One strategy firms employ to gain access to such data is by providing services to the state. This form of collaboration can be especially valuable for firms with substantial AI investment, as access to government data may enhance the effectiveness and success of their AI initiatives.

We build on the approach of Beraja et al. (2023) to identify firms with access to valuable government data using textual analysis. Specifically, we scrape all Chinese government procurement contracts with Python, link each contract to the awarded firm, and classify contracts likely to involve data access (e.g., intelligence, electronics). This procedure enables us to assess whether firms with substantial AI investment and access to government data achieve greater success in AI implementation. We then construct a dummy variable, *DataRich*, which equals one if a firm has access to sufficiently rich and high-quality data, and zero otherwise.

We interact *DataRich* with both market-driven and firm-specific AI investment to assess how data influences their effects on firm default risk. The results are reported in Table 11. Panel A shows that the interaction between market-driven AI investment and data is significantly positive across all four AI investment proxies, indicating that abundant and high-quality data amplifies the risk-reducing effect of AI investment aligned with market trends, thereby complementing such investment. By contrast, Panel B reports uniformly insignificant coefficients for the interaction between firm-specific AI investment and data, suggesting that AI strategies diverging from prevailing market trends do not lower default risk, even when supported by rich and reliable data. This highlights that high-quality data does not necessarily complement AI investment when the investment is excessive or strategically misaligned.

These findings underscore that the effectiveness of AI investment depends on its alignment with data. Recent studies (Mihet et al., 2025; Fedyk et al., 2022) indicate that excessive data without sufficient AI investment can lead to informational overload and diminishing returns. Consistent with this perspective, our results show that the complementarity between data and AI is contingent on the level of AI investment: when investment is aligned with the benchmark, data enhances risk mitigation, whereas excessive investment relative to the benchmark diminishes its effectiveness.

[Insert Table 11 Here]

5.2 Cash Flow and Data Quality

This section examines how interactions between AI investment and data—both market-driven and firm-specific—affect cash flow risk, thereby testing the assumptions outlined in Section 2. Tables 12 and 13 present the results for market-driven AI investment interacting with data, while Tables 14 and 15 report the results for firm-specific AI investment interacting with data.

Regarding the impact of market-driven AI investment and data on volatility, we find that across all four investment proxies, high-quality data combined with AI investment at the benchmark level significantly reduces cash flow volatility (Panel B of Table 12) and slightly lowers ROA volatility, although the latter effect is not statistically significant (three coefficients in Columns (2)-(4) of Panel B in Table 13 are negative). The coefficients on market-driven AI investment remain positive and significant, consistent with the results reported in Tables 7.

With respect to cash flow level and growth rate, Panel A of Tables 12 and 13 show that the interaction between AI investment and data is not statistically significant, with all interaction coefficients insignificant. Notably, these interaction coefficients generally turn negative, while the coefficients on market-driven AI investment alone remain positive and significant, consistent with the earlier findings.

These results suggest that the primary channel through which data complements market-driven AI investment is by stabilizing cash flow volatility, rather than by increasing cash flow level or growth. This finding is consistent with the modeling framework of

(Farboodi and Veldkamp, 2021; Veldkamp and Chung, 2024), which assumes that data can help smooth variations in demand shocks.

[Insert Table12 Here]

[Insert Table13 Here]

Turning to the effect of firm-specific AI investment and data, Tables 14 and 15 show that all coefficients on the interaction between firm-specific AI and data—regressed on cash flow level, growth rate, and volatility—are insignificant. This suggests that when a firm invests beyond the market benchmark, additional high-quality data does not complement its investment and therefore fails to mitigate cash flow risk.[19]

[Insert Table14 Here]

[Insert Table15 Here]

5.3 Alternative Measures of Data

Knowledge capital refers to a firm' s intangible assets essential for achieving business objectives, including employee expertise; data on processes, products, customers, and competitors; and intellectual property such as patents or regulatory licenses (Ewens et al., 2024). Firms with greater knowledge capital are better positioned to leverage AI and are likely more data-intensive (Fedyk et al., 2023). Therefore, we use knowledge capital intensity, a broader measure of data availability, as a proxy for access to high-quality data.

Following Aghion et al. (2013) and Ewens et al. (2024), we construct knowledge capital as the accumulated stock of past R&D investments that contribute to current knowledge, after accounting for depreciation. Formally, the knowledge stock (K) at time t is calculated using a perpetual inventory method:

[19] Data is generally costly, and the lack of complementarity when AI investment deviates from the benchmark may further amplify the negative impact of firm-specific AI investment on cash flow risk.

$$K_t = R\&D_t + (1 - \delta)K_{t-1},$$

where $R\&D_t$ denotes current-year R&D expenditures and δ is the depreciation rate of knowledge capital. This method allows us to capture the effective contribution of historical R&D to current knowledge.

Following this framework, Ewens et al. (2024) use industry-year-specific knowledge capital depreciation rates published by the BEA when available, and a 15% rate otherwise. Aghion et al. (2013) use a fixed private knowledge depreciation rate of 15% to construct the R&D stock. Based on these studies, we construct two measures, denoted as *KnowledgeStock* and *RDStock*, respectively.

Rather than relying on a dummy variable, we employ these continuous proxies for data quality to replicate the analysis in Section 5.1 and Section 5.2. The corresponding results are reported in Tables A5 to A14. The findings remain robust, consistent with Section 5.1 and Section 5.2, when broader measures of data quality are employed. Specifically, the interaction between market-driven AI investment and knowledge capital does not lead to an improvement in cash flow or ROA. However, it reduces cash flow volatility and ROA volatility, thereby increasing DD and lowering bankruptcy risk. In contrast, the interaction between firm-specific AI investment and knowledge capital does not produce favorable outcomes: it neither enhances cash flow and ROA nor mitigates volatility.

6 Who Deviates from AI Optimal Benchmark?

Thus far, we find that firms overinvesting in AI beyond industry benchmarks face a higher risk of default. In this section, we show that this behavior is primarily driven by two factors: managerial risk preference and competitive pressure.

6.1 Risk Preference

Risk-tolerant managers often invest in AI beyond market benchmarks. They are willing to accept higher default risk because they prioritize potential upside, betting that additional, firm-specific AI investments will generate above-average returns.

To identify firms with risk preference, we adopt two approaches. Firstly, following Laeven and Levine (2009), we use the volatility of ROA over the past five years as a proxy for risk preference ($RiskPreference$). Firms with higher risk preference will be motivated to overinvest AI for their risk-taking proclivities. Secondly, we consider CEO overconfidence. Overconfident CEOs tend to underestimate the risk of failure and are more likely to aggressively pursue novel solutions beyond what's common in the market (Galasso and Simcoe, 2011). Following Schrand and Zechman (2012), we infer CEO overconfidence by analyzing their various investing and financing decisions. We create an overconfidence indicator, $CEOOverconfidence$, to capture CEO overconfidence. It takes the value of 1 if the firm meets at least three of the following five criteria, and 0 otherwise: (i) Excess investment is in the top quartile of firms within industry-years, where excess investment is the residual from a regression of total asset growth on sales growth; (ii) Net acquisitions from the statement of cash flows are in the top quartile of firms within industry-years; (iii) The debt-to-equity ratio is in the top quartile of firms within industry-years, where the debt-to-equity ratio is defined as long-term plus short-term debt divided by total market value; (iv) Either convertible debt or preferred stock is greater than zero; (v) The dividend yield is zero.

We regress firm-specific AI investment on $RiskPreference$ and $CEOOverconfidence$, and present the results in Panel A of Table 16. The findings indicate that $RiskPreference$ is positively associated with both LLM-based and labor-based firm-specific AI investment, with coefficients of 0.062 and 0.063, respectively, significant at the 5% level. In addition, $CEOOverconfidence$ shows a positive association with asset-based and text-based firm-specific AI investment, with coefficients of 0.029 and 0.002, significant at the 5% and 1% levels, respectively.

These results suggest that firms are more likely to overinvest in firm-specific AI relative to the benchmark when managers exhibit risk tolerance or overconfidence. Such behavioral traits may lead managers to pursue AI projects beyond market trends—risk-tolerant managers willingly allocate more resources to AI, while overconfident managers overestimate their ability to achieve additional upside.

[Insert Table16 Here]

6.2 Industry Competition

Competition is a major driving force behind companies adopting new technologies like AI. Studies have shown that to stay ahead or even survive intense competition, companies need to embrace new technologies (Aghion and Howitt, 1996; Aghion et al., 2005). This means that businesses in highly competitive industries are likely to invest more heavily in AI compared to the average company.

We use the Herfindahl-Hirschman Index (HHI) to measure the level of competition in different industries. This index considers both the number of competitors in a market and their relative sizes. A higher HHI indicates fewer competitors or a larger gap between them, suggesting a less competitive market. Conversely, a lower HHI indicates a more competitive market with many players of similar size. We calculated the HHI by squaring the market share of each competitor in an industry. The data for these market shares came from the CSMAR database. We construct a variable, $RecHHI$, to measure competition facing a firm. $RecHHI$ represents logarithm of the reciprocal of the HHI. We also use $NumberOfFirms$, logarithm of the number of firms in the industry as an alternative measure of industry competition. Generally, a higher number of competitors indicates a more competitive market.

We investigate the relationship between competition measures and firm-specific AI investment. Panel C and D of Table16 reports the results. We find that $RecHHI$ is positively associated with text-based, LLM-based, and labor-based firm-specific AI investment, with coefficients of 0.145, 0.029, and 0.007, respectively, significant at the 1% and 5% levels. Similarly, $NumberOfFirms$ exhibits a positive relationship with text-based, LLM-based, and labor-based firm-specific AI investment, with coefficients of

0.215, 0.048, and 0.017, all significant at the 1% level.

Our results suggest that firms operating in more competitive environments are more inclined to allocate resources toward AI beyond market average, potentially as a strategic response to competitive pressures, either to maintain their market position or to seek competitive advantages through early AI adoption.

7 Conclusion

Our paper provides novel evidence on the implications of AI investment for firm default risk. Using four complementary firm-level measures of AI investment—asset-based, text-based, LLM-based, and labor-based—we show that the effect of AI investment on corporate default risk critically depends on investment intensity. When firms invest at the market benchmark level, AI investment enhances cash flows, stabilizes profitability, and reduces default risk. In contrast, excessive firm-specific investment beyond the benchmark undermines cash flow and increases the likelihood of default. These findings highlight that while AI investment offers substantial growth potential, firm-specific deviations expose firms to heightened financial vulnerability, providing an important corrective to the prevailing narrative that emphasizes only the benefits of AI.

Moreover, we demonstrate that access to high-quality data plays a complementary role in shaping the risk-return trade-off of AI investment. Abundant and reliable data amplifies the stabilizing effects of market-driven (benchmark) AI investment on cash flows but does not mitigate the adverse consequences of firm-specific (deviation) investment. This finding provides new insights into the interplay between AI investment and data quality and quantity.

Our findings contribute to the literature on AI, data economics, and corporate risk management by demonstrating that the benefits of AI are conditional on investment intensity and data availability. We show that while AI adoption aligned with market benchmarks can enhance financial stability, excessive investment, driven by managerial behaviors, may increase firm-level credit risk and potentially create systemic vulnerabilities. For policymakers, this highlights the importance of guiding AI-related investments to ensure they promote, rather than undermine, financial stability.

References

Acemoglu, D. (2025). The simple macroeconomics of ai. *Economic Policy*, 40(121):13–58.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

Aghion, P. and Howitt, P. (1996). Research and development in the growth process. *Journal of Economic Growth*, 1:49–73.

Aghion, P., Jones, B. F., and Jones, C. I. (2017). *Artificial Intelligence and Economic Growth*, volume 23928. National Bureau of Economic Research Cambridge, MA.

Aghion, P., Van Reenen, J., and Zingales, L. (2013). Innovation and institutional ownership. *American Economic Review*, 103(1):277–304.

Alderucci, D., Branstetter, L., Hovy, E., Runge, A., and Zolas, N. (2020). Quantifying the impact of ai on productivity and labor demand: Evidence from us census microdata. In *Allied social science associations—ASSA 2020 annual meeting*.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

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

Babina, T., Fedyk, A., He, A. X., and Hodson, J. (2023). Artificial intelligence and firms' systematic risk. *Available at SSRN 4868770*.

Bao, J., Pan, J., and Wang, J. (2011). The illiquidity of corporate bonds. *The Journal of Finance*, 66(3):911–946.

Beraja, M., Yang, D. Y., and Yuchtman, N. (2023). Data-intensive innovation and the state: Evidence from ai firms in china. *The Review of Economic Studies*, 90(4):1701–1723.

Bharath, S. T. and Shumway, T. (2008). Forecasting default with the merton distance to default model. *The Review of Financial Studies*, 21(3):1339–1369.

Brogaard, J., Li, D., and Xia, Y. (2017). Stock liquidity and default risk. *Journal of Financial Economics*, 124(3):486–502.

Brynjolfsson, E., Li, D., and Raymond, L. (2025). Generative ai at work. *The Quarterly Journal of Economics*, 140(2):889–942.

Brynjolfsson, E., Rock, D., and Syverson, C. (2019). Artificial intelligence and the modern productivity paradox. *The Economics of Artificial Intelligence: An Agenda*, 23:23–57.

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6):2899–2939.

Campbell, J. Y. and Taksler, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of Finance*, 58(6):2321–2350.

Chen, L., Collin-Dufresne, P., and Goldstein, R. S. (2009). On the relation between the credit spread puzzle and the equity premium puzzle. *The Review of Financial Studies*, 22(9):3367–3409.

Chen, W. and Srinivasan, S. (2023). Going digital: Implications for firm value and performance. *Review of Accounting Studies*, pages 1–47.

Cockburn, I. M., Henderson, R., and Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation*, volume 24449. National Bureau of Economic Research Cambridge, MA, USA.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? *the Journal of Finance*, 53(3):1131–1147.

Dixit, A. K. and Pindyck, R. S. (1994). *Investment under uncertainty.* Princeton university press.

Dougal, C., Parsons, C. A., and Titman, S. (2015). Urban vibrancy and corporate growth. *The Journal of Finance*, 70(1):163–210.

Eberhart, A., Maxwell, W., and Siddique, A. (2008). Reexamination of the tradeoff between the future benefit and riskiness of r&d increases. *Journal of Accounting Research*, 46(1):27–52.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

Ewens, M., Peters, R. H., and Wang, S. (2024). Measuring intangible capital with market prices. *Management Science*.

Farboodi, M., Mihet, R., Philippon, T., and Veldkamp, L. (2019). Big data and firm dynamics. In *AEA papers and proceedings*, volume 109, pages 38–42. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Farboodi, M. and Veldkamp, L. (2021). A model of the data economy. Technical report, National Bureau of Economic Research Cambridge, MA, USA.

Fedyk, A., Gomes, O., Mihet, R., and Rishabh, K. (2023). Data innovation complementarity and firm growth. *Swiss finance institute research paper*, (23-86).

Fedyk, A., Hodson, J., Khimich, N., and Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies*, 27(3):938–985.

Fracassi, C. (2017). Corporate finance policies and social networks. *Management Science*, 63(8):2420–2438.

Galasso, A. and Simcoe, T. S. (2011). Ceo overconfidence and innovation. *Management Science*, 57(8):1469–1484.

Goldstein, R., Ju, N., and Leland, H. (2001). An ebit-based model of dynamic capital structure. *The Journal of Business*, 74(4):483–512.

Gorbenko, A. S. and Strebulaev, I. A. (2010). Temporary versus permanent shocks: Explaining corporate financial policies. *The Review of Financial Studies*, 23(7):2591–2647.

Graetz, G. and Michaels, G. (2018). Robots at work. *Review of Economics and Statistics*, 100(5):753–768.

Hackbarth, D. and Mauer, D. C. (2012). Optimal priority structure, capital structure, and investment. *The Review of Financial Studies*, 25(3):747–796.

He, H., Li, S. X., and Han, Y. (2023). Labor market discrimination against family responsibilities: A correspondence study with policy change in china. *Journal of Labor Economics*, 41(2):361–387.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, 76(2):323–329.

Jensen, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *the Journal of Finance*, 48(3):831–880.

Jha, M., Qian, J., Weber, M., and Yang, B. (2024). Chatgpt and corporate policies. Technical report, National Bureau of Economic Research.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858.

Laeven, L. and Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93(2):259–275.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2):449–470.

Mihet, R., Rishabh, K., and Gomes, O. (2025). Is it ai or data that drives market power? *Swiss Finance Institute Research Paper*, (25-37).

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131.

Rock, D. (2019). Engineering value: The returns to technological talent and investments in artificial intelligence. *Available at SSRN 3427412*.

Schrand, C. M. and Zechman, S. L. (2012). Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and Economics*, 53(1-2):311–329.

Seamans, R. and Raj, M. (2018). Ai, labor, productivity and the need for firm-level data. Technical report, National Bureau of Economic Research.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, page 80.

Veldkamp, L. and Chung, C. (2024). Data and the aggregate economy. *Journal of Economic Literature*, 62(2):458–484.

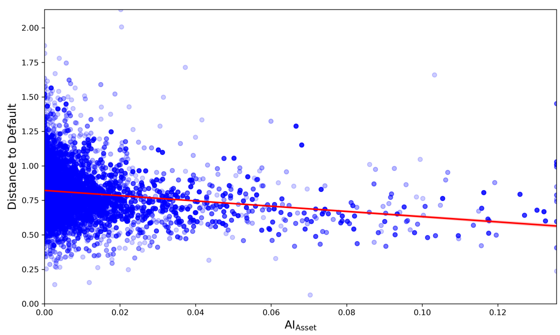
Westerman, G., Bonnet, D., and McAfee, A. (2014). *Leading Digital: Turning Technology into Business Transformation.* Harvard Business Press.

Figures

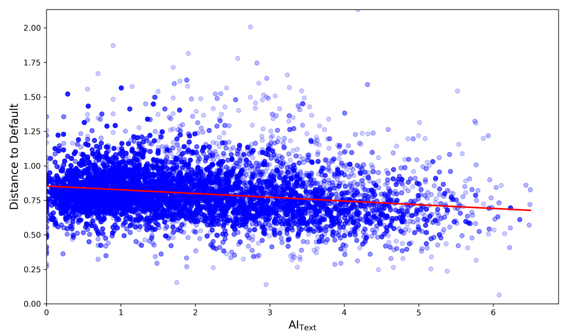
Figure 1

AI Investment and DD

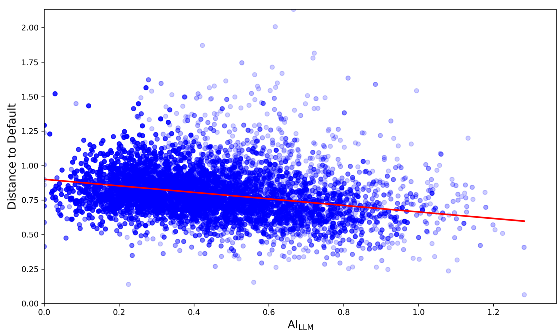
Note: This figure illustrates the relationship between AI investment and distance-to-default (DD) at the firm level. It consists of four subplots, each corresponding to a different measure of AI investment: asset-based (AI_{Asset}), text-based (AI_{Text}), LLM-based (AI_{LLM}), and labor-based ($AI_{Recruits}$). Each subplot displays the average values of the respective AI measure and DD across firms, along with a fitted regression line indicating their association.



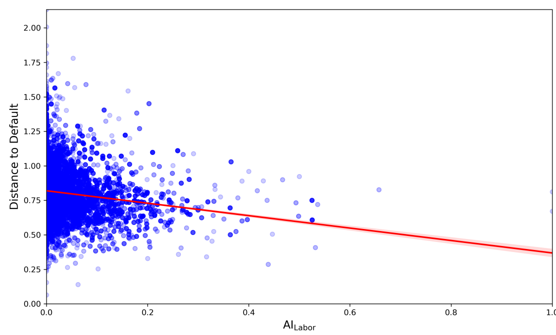
(a)
Asset-based AI investment and DD



(b)
Text-based AI investment and DD



(c)
LLM-based AI investment and DD

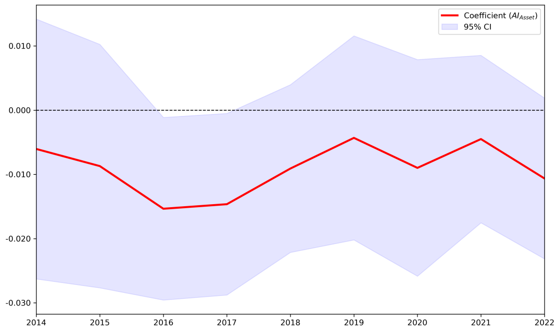


(d)
Labor-based AI investment and DD

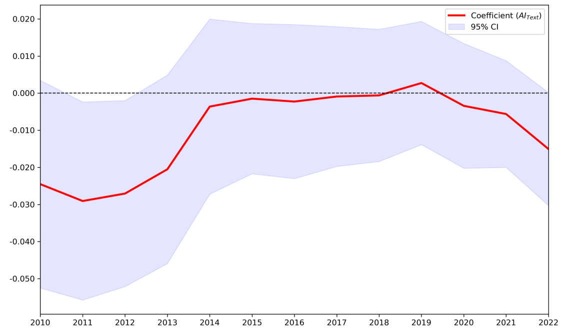
Figure 2

Time Varying Impact of AI Investment

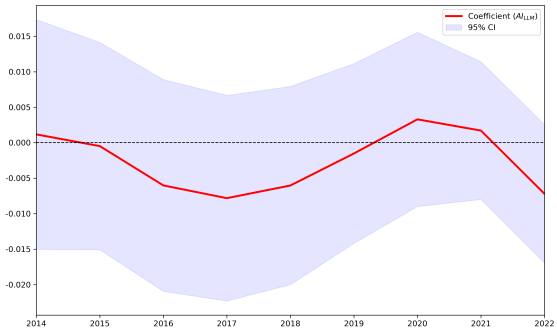
Note: This figure presents the rolling-window estimates of the regression coefficients depicting the relationship between AI investment and distance-to-default (DD). The analysis employs a five-year rolling window to compute time-varying coefficients. The figure contains four subplots, each corresponding to a different measure of AI investment: asset-based (AI_{Asset}), text-based (AI_{Text}), LLM-based (AI_{LLM}), and labor-based ($AI_{Recruits}$). Each subplot displays the evolution of the estimated coefficient between the respective AI measure and DD over time.



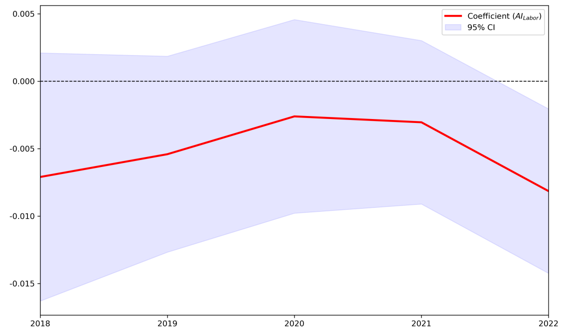
(a)
Asset-based AI investment and DD



(b)
Text-based AI investment and DD



(c)
LLM-based AI investment and DD



(d)
Labor-based AI investment and DD

Tables

Table 1

Summary Statistics

Note: This table reports summary statistics of the main variables used in the study. *DD* denotes the distance-to-default measure. AI variables capture firms' investments in AI across different dimensions: Asset, Text, LLM, and Labor. *MarketDrivenAI* and *FirmSpecificAI* represent the portions attributable to market-driven and firm-specific AI investment, respectively. Control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of the variables are provided in Table A1 in the Appendix. Observations (Obs), mean, standard deviation (SD), minimum (Min), median, and maximum (Max) are reported for each variable.

Variable	Obs	Mean	SD	Min	Median	Max
DD (log)	44006	0.783	0.352	-1.233	0.789	2.482
AI_{Asset} (log)	39101	0.011	0.043	-0.061	0.003	1.513
$\text{MarketDrivenAI}_{\text{Asset}}$	39101	0.001	0.000	0.000	0.001	0.001
$\text{FirmSpecificAI}_{\text{Asset}}$	39101	0.000	0.019	-0.946	-0.000	0.693
AI_{Text} (log)	44006	1.812	1.563	0.000	1.609	6.877
$\text{MarketDrivenAI}_{\text{Text}}$	44006	1.204	0.583	0.238	1.430	1.911
$\text{FirmSpecificAI}_{\text{Text}}$	44006	0.000	0.641	-3.344	0.014	3.854
AI_{LLM} (log)	39101	0.417	0.253	0.000	0.405	1.368
$\text{MarketDrivenAI}_{\text{LLM}}$	39101	0.318	0.103	0.122	0.332	0.450
$\text{FirmSpecificAI}_{\text{LLM}}$	39101	0.000	0.122	-0.893	0.000	0.668
AI_{Labor}	27097	0.032	0.078	0.000	0.000	1.000
$\text{MarketDrivenAI}_{\text{Labor}}$	27097	0.023	0.005	0.011	0.025	0.028
$\text{FirmSpecificAI}_{\text{Labor}}$	27097	0.000	0.055	-0.454	-0.002	0.890
Size (log)	44006	22.207	1.296	19.768	22.013	26.220
ROA	44006	0.034	0.063	-0.243	0.035	0.196
Leverage	44006	0.435	0.204	0.051	0.430	0.897
EPS	44006	0.395	0.678	-1.568	0.280	3.435
TobinQ	44006	2.418	1.705	0.830	1.881	10.827
Age (log)	44006	2.322	0.665	1.099	2.398	3.401
AssetGrowthRate	44006	0.156	0.338	-0.292	0.084	2.744

Table 2

Impact of AI Investment on Firm Default Risk

Note: This table reports regression estimates of a firm's default risk on AI investment. The dependent variable *DD* is one-year ahead distance-to-default. The key independent variables are asset-based (AI_{Asset}), text-based (AI_{Text}), LLM-based (AI_{LLM}) and labor-based (AI_{Labor}) AI investment. The control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	DD	DD	DD	DD	DD	DD	DD	DD
AI_{Asset}	-0.014***	-0.014***						
	(-4.27)	(-4.53)						
AI_{Text}			-0.004	-0.010***				
			(-1.07)	(-2.88)				
AI_{LLM}					-0.004	-0.007**		
					(-1.30)	(-2.24)		
AI_{Labor}							-0.005**	-0.006**
							(-2.01)	(-2.47)
Size		0.051***		0.043***		0.051***		0.055***
		(9.89)		(10.47)		(9.94)		(8.02)
ROA		0.910***		0.832***		0.924***		0.861***
		(16.80)		(16.87)		(17.04)		(14.57)
Leverage		-0.100***		-0.072***		-0.097***		-0.131***
		(-5.11)		(-4.46)		(-4.97)		(-5.51)
EPS		-0.099***		-0.089***		-0.099***		-0.080***
		(-15.58)		(-15.07)		(-15.58)		(-11.66)
TobinQ		-0.009***		-0.011***		-0.009***		-0.007***
		(-5.77)		(-7.67)		(-5.73)		(-4.01)
Age		0.081***		0.061***		0.081***		0.035**
		(7.11)		(6.30)		(7.16)		(2.23)
AssetGrowthRate		-0.048***		-0.052***		-0.049***		-0.034***
		(-8.96)		(-11.13)		(-9.07)		(-5.49)
Intercept	0.842***	-0.395***	0.806***	-0.220**	0.842***	-0.400***	0.832***	-0.402***
	(27127.96)	(-3.54)	(3380.92)	(-2.45)	(3460.35)	(-3.60)	(1.2e+05)	(-2.67)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	33,472	33,472	38,187	38,187	33,473	33,473	26,283	26,283
R^2	0.519	0.532	0.557	0.568	0.519	0.532	0.581	0.589

Table 3

Impact of Market-driven and Firm-specific AI on Firm Default Risk

Note: This table reports regression estimates of a firm's default risk on market-driven and firm-specific AI investment. The dependent variable *DD* is one-year ahead distance-to-default. The key independent variables are asset-based, text-based, LLM-based and labor-based market-driven and firm-specific AI investment. The control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Market-driven AI investment and firm default risk				
$\text{MarketDrivenAI}_{\text{Asset}}$	0.010***			
	(2.82)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.056***		
		(12.15)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.085***	
			(18.23)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.111***
				(32.56)
Controls	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes
N	33,472	38,187	33,473	26,283
R^2	0.300	0.318	0.308	0.393
Panel B: Firm-specific AI investment and firm default risk				
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.006***			
	(-4.53)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.004***		
		(-2.88)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.003**	
			(-2.24)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.004**
				(-2.47)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,472	38,187	33,473	26,283
R^2	0.532	0.568	0.532	0.589

Table 4

Impact of AI Investment on Firm Default Risk - 2SLS approach

Note: This table reports the second-stage results from the two-stage least squares (2SLS) regression of a firm's default risk on AI investment. The dependent variable is the one-year ahead distance-to-default (DD). The key independent variables are the instrumented measures of AI investment, $\widehat{AI}_{Asset}$, $\widehat{AI}_{Text}$, $\widehat{AI}_{LLM}$ and $\widehat{AI}_{Labor}$, which are the predicted value from the first-stage regression of the firm's AI investment on the number of AI-related research publications. Control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed variable definitions are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
$\widehat{AI}_{Asset}$	-0.312***			
	(-4.01)			
$\widehat{AI}_{Text}$		-0.006**		
		(-2.46)		
$\widehat{AI}_{LLM}$			-0.041**	
			(-2.58)	
$\widehat{AI}_{Labor}$				-0.086**
				(-2.21)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	25,801	30,722	25,801	14,101
R^2	0.525	0.569	0.525	0.607

Table 5

Robustness Tests: Alternative Measure of AI Decomposition

Note: This table reports regression estimates of a firm's default risk on market-driven and firm-specific AI investment. The dependent variable is the one-year ahead distance-to-default (DD). The key independent variables consist of alternative measures of asset-based, text-based, LLM-based and labor-based market-driven and firm-specific AI investment. These measures are constructed by substituting the market average with the industry mean in the decomposition framework specified in Equation 1. Control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed variable definitions are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Alternative measure of market-driven AI investment				
$\text{AltMarketDrivenAI}_{\text{Asset}}$	0.000			
	(0.04)			
$\text{AltMarketDrivenAI}_{\text{Text}}$		0.036***		
		(6.88)		
$\text{AltMarketDrivenAI}_{\text{LLM}}$			0.093***	
			(14.77)	
$\text{AltMarketDrivenAI}_{\text{Labor}}$				0.115***
				(13.27)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,473	38,187	33,473	26,283
R^2	0.300	0.316	0.306	0.374
Panel B: Alternative measure of firm-specific AI investment				
$\text{AltFirmSpecificAI}_{\text{Asset}}$	-0.006***			
	(-4.43)			
$\text{AltFirmSpecificAI}_{\text{Text}}$		-0.003*		
		(-1.85)		
$\text{AltFirmSpecificAI}_{\text{LLM}}$			-0.002	
			(-1.27)	
$\text{AltFirmSpecificAI}_{\text{Labor}}$				-0.004**
				(-2.32)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,473	38,187	33,473	26,283
R^2	0.532	0.568	0.532	0.589

Table 6

Robustness Tests: Alternative Measure of Firm Default Risk

Note: This table reports regression estimates of a firm's default risk on AI investment. The dependent variable are the one-year ahead Altman (1968)'s Z-score (*ZScore*) and the negative of Ohlson (1980)'s O-score (*OScore*). The key independent variables are asset-based, text-based, LLM-based and labor-based AI investment. Control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed variable definitions are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Alternative measure of firm default risk: Z-score				
	ZScore	ZScore	ZScore	ZScore
AI_{Asset}	-0.008			
	(-0.15)			
AI_{Text}		-0.176***		
		(-3.19)		
AI_{LLM}			-0.052	
			(-1.26)	
AI_{Labor}				-0.063**
				(-2.09)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,472	38,185	33,472	26,283
R^2	0.743	0.727	0.743	0.768
Panel B: Alternative measure of firm default risk: O-score				
	OScore	OScore	OScore	OScore
AI_{Asset}	-0.095***			
	(-3.97)			
AI_{Text}		-0.045**		
		(-2.20)		
AI_{LLM}			-0.054***	
			(-3.48)	
AI_{Labor}				-0.011
				(-0.97)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,473	38,187	33,473	26,283
R^2	0.734	0.727	0.733	0.749

Table 7

Impact of Market-driven AI on Cash Flow

Note: This table presents regression estimates examining the relationship between cash flow variables and market-driven AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). The key independent variables are asset-based, text-based, LLM-based and labor-based market-driven AI investment. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Impact of market-driven AI investment on cash flow level				
	CF	CF	CF	CF
$\text{MarketDrivenAI}_{\text{Asset}}$	0.003***			
	(3.19)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.004***		
		(3.25)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.005***	
			(4.02)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.002**
				(1.98)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,447	38,162	33,448	26,260
R^2	0.413	0.381	0.414	0.459
Panel B: Impact of market-driven AI investment on cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.001			
	(-1.47)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.001*		
		(-1.89)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.002**	
			(-2.47)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.001***
				(-3.23)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	29,054	33,729	29,054	21,885
R^2	0.614	0.577	0.615	0.699

Table 8

Impact of Market-driven AI on ROA

Note: This table presents regression estimates examining the relationship between ROA variables and market-driven AI investment. The dependent variables are ROA (*ROA*) and ROA volatility (*ROAvol*). The key independent variables are asset-based, text-based, LLM-based and labor-based market-driven AI investment. The control variables include *Size*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Impact of market-driven AI investment on ROA				
	ROA	ROA	ROA	ROA
$\text{MarketDrivenAI}_{\text{Asset}}$	0.006***			
	(6.03)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.006***		
		(3.87)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.008***	
			(5.65)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.007***
				(6.57)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,472	38,187	33,473	26,283
R^2	0.418	0.379	0.419	0.426
Panel B: Impact of market-driven AI investment on ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.001*			
	(-1.67)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.004***		
		(-3.73)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.005***	
			(-5.11)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.003***
				(-5.47)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	29,054	33,729	29,054	21,885
R^2	0.632	0.577	0.633	0.780

Table 9

Impact of Firm-specific AI on Cash Flow

Note: This table presents regression estimates examining the relationship between cash flow variables and firm-specific AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). The key independent variables are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Impact of firm-specific AI investment on cash flow level				
	CF	CF	CF	CF
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.001**			
	(-2.23)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001***		
		(-2.67)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.001***	
			(-3.53)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000
				(-0.60)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,447	38,162	33,448	26,260
R^2	0.421	0.392	0.421	0.466
Panel B: Impact of firm-specific AI investment on cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{FirmSpecificAI}_{\text{Asset}}$	0.000			
	(0.59)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.000		
		(-0.13)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000	
			(-0.25)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000
				(-0.89)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	29,054	33,729	29,054	21,885
R^2	0.621	0.584	0.621	0.705

Table 10

Impact of Firm-specific AI on ROA

Note: This table presents regression estimates examining the relationship between ROA variables and firm-specific AI investment. The dependent variables are ROA (*ROA*) and ROA volatility (*ROAvol*). The key independent variables are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. The control variables include *Size*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Impact of firm-specific AI investment on ROA				
	ROA	ROA	ROA	ROA
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.002*** (-3.18)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001** (-1.99)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.001** (-2.24)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.001* (-1.84)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,472	38,187	33,473	26,283
R^2	0.421	0.384	0.421	0.428
Panel B: Impact of firm-specific AI investment on ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$\text{FirmSpecificAI}_{\text{Asset}}$	0.001* (1.68)			
$\text{FirmSpecificAI}_{\text{Text}}$		0.000 (0.93)		
$\text{FirmSpecificAI}_{\text{LLM}}$			0.000* (1.74)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000 (-0.02)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	29,054	33,729	29,054	21,885
R^2	0.651	0.594	0.651	0.789

Table 11
AI decomposition, Data and Firm Default Risk

Note: This table reports regression estimates analyzing how a firm's data quality moderates the relationship between its default risk and market-driven as well as firm-specific AI investment. The dependent variable DD is one-year ahead distance-to-default. $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $DataRich$ is a dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Market-driven AI investment, data and DD				
$\text{MarketDrivenAI}_{\text{Asset}} \times$ DataRich	0.474***			
	(5.99)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ DataRich		0.106***		
		(3.63)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ DataRich			0.146***	
			(5.25)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ DataRich				0.079**
				(2.56)
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.122***			
	(-12.27)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.177***		
		(-16.24)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.026***	
			(2.73)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.160***
				(-29.66)
DataRich	-2.293***	-0.522***	-0.704***	-0.403**
	(-5.91)	(-3.38)	(-5.19)	(-2.27)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,192	22,193	22,193	22,193
R^2	0.419	0.423	0.414	0.444
Panel B: Firm-specific AI investment, data and DD				
$\text{FirmSpecificAI}_{\text{Asset}} \times$ DataRich	-0.001			
	(-0.36)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ DataRich		0.002		
		(0.85)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ DataRich			0.001	
			(0.30)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ DataRich				-0.000
				(-0.01)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.007**			
	(-2.45)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.004*		
		(-1.69)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000	
			(-0.12)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.004*
				(-1.94)
DataRich	0.014	0.013	0.013	0.015
	(1.23)	(1.09)	(1.11)	(1.05)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,192	22,193	22,193	22,193
R^2	0.413	0.413	0.413	0.413

Table 12
Market-driven AI, Data and Cash Flow

Note: This table reports regression estimates analyzing how a firm's data quality moderates the relationship between its cash flow and market-driven AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $DataRich$ is a dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, data and cash flow level				
	CF	CF	CF	CF
$\text{MarketDrivenAI}_{\text{Asset}} \times$ DataRich	-0.008 (-0.59)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ DataRich		-0.006 (-1.15)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ DataRich			-0.003 (-0.56)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ DataRich				-0.010 (-1.62)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.002 (0.88)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.009*** (4.02)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.007*** (3.40)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.003** (2.45)
DataRich	0.034 (0.50)	0.024 (0.89)	0.006 (0.25)	0.054 (1.46)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	21,666	21,666	21,666	21,666
R^2	0.479	0.480	0.480	0.479
Panel B: Market-driven AI investment, data and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{MarketDrivenAI}_{\text{Asset}} \times$ DataRich	-0.001 (-0.41)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ DataRich		-0.005** (-2.49)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ DataRich			-0.005** (-2.30)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ DataRich				-0.003** (-2.20)
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.005*** (-5.52)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.008*** (-6.66)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.008*** (-7.23)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.004*** (-6.38)
DataRich	0.005 (0.34)	0.022** (2.48)	0.023** (2.31)	0.016** (2.17)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	17,833	17,833	17,833	17,833
R^2	0.762	0.763	0.764	0.763

Table 13
Market-driven AI, Data and ROA

Note: This table reports regression estimates analyzing how a firm's data quality moderates the relationship between its ROA and market-driven AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $DataRich$ is a dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. The control variables include $Size$, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, data and ROA				
	ROA	ROA	ROA	ROA
$\text{MarketDrivenAI}_{\text{Asset}} \times$ DataRich	-0.001 (-0.06)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ DataRich		-0.002 (-0.39)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ DataRich			-0.004 (-0.94)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ DataRich				0.000 (0.01)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.016*** (4.51)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.015*** (3.57)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.015*** (4.44)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.007*** (3.72)
DataRich	-0.003 (-0.05)	0.004 (0.16)	0.013 (0.63)	-0.005 (-0.20)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,192	22,193	22,193	22,193
R^2	0.507	0.508	0.508	0.508
Panel B: Market-driven AI investment, data and ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$\text{MarketDrivenAI}_{\text{Asset}} \times$ DataRich	0.007 (1.18)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ DataRich		-0.003 (-0.96)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ DataRich			-0.002 (-0.71)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ DataRich				-0.001 (-0.43)
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.013*** (-9.83)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.017*** (-10.55)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.017*** (-11.17)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.009*** (-10.51)
DataRich	-0.033 (-1.18)	0.015 (1.04)	0.013 (0.79)	0.007 (0.49)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	17,833	17,833	17,833	17,833
R^2	0.822	0.824	0.824	0.823

Table 14
Firm-specific AI, Data and Cash Flow

Note: This table reports regression estimates analyzing how a firm's data quality moderates the relationship between its cash flow and firm-specific AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $DataRich$ is a dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, data and cash flow level				
	CF	CF	CF	CF
$\text{FirmSpecificAI}_{\text{Asset}} \times$ DataRich	0.000 (0.71)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ DataRich		0.000 (0.64)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ DataRich			0.000 (0.69)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ DataRich				-0.000 (-0.45)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.000 (-0.72)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001 (-1.39)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.001*** (-2.71)	
$\text{FirmSpecificAI}_{\text{Labor}}$				0.000 (1.07)
DataRich	-0.004 (-1.56)	-0.004 (-1.58)	-0.004 (-1.54)	-0.004 (-1.56)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,171	22,172	22,172	22,172
R^2	0.501	0.501	0.501	0.501
Panel B: Firm-specific AI investment, data and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{FirmSpecificAI}_{\text{Asset}} \times$ DataRich	0.000 (0.76)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ DataRich		0.000 (0.63)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ DataRich			0.000 (0.67)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ DataRich				-0.000 (-0.55)
$\text{FirmSpecificAI}_{\text{Asset}}$	0.000 (0.83)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.000 (-0.96)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000 (-0.68)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000 (-1.24)
DataRich	0.000 (0.20)	0.000 (0.18)	0.000 (0.17)	0.000 (0.24)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	17,833	17,833	17,833	17,833
R^2	0.778	0.778	0.778	0.778

Table 15
Firm-specific AI, Data and ROA

Note: This table reports regression estimates analyzing how a firm's data quality moderates the relationship between its ROA and firm-specific AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $DataRich$ is a dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise. The control variables include $Size$, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, data and ROA				
	ROA	ROA	ROA	ROA
$FirmSpecificAI_{Asset}$ × DataRich	0.000 (0.48)			
$FirmSpecificAI_{Text}$ × DataRich		0.000 (1.04)		
$FirmSpecificAI_{LLM}$ × DataRich			0.000 (0.75)	
$FirmSpecificAI_{Labor}$ × DataRich				0.000 (0.48)
$FirmSpecificAI_{Asset}$	-0.004*** (-2.61)			
$FirmSpecificAI_{Text}$		-0.001 (-1.34)		
$FirmSpecificAI_{LLM}$			-0.001 (-0.86)	
$FirmSpecificAI_{Labor}$				-0.001 (-1.40)
DataRich	-0.006** (-2.52)	-0.007*** (-2.79)	-0.007*** (-2.78)	-0.007*** (-2.70)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22192	22193	22193	22193
R^2	0.509	0.510	0.510	0.510
Panel B: Firm-specific AI investment, data and ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$FirmSpecificAI_{Asset}$ × DataRich	0.000 (0.38)			
$FirmSpecificAI_{Text}$ × DataRich		0.000 (0.95)		
$FirmSpecificAI_{LLM}$ × DataRich			0.000 (0.34)	
$FirmSpecificAI_{Labor}$ × DataRich				-0.000 (-1.24)
$FirmSpecificAI_{Asset}$	-0.000 (-0.49)			
$FirmSpecificAI_{Text}$		-0.000 (-1.24)		
$FirmSpecificAI_{LLM}$			-0.000 (-0.74)	
$FirmSpecificAI_{Labor}$				0.000 (1.39)
DataRich	0.002* (1.85)	0.002* (1.75)	0.002* (1.84)	0.002* (1.84)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	17,833	17,833	17,833	17,833
R^2	0.826	0.826	0.826	0.826

Table 16

Motivations for Firm-specific AI

Note: This table reports the motivation for firm to invest firm-specific AI. The key dependent variables are one-year-ahead asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $RiskPreference$ is the volatility of ROA over the past five years as a proxy for risk preference. $CEOOverconfidence$ is a proxy for CEO overconfidence using Schrand and Zechman (2012)'s method. $RecHHI$ represents the reciprocal of the HHI. $NumberOfFirms$, the number of firms in the industry. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	$\text{FirmSpecificAI}_{\text{Asset}}$	$\text{FirmSpecificAI}_{\text{Text}}$	$\text{FirmSpecificAI}_{\text{LLM}}$	$\text{FirmSpecificAI}_{\text{Labor}}$
Panel A: Risk preference				
RiskPreference	0.001	-0.040	0.062**	0.033**
	(0.33)	(-0.23)	(1.98)	(2.16)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	28854	29836	28854	20629
R^2	0.082	0.104	0.057	0.053
Panel B: CEO overconfidence				
CEOOverconfidence	0.002***	0.029**	0.003	0.000
	(2.85)	(2.48)	(1.12)	(0.34)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	34755	38162	34756	24177
R^2	0.020	0.029	0.024	0.028
Panel C: HHI				
RecHHI	0.002	0.145***	0.029***	0.007**
	(1.31)	(4.62)	(5.68)	(2.23)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	34083	37337	34084	23771
R^2	0.020	0.034	0.028	0.028
Panel D: Number of firms				
NumberOfFirms	0.003	0.215***	0.048***	0.017***
	(1.05)	(5.56)	(7.75)	(3.94)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	34085	37342	34086	23772
R^2	0.019	0.036	0.030	0.029

Appendix

Table A1

Variable Definitions

Variable Name	**Definition**
Panel A: Firm default risk	
DD	Logarithm of the default distance calculated by the KMV model; higher DD indicates lower bankruptcy risk.
Panel B: AI-related variables	
$\mathrm{AI}_{\mathrm{Asset}}$	An asset-based measure of AI investment, calculated as the natural logarithm of the sum of a firm's AI-related intangible and fixed assets, scaled by its total sales.
$\mathrm{AI}_{\mathrm{Text}}$	A text-based measure of AI investment, the natural logarithm of this frequency of AI-related terms in the firm's annual report.
$\mathrm{AI}_{\mathrm{LLM}}$	An LLM-based measure of AI investment, the natural logarithm of the sum of AI investment scores across firm's annual report segments, weighted by their corresponding confidence scores.
$\mathrm{AI}_{\mathrm{Labor}}$	An labor-based measure of AI investment, the ratio of the number recruits of AI-related jobs to the number recruits of total jobs
$\mathrm{MarketDrivenAI}_{\mathrm{Asset;Text;LLM;Labor}}$	β_{Market} in the regression of model 1, defined as the AI investment due to market trends, measured using the asset-based, text-based, LLM-based and labor-based AI investment measure.
$\mathrm{FirmSpecificAI}_{\mathrm{Asset;Text;LLM;Labor}}$	Stochastic error term in the regression of model 1, defined as the AI investment beyond market trends, measured using the asset-based, text-based, LLM-based and labor-based AI investment measure.
Panel C: Control variables	
Size	Logarithm of firm's total assets.
ROA	Firm's return on assets.
Leverage	The ratio of the sum of short-term and long-term debt to the book value of assets.
EPS	Earnings per share.
TobinQ	Tobin's Q value.
Age	Logarithm of the listed age of a company.
AssetGrowthRate	Annual growth rate of total assets.
Panel D: Channel test variables	
CF	Firm's operating cash flows, scaled by total assets.
CFvol	Cash flow volatility, measured as the standard deviation of CF over the next five years.
ROAvol	ROA volatility, measured as the standard deviation of ROA over the next five years.
DataRich	A dummy variable which equals 1 if a firm gains access to sufficiently rich and high-quality data, and zero otherwise.
KnowledgeStock	Firm's knowledge stock level, constructed by Ewens et al. (2024)' method.
RDStock	Firm's R&D stock level, constructed by Aghion et al. (2013)' method.
Panel E: Motivation test variables	

Continued on next page

Table A1 Continued from previous page

Variable Name	Definition
RiskPreference	Volatility of ROA over past five years.
CEOOverconfidence	Proxy for CEO overconfidence (Schrand and Zechman, 2012).
NumberOfFirms	Logarithm of number of firms in the industry.
RecHHI	Logarithm of reciprocal of firm's Herfindahl-Hirschman Index.
Panel F: Robustness test variables	
ResearchOutcomeAI	Logarithm of number of AI-related papers from CNKI/WOS, 2006-2023.
$\text{AltMarketDrivenAI}_{\text{Asset;Text;LLM;Labor}}$	Alternative measures of $\text{MarketDrivenAI}_{\text{Asset;Text;LLM;Labor}}$. These measures are constructed by substituting the market average with the industry mean in the decomposition framework specified in model 1
$\text{AltFirmSpecificAI}_{\text{Asset;Text;LLM;Labor}}$	Alternative measures of $\text{FirmSpecificAI}_{\text{Asset;Text;LLM;Labor}}$. These measures are constructed by substituting the market average with the industry mean in the decomposition framework specified in model 1
$\text{IndustryAverageAI}_{\text{Asset;Text;LLM;Labor}}$	Industry average of AI investment, alternative measures of $\text{MarketDrivenAI}_{\text{Asset;Text;LLM;Labor}}$.
$\text{IndustryDeviationAI}_{\text{Asset;Text;LLM;Labor}}$	Alternative measures of $\text{FirmSpecificAI}_{\text{Asset;Text;LLM;Labor}}$. These measures are the difference between firm overall AI investment and industry average AI.
ZScore	Altman's Z-score.
OScore	Negative Ohlson's O-score.

Table A2

Dictionary of AI terms

AI	Cloud Computing	Intelligence	Natural Language
Artificial Intelligence	Data Mining	Human-computer Interaction	Neural Network
Speech Recognition	Deep Learning	Knowledge Graph	Robot
Big Data	Image Recognition	Machine Learning	Virtual Reality

Table A3

Correlation Matrix for AI Investment Measures

Note: This table reports pairwise correlations among four firm-level AI investment measures: asset-based (AI_{Asset}), text-based (AI_{Text}), LLM-based (AI_{LLM}), and labor-based (AI_{Labor}). Panel A reports the pooled correlation coefficients across the full sample. Panel B presents the average of firm-level correlation coefficients, and Panel C presents the median of firm-level correlation coefficients. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Correlation matrix for AI Measures

	AI_{Asset}	AI_{Text}	AI_{LLM}	AI_{Labor}
AI_{Asset}	1			
AI_{Text}	0.301***	1		
AI_{LLM}	0.289***	0.668***	1	
AI_{Labor}	0.241***	0.502***	0.397***	1

Panel B: Mean of correlation coefficients across all firms

	AI_{Asset}	AI_{Text}	AI_{LLM}	AI_{Labor}
AI_{Asset}	1.000			
AI_{Text}	0.449***	1.000		
AI_{LLM}	0.404***	0.591***	1.000	
AI_{Labor}	0.344***	0.386***	0.357***	1.000

Panel C: Median of correlation coefficients across all firms

	AI_{Asset}	AI_{Text}	AI_{LLM}	AI_{Labor}
AI_{Asset}	1.000			
AI_{Text}	0.465***	1.000		
AI_{LLM}	0.394***	0.660***	1.000	
AI_{Labor}	0.299***	0.381***	0.328***	1.000

Distance to Default

We measure default risk by the distance to default (DD) based on Merton (1974). Suppose that the firm asset value V follows a geometric Brownian motion:

$$\mathrm{d}V = \mu_V V\,\mathrm{d}t + \sigma_V V\,\mathrm{d}W, \tag{A.1}$$

where μ_V is the expected firm asset return and σ_V is the volatility.

The firm issues a zero-coupon bond with a face value D and maturity of T years. The firm equity value E can be viewed as a call option on the firm asset value V, with a strike price equal to D. The equity value is given by

$$E = V\,\Phi(\delta_1) - e^{-rT} D\,\Phi(\delta_2), \tag{A.2}$$

with

$$\delta_1 = \frac{\ln(V/D) + (r + 0.5\,\sigma_V^2)T}{\sigma_V\sqrt{T}}, \quad \delta_2 = \delta_1 - \sigma_V\sqrt{T}, \tag{A.3}$$

where r is the risk-free interest rate, and $\Phi(\cdot)$ is the cumulative standard normal distribution function.

Applying Ito's lemma to the equity value, we obtain

$$\sigma_E = \frac{V}{E}\frac{\partial E}{\partial V}\sigma_V. \tag{A.4}$$

Because $\partial E/\partial V = \Phi(\delta_1)$, Equation (A.4) can be rewritten as

$$\sigma_E = \frac{V}{E}\Phi(\delta_1)\,\sigma_V. \tag{A.5}$$

Default is assumed to occur at maturity T if the asset value falls below the debt level, i.e., when $\{V_T < D\}$. Since V_T follows a lognormal distribution, the probability of default can be expressed as

$$P(V_T < D) = \Phi(-DD), \tag{A.6}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal variable, and DD is defined as

$$\text{DD} = \frac{\ln\left(\frac{V}{D}\right) + \left(\mu - \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}} \;\approx\; \frac{V - D}{V\,\sigma_V}. \tag{A.7}$$

The approximation in (A.7) follows from the fact that $\mu - \frac{1}{2}\sigma^2$ is typically small over short periods in practice, together with a first-order Taylor expansion.

In Equation (A.7), V and σ_V are unknown. Following Bharath and Shumway (2008), we solve for these variables using an iterative procedure based on Equations (A.2) and (A.5) to calculate the annual DD for all firms from 2006 to 2023. The face value of debt D is defined as the sum of current liabilities and half of non-current liabilities, while the market equity value is taken as the firm's market capitalization. The one-year historical daily return is used to estimate σ_E, and the one-year government bond rate serves as the risk-free rate. The resulting V and σ_V are then used to compute the implied distance to default (DD) from Equation (A.7), with lower DD indicating higher default risk.

Table A4

Impact of Industry-average and Industry-deviation AI on Firm Default Risk

Note: This table reports regression estimates of a firm's default risk on industry-average and industry-deviation AI investment. The dependent variable *DD* is one-year ahead distance-to-default. The key independent variables are asset-based, text-based, LLM-based and labor-based industry-average and industry-deviation AI investment. The control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Industry-average AI investment and firm default risk				
$\text{IndustryAverageAI}_{\text{Asset}}$	-0.003			
	(-0.30)			
$\text{IndustryAverageAI}_{\text{Text}}$		0.037***		
		(6.97)		
$\text{IndustryAverageAI}_{\text{LLM}}$			0.092***	
			(14.64)	
$\text{IndustryAverageAI}_{\text{Labor}}$				0.115***
				(13.26)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,473	38,187	33,473	26,283
R^2	0.300	0.316	0.306	0.374
Panel B: Industry-deviation AI investment and firm default risk				
$\text{IndustryDeviationAI}_{\text{Asset}}$	-0.013***			
	(-4.41)			
$\text{IndustryDeviationAI}_{\text{Text}}$		-0.005*		
		(-1.68)		
$\text{IndustryDeviationAI}_{\text{LLM}}$			-0.003	
			(-1.14)	
$\text{IndustryDeviationAI}_{\text{Labor}}$				-0.005**
				(-2.32)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	33,473	38,187	33,473	26,283
R^2	0.532	0.568	0.532	0.589

Table A5
AI decomposition, Knowledge Stock and Firm Default Risk

Note: This table reports regression estimates analyzing how a firm's knowledge stock moderates the relationship between its default risk and market-driven, firm-specific AI investment. The dependent variable DD is one-year ahead distance-to-default. $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $KnowledgeStock$ is firm's knowledge stock level, constructed by Ewens et al. (2024)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Market-driven AI investment, knowledge stock and DD				
$\text{MarketDrivenAI}_{\text{Asset}}$ × KnowledgeStock	0.330***			
	(5.16)			
$\text{MarketDrivenAI}_{\text{Text}}$ × KnowledgeStock		0.098***		
		(4.53)		
$\text{MarketDrivenAI}_{\text{LLM}}$ × KnowledgeStock			0.184***	
			(5.71)	
$\text{MarketDrivenAI}_{\text{Labor}}$ × KnowledgeStock				-0.022
				(-1.17)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.042***			
	(7.71)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.075***		
		(11.62)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.121***	
			(18.89)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.102***
				(23.60)
KnowledgeStock	-0.000***	-0.000***	-0.000***	0.000
	(-5.02)	(-4.34)	(-5.86)	(1.18)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.328	0.333	0.340	0.401
Panel B: Firm-specific AI investment, knowledge stock and DD				
$\text{FirmSpecificAI}_{\text{Asset}}$ × KnowledgeStock	0.000			
	(0.16)			
$\text{FirmSpecificAI}_{\text{Text}}$ × KnowledgeStock		0.002		
		(0.65)		
$\text{FirmSpecificAI}_{\text{LLM}}$ × KnowledgeStock			0.003	
			(1.31)	
$\text{FirmSpecificAI}_{\text{Labor}}$ × KnowledgeStock				-0.002
				(-1.28)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.008***			
	(-4.43)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.006***		
		(-2.68)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.004*	
			(-1.87)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.002
				(-0.90)
KnowledgeStock	-0.000	-0.000	-0.000	0.000
	(-0.28)	(-0.46)	(-0.39)	(0.24)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.542	0.547	0.542	0.589

Table A6
AI decomposition, R&D Stock and Firm Default Risk

Note: This table reports regression estimates analyzing how a firm's R&D stock moderates the relationship between its default risk and market-driven, firm-specific AI investment. The dependent variable DD is one-year ahead distance-to-default. $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $RDStock$ is firm's R&D stock level, constructed by Aghion et al. (2013)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. Regressions in panel A control for firm fixed effects and regressions in panel B control for year and firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	DD	DD	DD	DD
Panel A: Market-driven AI investment, R&D stock and DD				
$\text{MarketDrivenAI}_{\text{Asset}} \times \text{RDStock}$	0.308***			
	(5.66)			
$\text{MarketDrivenAI}_{\text{Text}} \times \text{RDStock}$		0.081***		
		(4.80)		
$\text{MarketDrivenAI}_{\text{LLM}} \times \text{RDStock}$			0.151***	
			(6.64)	
$\text{MarketDrivenAI}_{\text{Labor}} \times \text{RDStock}$				-0.014
				(-0.83)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.042***			
	(7.55)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.075***		
		(11.61)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.120***	
			(18.77)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.102***
				(23.69)
RDStock	-0.000***	-0.000***	-0.000***	0.000
	(-5.59)	(-4.65)	(-6.91)	(0.69)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.328	0.333	0.340	0.401
Panel B: Firm-specific AI investment, R&D stock and DD				
$\text{FirmSpecificAI}_{\text{Asset}} \times \text{RDStock}$	-0.001			
	(-0.30)			
$\text{FirmSpecificAI}_{\text{Text}} \times \text{RDStock}$		0.003		
		(1.03)		
$\text{FirmSpecificAI}_{\text{LLM}} \times \text{RDStock}$			0.005**	
			(2.13)	
$\text{FirmSpecificAI}_{\text{Labor}} \times \text{RDStock}$				-0.001
				(-0.37)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.005**			
	(-2.24)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.004*		
		(-1.77)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.004	
			(-1.58)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.003
				(-1.06)
RDStock	0.000***	0.000**	0.000**	0.000***
	(2.59)	(2.28)	(2.15)	(3.01)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.323	0.327	0.323	0.383

Table A7
Market-driven AI, Knowledge Stock and Cash Flow

Note: This table reports regression estimates analyzing how a firm's knowledge stock moderates the relationship between its cash flow and market-driven AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $KnowledgeStock$ is firm's knowledge stock level, constructed by Ewens et al. (2024)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, knowledge stock and cash flow level				
	CF	CF	CF	CF
$\text{MarketDrivenAI}_{\text{Asset}} \times$ KnowledgeStock	-0.004			
	(-0.59)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ KnowledgeStock		0.002		
		(0.73)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ KnowledgeStock			-0.002	
			(-0.49)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ KnowledgeStock				0.004
				(1.35)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.004***			
	(3.53)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.007***		
		(4.52)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.006***	
			(4.11)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.002**
				(2.37)
KnowledgeStock	0.000	-0.000	0.000	-0.000
	(0.65)	(-0.97)	(0.49)	(-1.09)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,575	27,405	26,576	22,264
R^2	0.450	0.441	0.450	0.481
Panel B: Market-driven AI investment, knowledge stock and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{MarketDrivenAI}_{\text{Asset}} \times$ KnowledgeStock	-0.007**			
	(-2.15)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ KnowledgeStock		-0.002		
		(-1.64)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ KnowledgeStock			-0.003**	
			(-2.31)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ KnowledgeStock				-0.003*
				(-1.82)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.001			
	(1.06)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.001		
		(-0.60)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.000	
			(-0.57)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.001
				(-1.36)
KnowledgeStock	0.000*	0.000	0.000*	0.000
	(1.91)	(1.10)	(1.95)	(1.51)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.634	0.618	0.634	0.707

Table A8
Market-driven AI, Knowledge Stock and ROA

Note: This table reports regression estimates analyzing how a firm's knowledge stock moderates the relationship between its ROA and market-driven AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $KnowledgeStock$ is firm's knowledge stock level, constructed by Ewens et al. (2024)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, knowledge stock and ROA				
	ROA	ROA	ROA	ROA
$\text{MarketDrivenAI}_{\text{Asset}} \times$ KnowledgeStock	-0.011**			
	(-2.42)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ KnowledgeStock		-0.005***		
		(-2.67)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ KnowledgeStock			-0.007***	
			(-3.18)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ KnowledgeStock				0.004
				(1.57)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.010***			
	(7.20)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.011***		
		(5.94)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.011***	
			(6.26)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.007***
				(6.38)
KnowledgeStock	0.000***	0.000***	0.000***	-0.000
	(2.88)	(3.30)	(3.74)	(-0.12)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.454	0.453	0.455	0.466
Panel B: Market-driven AI investment, knowledge stock and ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$\text{MarketDrivenAI}_{\text{Asset}} \times$ KnowledgeStock	-0.006*			
	(-1.82)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ KnowledgeStock		-0.003**		
		(-2.27)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ KnowledgeStock			-0.004**	
			(-2.32)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ KnowledgeStock				-0.002
				(-0.92)
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.002			
	(-1.61)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.007***		
		(-5.04)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.006***	
			(-4.55)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.003***
				(-4.01)
KnowledgeStock	0.000	0.000	0.000	-0.000
	(0.32)	(0.16)	(0.06)	(-0.34)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.655	0.641	0.656	0.774

Table A9
Market-driven AI, R&D Stock and Cash Flow

Note: This table reports regression estimates analyzing how a firm's R&D stock moderates the relationship between its cash flow and market-driven AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $RDStock$ is firm's R&D stock level, constructed by Aghion et al. (2013)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, R&D stock and cash flow level				
	CF	CF	CF	CF
$MarketDrivenAI_{Asset} \times RDStock$	-0.001 (-0.20)			
$MarketDrivenAI_{Text} \times RDStock$		0.001 (0.43)		
$MarketDrivenAI_{LLM} \times RDStock$			-0.001 (-0.37)	
$MarketDrivenAI_{Labor} \times RDStock$				0.003 (1.21)
$MarketDrivenAI_{Asset}$	0.004*** (3.54)			
$MarketDrivenAI_{Text}$		0.007*** (4.50)		
$MarketDrivenAI_{LLM}$			0.006*** (4.12)	
$MarketDrivenAI_{Labor}$				0.002** (2.42)
RDStock	0.000 (0.27)	-0.000 (-0.66)	0.000 (0.37)	-0.000 (-0.96)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,575	27,405	26,576	22,264
R^2	0.450	0.441	0.450	0.481
Panel B: Market-driven AI investment, R&D stock and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$MarketDrivenAI_{Asset} \times RDStock$	-0.007** (-2.50)			
$MarketDrivenAI_{Text} \times RDStock$		-0.002* (-1.70)		
$MarketDrivenAI_{LLM} \times RDStock$			-0.003** (-2.31)	
$MarketDrivenAI_{Labor} \times RDStock$				-0.002 (-1.34)
$MarketDrivenAI_{Asset}$	0.001 (1.07)			
$MarketDrivenAI_{Text}$		-0.001 (-0.64)		
$MarketDrivenAI_{LLM}$			-0.000 (-0.57)	
$MarketDrivenAI_{Labor}$				-0.001 (-1.47)
RDStock	0.000** (2.29)	0.000 (1.21)	0.000** (1.99)	0.000 (1.09)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.634	0.618	0.634	0.707

Table A10
Market-driven AI, R&D Stock and ROA

Note: This table reports regression estimates analyzing how a firm's R&D stock moderates the relationship between its ROA and market-driven AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $MarketDrivenAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based market-driven AI investment. $RDStock$ is firm's R&D stock level, constructed by Aghion et al. (2013)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Market-driven AI investment, R&D stock and ROA				
	ROA	ROA	ROA	ROA
$\text{MarketDrivenAI}_{\text{Asset}} \times$ RDStock	-0.007 (-1.43)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ RDStock		-0.004** (-2.25)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ RDStock			-0.006** (-2.56)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ RDStock				0.003 (1.29)
$\text{MarketDrivenAI}_{\text{Asset}}$	0.010*** (7.23)			
$\text{MarketDrivenAI}_{\text{Text}}$		0.011*** (6.00)		
$\text{MarketDrivenAI}_{\text{LLM}}$			0.011*** (6.32)	
$\text{MarketDrivenAI}_{\text{Labor}}$				0.007*** (6.44)
RDStock	0.000** (2.08)	0.000*** (3.15)	0.000*** (3.41)	0.000 (0.45)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.454	0.453	0.455	0.466
Panel B: Market-driven AI investment, R&D Stock and ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$\text{MarketDrivenAI}_{\text{Asset}} \times$ RDStock	-0.007** (-2.42)			
$\text{MarketDrivenAI}_{\text{Text}} \times$ RDStock		-0.003** (-2.25)		
$\text{MarketDrivenAI}_{\text{LLM}} \times$ RDStock			-0.003** (-2.34)	
$\text{MarketDrivenAI}_{\text{Labor}} \times$ RDStock				-0.000 (-0.22)
$\text{MarketDrivenAI}_{\text{Asset}}$	-0.002* (-1.68)			
$\text{MarketDrivenAI}_{\text{Text}}$		-0.007*** (-5.12)		
$\text{MarketDrivenAI}_{\text{LLM}}$			-0.006*** (-4.65)	
$\text{MarketDrivenAI}_{\text{Labor}}$				-0.003*** (-4.15)
RDStock	0.000 (0.72)	-0.000 (-0.42)	-0.000 (-0.40)	-0.000 (-1.64)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.655	0.641	0.656	0.774

Table A11
Firm-specific AI, Knowledge Stock and Cash Flow

Note: This table reports regression estimates analyzing how a firm's knowledge stock moderates the relationship between its cash flow and firm-specific AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $KnowledgeStock$ is firm's knowledge stock level, constructed by Ewens et al. (2024)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm and year fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, knowledge stock and cash flow level				
	CF	CF	CF	CF
$\text{FirmSpecificAI}_{\text{Asset}}$ × KnowledgeStock	-0.000			
	(-0.41)			
$\text{FirmSpecificAI}_{\text{Text}}$ × KnowledgeStock		0.001		
		(1.32)		
$\text{FirmSpecificAI}_{\text{LLM}}$ × KnowledgeStock			0.000	
			(0.69)	
$\text{FirmSpecificAI}_{\text{Labor}}$ × KnowledgeStock				0.000
				(0.63)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.000			
	(-0.39)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001**		
		(-2.33)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.001***	
			(-3.46)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000
				(-0.55)
KnowledgeStock	0.000	-0.000	0.000	0.000
	(0.15)	(-0.99)	(0.33)	(0.20)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	26,575	27,405	26,576	22,264
R^2	0.458	0.450	0.459	0.489
Panel B: Firm-specific AI investment, knowledge stock and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{FirmSpecificAI}_{\text{Asset}}$ × KnowledgeStock	0.000			
	(0.44)			
$\text{FirmSpecificAI}_{\text{Text}}$ × KnowledgeStock		0.000		
		(0.13)		
$\text{FirmSpecificAI}_{\text{LLM}}$ × KnowledgeStock			0.000*	
			(1.66)	
$\text{FirmSpecificAI}_{\text{Labor}}$ × KnowledgeStock				-0.000
				(-0.19)
$\text{FirmSpecificAI}_{\text{Asset}}$	0.000			
	(0.55)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.000		
		(-0.56)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000	
			(-0.77)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000
				(-0.55)
KnowledgeStock	0.000	-0.000	0.000	0.000
	(0.34)	(-0.08)	(0.14)	(0.61)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.642	0.626	0.642	0.713

Table A12
Firm-specific AI, Knowledge Stock and ROA

Note: This table reports regression estimates analyzing how a firm's knowledge stock moderates the relationship between its ROA and firm-specific AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $KnowledgeStock$ is firm's knowledge stock level, constructed by Ewens et al. (2024)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm and year fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, knowledge stock and ROA				
	ROA	ROA	ROA	ROA
$FirmSpecificAI_{Asset} \times KnowledgeStock$	-0.002*** (-2.79)			
$FirmSpecificAI_{Text} \times KnowledgeStock$		0.000 (0.77)		
$FirmSpecificAI_{LLM} \times KnowledgeStock$			-0.000 (-1.10)	
$FirmSpecificAI_{Labor} \times KnowledgeStock$				-0.000 (-0.85)
$FirmSpecificAI_{Asset}$	-0.001* (-1.79)			
$FirmSpecificAI_{Text}$		-0.001* (-1.80)		
$FirmSpecificAI_{LLM}$			-0.000 (-0.69)	
$FirmSpecificAI_{Labor}$				-0.001 (-1.27)
KnowledgeStock	0.000** (2.25)	0.000* (1.83)	0.000*** (2.89)	0.000*** (2.72)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.457	0.457	0.458	0.468
Panel B: Firm-specific AI investment, knowledge stock and ROA volatility				
	ROAvol	ROAvol	ROAvol	ROAvol
$FirmSpecificAI_{Asset} \times KnowledgeStock$	0.000 (0.38)			
$FirmSpecificAI_{Text} \times KnowledgeStock$		0.000 (1.40)		
$FirmSpecificAI_{LLM} \times KnowledgeStock$			0.000 (1.33)	
$FirmSpecificAI_{Labor} \times KnowledgeStock$				-0.000 (-0.31)
$FirmSpecificAI_{Asset}$	0.000 (0.17)			
$FirmSpecificAI_{Text}$		0.000 (0.07)		
$FirmSpecificAI_{LLM}$			0.000 (0.31)	
$FirmSpecificAI_{Labor}$				0.000 (0.79)
KnowledgeStock	-0.000** (-2.55)	-0.000*** (-2.66)	-0.000*** (-2.67)	-0.000 (-0.91)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.674	0.659	0.674	0.780

Table A13
Firm-specific AI, R&D Stock and Cash Flow

Note: This table reports regression estimates analyzing how a firm's R&D stock moderates the relationship between its cash flow and firm-specific AI investment. The dependent variables are cash flow (CF) and cash flow volatility ($CFvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. *RDStock* is firm's R&D stock level, constructed by Aghion et al. (2013)' method. The control variables include *Size*, *ROA*, *Leverage*, *EPS*, *TobinQ*, *Age*, and *AssetGrowthRate*. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm and year fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, R&D stock and cash flow level				
	CF	CF	CF	CF
$\text{FirmSpecificAI}_{\text{Asset}} \times$ RDStock	-0.000 (-0.61)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ RDStock		0.001* (1.89)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ RDStock			0.000 (1.04)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ RDStock				0.000 (0.46)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.000 (-0.39)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001** (-2.15)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.001*** (-3.52)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000 (-0.46)
RDStock	0.000 (1.01)	-0.000 (-0.23)	0.000 (1.12)	0.000 (1.20)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,575	27,405	26,576	22,264
R^2	0.449	0.441	0.450	0.481
Panel B: Firm-specific AI investment, R&D stock and cash flow volatility				
	CFvol	CFvol	CFvol	CFvol
$\text{FirmSpecificAI}_{\text{Asset}} \times$ RDStock	0.000 (0.44)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ RDStock		0.000 (0.09)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ RDStock			0.000 (1.32)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ RDStock				-0.000 (-0.55)
$\text{FirmSpecificAI}_{\text{Asset}}$	0.000 (0.32)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.000 (-0.27)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000 (-0.11)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.000 (-0.22)
RDStock	-0.000 (-0.95)	-0.000 (-1.27)	-0.000 (-1.14)	-0.000 (-1.16)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.634	0.618	0.634	0.707

Table A14
Firm-specific AI, R&D Stock and ROA

Note: This table reports regression estimates analyzing how a firm's R&D stock moderates the relationship between its ROA and firm-specific AI investment. The dependent variables are ROA (ROA) and ROA volatility ($ROAvol$). $FirmSpecificAI_{Asset;Text;LLM;Labor}$ are asset-based, text-based, LLM-based and labor-based firm-specific AI investment. $RDStock$ is firm's R&D stock level, constructed by Aghion et al. (2013)' method. The control variables include $Size$, ROA, $Leverage$, EPS, $TobinQ$, Age, and $AssetGrowthRate$. Detailed descriptions of all the variables are provided in Table A1 in the Appendix. All regressions control for firm and year fixed effects. The standard error clusters at the firm level. The t-statistics are reported in brackets. Single, double, and triple * indicate significance at the indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
Panel A: Firm-specific AI investment, R&D stock and ROA				
	ROA	ROA	ROA	ROA
$\text{FirmSpecificAI}_{\text{Asset}} \times$ RDStock	-0.002***			
	(-2.97)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ RDStock		0.001		
		(1.36)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ RDStock			-0.000	
			(-0.89)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ RDStock				-0.000
				(-0.79)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.001			
	(-1.49)			
$\text{FirmSpecificAI}_{\text{Text}}$		-0.001*		
		(-1.82)		
$\text{FirmSpecificAI}_{\text{LLM}}$			-0.000	
			(-0.82)	
$\text{FirmSpecificAI}_{\text{Labor}}$				-0.001
				(-1.29)
RDStock	0.000***	0.000***	0.000***	0.000***
	(4.54)	(4.07)	(4.98)	(4.96)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	26,589	27,419	26,590	22,278
R^2	0.453	0.452	0.454	0.465
Panel B: Firm-specific AI investment, R&D stock and ROA volatility				
	ROAVOL	ROAVOL	ROAVOL	ROAVOL
$\text{FirmSpecificAI}_{\text{Asset}} \times$ RDStock	0.000			
	(0.49)			
$\text{FirmSpecificAI}_{\text{Text}} \times$ RDStock		0.000		
		(0.75)		
$\text{FirmSpecificAI}_{\text{LLM}} \times$ RDStock			0.000	
			(0.20)	
$\text{FirmSpecificAI}_{\text{Labor}} \times$ RDStock				-0.000
				(-0.73)
$\text{FirmSpecificAI}_{\text{Asset}}$	-0.000			
	(-0.24)			
$\text{FirmSpecificAI}_{\text{Text}}$		0.000		
		(0.41)		
$\text{FirmSpecificAI}_{\text{LLM}}$			0.000	
			(1.38)	
$\text{FirmSpecificAI}_{\text{Labor}}$				0.000
				(1.20)
RDStock	-0.000***	-0.000***	-0.000***	-0.000***
	(-5.73)	(-5.55)	(-5.69)	(-3.74)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	22,640	23,465	22,640	18,335
R^2	0.657	0.641	0.657	0.777