

Can Digital Transformation be a Driver of R&D Intensity? Evidence from China

Tingqian Pu^{1*}

¹*School of Economics and Finance, Guizhou University of Commerce, China.*

Abstract: **Research Question:** Does digital transformation (DT) enhance firms' R&D intensity (RDI)? If so, which subdimensions of DT drive this effect? Does the DT-RDI relationship exhibit heterogeneity based on ownership, profitability, technological orientation, and geographic location? **Motivation:** While prior research has explored the broader economic impacts of DT, there is a gap in understanding which DT subdimensions drive RDI and how firm characteristics (e.g., ownership, profitability, technological orientation, and geographic location) shape this relationship. **Idea:** This study investigates whether DT is a driving factor of RDI through the lenses of resource-based theory, agency theory, and innovation diffusion theory. **Data:** The sample consists of 3,920 Chinese publicly listed firms from 2012 to 2021, with DT data sourced from annual report text analysis. **Method/Tools:** A series of panel analyses, including grouped t-tests and two-way fixed effects, were conducted. Robust standard errors clustered at the firm level were applied to address heterogeneity and cross-sectional dependence. Endogeneity tests using PSM, DID, and GMM estimations confirm the robustness of the findings. **Findings:** This study finds a positive relationship between DT and RDI, with key drivers being subdimensions such as artificial intelligence, cloud computing, and big data. The positive effect is stronger in state-owned, profitable, high-tech firms, and those located in eastern and central China. **Contributions:** First, this research expands the limited literature on how DT influences R&D strategies, focusing on R&D decision-making rather than just operational outcomes. Second, it unveils the role of DT's subdimensions in shaping R&D, offering insights into the strategic implications of the digital era. Third, by highlighting the heterogeneous effects of DT on RDI, the study provides actionable recommendations for enhancing firms' R&D competitiveness through tailored digital strategies.

Keywords: Digital transformation, R&D intensity, artificial intelligence, cloud computing, big data, China.

JEL Classification: L86, O25, O32

1. Introduction

In the face of intensifying global competition, the research and development (R&D) intensity of firms has become a crucial metric for evaluating their innovation capabilities and market competitiveness (Lin *et al.*, 2021; Zheng *et al.*, 2023). R&D intensity not only reflects a company's commitment to technological advancement and new product

* Corresponding author: Tingqian Pu. Tel.: +0851-84870793.

Email: tq_pu@outlook.com

Received 20 Sep 2024; Final revised 27 Dec 2024; Accepted 6 Mar 2025; Available online 31 Mar 2025.

To link to this article: https://www.mfa.com.my/cmrv33_i1_a1/

© Malaysian Finance Association, 2025. This work is licensed under the terms of the Creative Commons Attribution (CC BY) (<http://creativecommons.org/licenses/by/4.0/>).

development but also determines its ability to maintain long-term competitive advantage in a rapidly evolving market environment (Aksoy *et al.*, 2024). This issue takes on heightened importance in the context of China, where firms are navigating the transition from being the "world's factory" to becoming an "innovation-driven economy." As global technological rivalry escalates and the domestic economy undergoes restructuring, the R&D intensity of firms is closely tied to the country's broader goal of achieving technological self-sufficiency (Li *et al.*, 2024). Enhancing R&D investment is thus pivotal for companies to improve technological capabilities, foster innovation, and reduce reliance on foreign technologies, ultimately driving domestic breakthroughs and expanding global market influence.

The rise of digital technologies is reshaping the global landscape of R&D. With the rapid proliferation of technologies such as artificial intelligence, cloud computing, and big data, digital transformation is fundamentally altering how firms operate and allocate resources. Digitalization not only enhances operational efficiency but also transforms decision-making processes related to R&D (Vial, 2019). For instance, AI and data analytics can optimize R&D workflows, shortening development cycles and reducing costs, while cloud computing supports distributed collaboration and data sharing (Butschan *et al.*, 2019). These technologies significantly enhance firms' agility in global markets, forcing them to reconsider how they allocate and manage R&D resources. However, whether digital transformation alters innovation patterns in a way that strengthens R&D intensity remains a question that warrants deeper investigation.

While prior studies have explored the broad impact of digital transformation on innovation (e.g., Hao *et al.*, 2023), many of these studies treat digital transformation as a unidimensional construct, paying insufficient attention to the distinct roles of its subdimensions, such as artificial intelligence, cloud computing, and big data. Moreover, existing research rarely examines how digital transformation influences R&D strategies across various firm characteristics—such as ownership structures, profitability levels, industry classifications, and geographical regions—despite growing evidence that organizational context significantly shapes the outcomes of technology adoption (Sun *et al.*, 2024). This gap is critical to address because R&D intensity is a key determinant of corporate innovation, particularly in China's journey toward an innovation-driven economy. By systematically exploring these subdimensions and their heterogeneous effects, researchers and practitioners can gain deeper insights into how digital transformation supports or constrains R&D investment across diverse firm contexts, ultimately guiding more effective innovation strategies.

Given the critical role that R&D intensity plays in driving corporate innovation, this study aims to fill this gap by systematically analyzing whether and how digital transformation impacts R&D intensity. Using a dataset of Chinese listed firms from 2012 to 2021, this paper finds a positive relationship between digital transformation and R&D intensity. Further analysis reveals that subdimensions of digital transformation, such as artificial intelligence, cloud computing, and big data, are key drivers of this effect. Moreover, this positive impact is more pronounced among state-owned enterprises, profitable firms, high-tech companies, and firms located in China's eastern and central regions.

The contributions of this study are threefold. First, while existing research on the economic consequences of digital transformation largely focuses on its effects on operational performance and productivity, few studies examine its impact on R&D strategy and decision-making. This research not only extends the literature on the economic outcomes of digital transformation but also enriches the exploration of factors that influence corporate R&D. Second, although many studies have examined digital transformation, few have explored the relationship between its subdimensions and R&D strategy in the context

of the new technological revolution. By doing so, this study sheds light on the "black box" of how digital transformation affects R&D decisions. Third, by analyzing the heterogeneous effects of digital transformation on R&D intensity, this research provides more targeted insights that can inform the digital transformation strategies of Chinese firms, offering practical policy recommendations to further enhance their competitiveness.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature and presents the hypotheses. Section 3 outlines the methodology used in this study. Section 4 discusses the findings, robustness checks, and additional analyses. Finally, Section 5 concludes the paper with a summary of key insights.

2. Theory Foundation and Hypothesis Development

2.1 The View of Resource-Based

According to the Resource-Based Theory (RBT), differences in firms' resources are a critical factor in explaining variations in profitability (Das and Teng, 2000). Firms with superior resources can achieve a competitive advantage, enabling them to capture economic rents (Mahoney, 2001). In the modern business landscape, digital technologies—such as big data, artificial intelligence (AI), and cloud computing—have become key resources that allow firms to optimize resource allocation, enhance productivity, and reduce operational costs (Giustiziero *et al.*, 2021). For instance, big data can analyze vast amounts of information to identify market trends, enabling firms to develop more precise R&D strategies, while AI accelerates product design and testing in the innovation process (Sestino *et al.*, 2020).

On the one hand, digital transformation strengthens a firm's ability to integrate and utilize resources. This involves not only the adoption of new technological resources but also the optimization of existing ones. For example, through digital tools, firms can streamline various stages of the supply chain, minimize resource waste, and enhance R&D efficiency (Bejlegaard *et al.*, 2021). By effectively acquiring and utilizing these scarce resources, firms can increase their investment in innovation and boost their R&D intensity.

On the other hand, digital transformation provides firms with opportunities to develop proprietary technologies, such as in-house big data platforms, AI algorithms, or IoT solutions (Mariani *et al.*, 2023). These technologies serve as rare and irreplaceable resources that offer long-term technical advantages and elevate R&D intensity. Such proprietary technologies not only enhance the firm's R&D capabilities but also create a competitive advantage that is difficult for rivals to replicate. These scarce digital resources not only directly strengthen a firm's R&D intensity but also contribute to its ability to innovate products and services, leading to a competitive edge in the market.

2.2 The View of Agency

One of the core challenges in the principal-agent problem is information asymmetry, where agents possess more knowledge about the firm's operations than principals, making it difficult for principals to effectively monitor the actions of agents (Eisenhardt, 1989; Hooy *et al.*, 2019). According to the Agency Theory (AT), digital transformation can mitigate information asymmetry by enhancing transparency and enabling data visualization, thus improving corporate governance. Specifically, digital tools such as big data analytics, real-time monitoring systems, and intelligent management platforms enable firms to achieve greater transparency in their operations (Matheus *et al.*, 2021). These technologies allow top management to monitor company performance in real time, increasing oversight of the managerial team and reducing the likelihood of agents deviating from the firm's objectives. For example, digital tools can track the progress of R&D projects, monitor the use of financial resources, and assess team efficiency, ensuring that resources are allocated

efficiently and reducing managerial discretion in resource allocation (Marion and Fixson, 2020). This heightened transparency allows principals to better supervise and evaluate managerial decisions, helping to reduce agency problems and strengthening support for innovation and R&D efforts.

Moreover, agency problems often lead to increased agency costs, including monitoring costs, incentive costs, and opportunity costs. Digital transformation can enhance the efficiency of innovation management, reducing these agency costs and freeing up more resources for R&D investment. For instance, digital technologies can streamline innovation management processes, reducing unnecessary bureaucracy and improving communication across departments, leading to more efficient allocation of resources for R&D (Urbinati *et al.*, 2020). In traditional R&D management models, complex approval processes and poor interdepartmental collaboration often result in wasted innovation resources and higher opportunity costs. By leveraging digital transformation, firms can utilize smart management systems and collaboration platforms to facilitate real-time cross-departmental cooperation, accelerating decision-making and minimizing delays (Gupta *et al.*, 2019). This not only reduces the agency costs associated with innovation management but also enhances overall innovation efficiency and strengthens R&D intensity.

2.3 The View of Innovation Diffusion

The Diffusion of Innovation Theory (DIT) posits that innovation dissemination is a social process involving the spread, adoption, and eventual widespread use of new technologies (Dearing, 2009). Digital transformation refers to the fundamental changes that businesses undergo when they leverage digital technologies to reshape their business models, operations, and product offerings. This transformation involves the deep application of technologies such as big data, cloud computing, artificial intelligence, and the Internet of Things (e.g., Gill *et al.*, 2019; Liu *et al.*, 2021). These rapidly evolving technologies not only drive companies' digital transformation efforts but also have a significant impact on their R&D investment and innovation intensity.

According to DIT, the early stages of innovation diffusion rely heavily on "innovators" and "early adopters." In the context of digital transformation, pioneering companies play the role of innovators. They are the first to adopt digital technologies, using AI, big data analytics, and other tools to optimize R&D processes and improve innovation efficiency (Niebel *et al.*, 2018). These pioneers are not just adopters of technology but also experimenters, committing resources to innovation when the technology is still nascent. They pave the way for subsequent firms by providing successful examples and lessons learned.

Early adopters, quickly follow the innovators once early success has been demonstrated. These firms typically possess strong market insights and resource integration capabilities, allowing them to learn from innovators' experiences and combine them with their own advantages to drive further technological diffusion and application (Antony *et al.*, 2023). As early adopters often have a significant influence within their industries, their success in digital transformation serves to enhance other firms' confidence in digital technologies, fostering an environment of R&D and innovation throughout the sector (Fratini *et al.*, 2013).

DIT highlights that the success of innovation diffusion depends on factors such as relative advantage, compatibility, complexity, trialability, and observability. The application of digital technologies in R&D processes offers notable relative advantages. For example, big data analytics and AI algorithms enable companies to predict market trends and consumer preferences more accurately, allowing them to develop more targeted R&D strategies (Vassakis *et al.*, 2017). Cloud computing reduces R&D costs and improves data

processing efficiency (Attaran and Woods, 2018). These relative advantages motivate companies to pursue digital transformation, leading to increased R&D investment.

Additionally, the use of digital technologies significantly improves R&D efficiency. For instance, virtual reality (VR) and augmented reality (AR) allow companies to conduct virtual product testing in the early stages of development, reducing the need for physical prototypes and accelerating innovation cycles (Steffen *et al.*, 2019). These relative advantages make companies more willing to increase R&D intensity during their digital transformation journey, as they strive to maintain their competitive edge.

2.4 Hypothesis Development

Building on RBT, AT, and DIT, I posit that digital transformation enhances firms' R&D intensity by providing them with valuable technological resources, improving governance mechanisms, and facilitating the diffusion of innovations. As discussed, digital transformation encompasses a range of technologies—such as Artificial Intelligence (AI), Blockchain (BC), Cloud Computing (CC), Big Data (BD), and Digital Technology Application (DTA)—each potentially offering distinct advantages to firms aiming to strengthen innovation capabilities.

First, AI can serve as a valuable, rare, inimitable, and non-substitutable resource (Agnihotri and Gabler, 2024), aiding in predictive analytics, streamlining product design, and optimizing resource allocation (Wang *et al.*, 2021). From an AT viewpoint, AI-driven real-time monitoring tools reduce information asymmetry, channeling more resources into innovation (Chaudhuri *et al.*, 2022). Moreover, DIT suggests that AI's clear relative advantages—such as automating labor-intensive processes—accelerate its adoption, further boosting R&D efficiency (Besiroglu *et al.*, 2024).

Second, BC's ability to securely record transactions and data creates transparent ledgers that can reduce information asymmetry and agency costs (Onjewu *et al.*, 2023). As a unique digital resource, blockchain can enhance collaboration in R&D by protecting intellectual property and streamlining supply chain processes (Chang *et al.*, 2019). In line with DIT, early adopters of blockchain demonstrate how the technology can facilitate secure data sharing, encouraging broader adoption and spurring R&D investment (e.g., Kwok and Treiblmaier, 2023; Xu and Guan, 2022).

Third, CC offers on-demand access to scalable computing resources, reducing the need for costly infrastructure investments (Schniederjans and Hales, 2016). Under RBT, these resources can be transformed into VRIN assets if they enable rapid experimentation, faster product development cycles, and superior collaboration among R&D teams (Hong *et al.*, 2024). From an AT standpoint, cloud-based platforms provide centralized data oversight, decreasing managerial opportunism and aligning resources toward innovative projects (Schneckenberg *et al.*, 2021). DIT further suggests that CC's relative advantage and trialability foster its widespread adoption, thus amplifying its impact on R&D activities (Kruger and Steyn, 2023).

Fourth, BD analytics allows firms to process extensive datasets for real-time insights, helping identify market trends and potential innovation paths (Bresciani *et al.*, 2021). This capacity can constitute a VRIN resource, enhancing competitive advantage by guiding targeted R&D (Ranjan and Foropon, 2020). In terms of AT, transparent analytics dashboards reduce information asymmetry and encourage efficient resource allocation (Manroop *et al.*, 2024). From a DIT perspective, the observable success of early BD adopters motivates other firms to invest in analytics tools, further intensifying R&D efforts (Park and Kim, 2019).

Fifth, beyond individual solutions, a holistic application of digital technologies (e.g., VR, AR, IoT) can reshape business models and operations, creating synergies that drive

ongoing R&D investment (Ancillai *et al.*, 2023). By reducing operational frictions and enabling cross-functional collaboration, these broader digital tools also mitigate agency costs and free up resources for experimentation (Yuan and Pan, 2023). Meanwhile, DIT underscores that the successful integration of multiple technologies can serve as a powerful demonstration effect, spurring broader adoption and intensifying industry-wide R&D (De Propriis and Bailey, 2021).

Based on the above discussion, I propose the following hypothesis:

H1: All else being equal, digital transformation drives an increase in firms' R&D intensity.

H1a: AI, as a sub-dimension of digital transformation, positively influences firms' R&D intensity.

H1b: BC, as a sub-dimension of digital transformation, positively influences firms' R&D intensity.

H1c: CC, as a sub-dimension of digital transformation, positively influences firms' R&D intensity.

H1d: BD, as a sub-dimension of digital transformation, positively influences firms' R&D intensity.

H1e: DTA, as a sub-dimension of digital transformation, positively influences firms' R&D intensity.

State-owned enterprises (SOEs) often enjoy preferential access to capital, policy support, and governmental resources, which can amplify the benefits they derive from digital transformation (Wu *et al.*, 2024). From an RBT perspective, these advantages enable SOEs to acquire, integrate, and leverage digital technologies with greater efficiency, thereby enhancing their R&D capabilities more substantially than non-SOEs. Additionally, their privileged standing may provide them with broader opportunities for developing proprietary digital solutions—such as AI or big data platforms—further elevating their innovation potential (Muldoon *et al.*, 2024). Thus, I propose the following hypothesis.

H2: Compared with non-SOEs, the positive impact of digital transformation on firms' R&D intensity is stronger in SOEs.

Profitable firms often have the financial slack and stronger internal resources needed to integrate advanced digital technologies into their innovation processes (Babalola *et al.*, 2025). Drawing on RBT, these resources enable profitable firms to acquire and leverage proprietary digital solutions—such as AI-driven analytics or specialized big data systems—leading to greater improvements in R&D productivity (Ardito *et al.*, 2024). In contrast, loss-making enterprises typically face tighter budget constraints, limiting their ability to invest in and capitalize on the potential of digital transformation. Consequently, I posit the following hypothesis:

H3: Compared with loss-making enterprises, the positive impact of digital transformation on firms' R&D intensity is stronger for profitable firms.

High-tech enterprises typically operate in rapidly evolving, knowledge-intensive markets, where digital transformation initiatives—such as data-driven prototyping, real-time analytics, and advanced automation—can be more seamlessly integrated into R&D processes (Wang and Li, 2024). This integration can accelerate product development cycles and enable more radical innovation, giving high-tech firms a distinct competitive edge. In contrast, non-high-tech enterprises often prioritize incremental improvements over disruptive innovation, limiting the immediate impact of digital tools on their R&D activities. Consequently, I posit the following hypothesis:

H4: Compared with non-high-tech firms, the positive impact of digital transformation on firms' R&D intensity is stronger for high-tech firms.

In China, regional disparities in economic development, infrastructure, and resource availability can significantly influence how effectively firms leverage digital technologies for innovation. In the central and eastern regions, companies tend to have better access to advanced digital infrastructure, specialized talent, and supportive local policies that facilitate the adoption of AI, cloud computing, and other emerging technologies (Liu *et al.*, 2023). These factors can enable central and eastern firms to extract greater value from digital transformation efforts, resulting in faster product development cycles, enhanced collaboration, and a stronger emphasis on cutting-edge R&D projects. In contrast, western region enterprises may face limitations such as weaker digital infrastructure and a relative scarcity of high-skilled labor, constraining their ability to reap the full benefits of digital transformation.

Furthermore, the presence of technology clusters and innovation ecosystems in the central and eastern regions often fosters synergies among industry, academia, and government, thereby amplifying the positive impact of digital transformation on R&D intensity. Firms located in these areas can more readily collaborate with top-tier universities, research institutes, and venture capital networks, accelerating the diffusion of advanced digital tools throughout their R&D processes (Chen and Wang, 2024). By comparison, enterprises in the western region may lack similar network advantages and be slower to adopt or optimize these technologies. Consequently, I posit the following hypothesis:

H5: Compared with firms in the western region, the positive impact of digital transformation on R&D intensity is stronger for firms in the central and eastern regions of China.

3. Methodology

3.1 Data Source

Based on data from Chinese A-share listed companies spanning from 2012 to 2021, this study utilizes R&D data from the Wind Database, digital transformation information from company annual reports, and additional data from the CSMAR Database. The data preprocessing involved several steps to ensure accuracy and relevance: (1) Excluding financial institutions due to their distinct governance and performance systems compared to non-financial firms in China; (2) Eliminating companies designated as "Special Treatment" (ST) under Chinese securities regulations, which face potential delisting due to consecutive years of losses, to avoid the impact of anomalous financial conditions; (3) Removing companies with missing information to enhance the accuracy and usability of the dataset. After excluding missing data and outliers, the final sample consists of 27,163 firm-year observations from 3,920 unique companies.

3.2 Variable Selection and Measurement

(I) **Dependent Variable:** The dependent variable in this study is R&D intensity (RDI). Following the approach of Padgett and Galan (2009), RDI is proxied by dividing R&D expenditures by total sales. For robustness checks, I also calculate RDI as the ratio of R&D expenditures to total assets to ensure consistency in the measurement of R&D intensity.

(II) **Independent Variable:** The independent variable of interest is digital transformation (DT). Corporate annual reports provide valuable insights into a firm's strategy and future direction, making them an effective source for evaluating strategic priorities (Kindermann *et al.*, 2020). To capture digital transformation, I adopt a Python-based web scraping approach,

following the methodology of Tu and He (2022), to systematically analyze the textual content of company annual reports.

The construction of the digital transformation metric follows these steps: First, I selected keywords related to digital transformation, such as “artificial intelligence,” “cloud computing,” “blockchain,” “big data,” and “digital technology applications.” Next, using Python libraries such as Jieba, Gensim, and Re, I processed the textual data by tokenizing it, removing stop words, and employing the Skip-Gram and Continuous Bag-of-Words (CBOW) models for word embedding (Rong, 2014). Negative sampling was applied during the training phase to refine the neural network model, which was built to detect terms semantically related to the selected seed words.

Subsequently, I constructed a research corpus and used the Word2Vec model to generate word vectors and measure word similarities, thereby identifying additional terms linked to the core digital transformation concepts. These terms were categorized into five dimensions of digital transformation (outlined in Table 1). To quantify the level of digital transformation for each firm, I calculated the frequency of digital transformation-related keywords. Since the keyword frequencies exhibit a typical right-skewed distribution, I applied a natural logarithmic transformation to the data (i.e., $\text{LN}(\text{keyword frequency} + 1)$) to standardize the metric. This transformed value serves as the digital transformation score for all Chinese A-share listed firms in my sample.

(III) Control Variables. In order to isolate the impact of DT on RDI, I account for several control variables that could influence RDI, drawing on both prior empirical findings and theoretical perspectives such as RBT and corporate governance literature (e.g., Pu and Zulkafli, 2024; Yuan and Wen, 2018). Based on RBT, firm-specific resources and capabilities—often linked to size, age, and financial health—can shape a firm’s capacity and strategic choice to invest in innovation (Wernerfelt, 1984). Meanwhile, corporate governance theories (e.g., Agency Theory) suggest that board structure and ownership concentration may affect firms’ decision-making processes for R&D expenditure (e.g., Appio *et al.*, 2021; Urbinati *et al.*, 2020; Wang and Li, 2024).

Specifically, the first control variable is Firm Size (Size), measured as the natural logarithm of total assets, which captures the effect of scale on R&D activities. Next, Firm Age (FirmAge), expressed as the logarithm of the number of years since the firm’s founding, reflects the impact of firm maturity on its capacity to invest in R&D. I also control for Financial Leverage (Lev)—the ratio of total debt to total assets—as higher leverage may constrain a firm’s ability to engage in R&D due to debt servicing obligations. Return on Assets (ROA), defined as net income divided by total assets, indicates firm profitability, which can shape the resources available for R&D investment. Meanwhile, Sales Growth (Growth)—the percentage change in operating income relative to the previous year—serves as a proxy for market performance and growth opportunities, potentially influencing R&D spending decisions.

In terms of governance-related factors, I include Board Size (Board), measured as the natural logarithm of the total number of directors, because larger boards may enhance decision-making capacity and strategic oversight. Finally, Ownership Concentration (TOP1)—the proportion of shares held by the largest shareholder—is considered, since firms with concentrated ownership often experience distinct governance dynamics that can affect their R&D strategies.

Table 1: Keywords of digital transformation

Dimension	Keywords
Artificial Intelligence (AI)	artificial intelligence, business intelligence, intelligent robotics, machine learning, deep learning, image understanding, investment decision aids, intelligent data analysis, semantic search, biometrics, face recognition face recognition, voice recognition, identity verification, autonomous driving, natural language processing
Blockchain (BC)	blockchain, digital currency, differential privacy technologies, smart financial contracts, distributed computing
Cloud Computing (CC)	cloud computing, streaming computing, converged architecture, EB-class storage, billion concurrency, graph computing, in-memory computing, multiparty secure computing, brain-like computing, green computing, cognitive computing, internet of things, information physical systems
Big Data (BD)	big data, data mining, text mining, data visualization, heterogeneous data, credit reporting, augmented reality, mixed reality, virtual reality
Digital Technology Application (DTA)	mobile internet, B2B, B2C, C2B, C2C, O2O, e-commerce, industrial internet, mobile internet, internet healthcare, mobile payment, third-party payment, smart customer service, smart marketing, quantitative finance, NFC payment, smart energy, netlink, smart wearing, smart agriculture, smart transportation, smart healthcare, smart home, smart investment, smart tourism, smart environmental protection, smart grid, digital marketing, unmanned retail, internet finance, digital finance, fintech, financial technology, open banking

3.3 Empirical Model

To empirically investigate the effect of DT on RDI, I first estimate the following base regression model:

$$RDI_{i,t} = \alpha_0 + \alpha_1 DT_{i,t} + \alpha_2 Controls_{i,t} + \alpha_3 Year_{i,t} + \alpha_4 Firm_{i,t} + \varepsilon \quad (1)$$

where $RDI_{i,t}$ denotes R&D intensity of firm; $DT_{i,t}$ measures digital transformation for firm; $Controls_{i,t}$ is a vector of firm-level control variables, such as firm size (Size), profitability (ROA), leverage (LEV), growth opportunities (Growth), and other factors that can influence RDI; $Year_{i,t}$ denotes year fixed effects, controlling for macroeconomic shocks that vary by year and could affect all firms; $Firm_{i,t}$ indicates firm fixed effects, absorbing firm-specific characteristics that might otherwise bias the estimated impact of DT on RDI.

Furthermore, I winsorize all continuous variables at the 1st and 99th percentiles to mitigate the influence of outliers, thus preventing a small number of extreme observations from unduly skewing our estimates. This approach ensures that unusually large or small values do not dominate the results, while still retaining most of the underlying variation in the data. In addition, i cluster standard errors at the firm level to account for potential intra-firm correlations over time, a practice that helps correct for serial dependence within firms and provides more reliable statistical inferences in a panel data context.

4. Empirical Results

4.1 Summary Statistics

Table 2 presents the descriptive statistics for the key variables used in the analysis, including the number of observations, mean, standard deviation, and the minimum and maximum values for each variable.

For the R&D-related variables (RDI and RDI2), the mean value of RDI is 0.041, with a standard deviation of 0.046, and it ranges from 0.000 to 0.256. This wide range reflects the substantial variation in R&D intensity across firms in the sample. The alternative dependent variables used in the robustness tests (RDI2) exhibit means and standard deviations that also indicate considerable variability (0.000 to 0.101), underscoring the diverse levels of R&D intensity among firms.

Regarding digital transformation, the variable DT shows a mean of 1.415 and a standard deviation of 1.388, with values spanning from 0 to 5.056. These statistics suggest that the

digital transformation efforts of Chinese listed firms are generally at an early stage, with most firms displaying relatively low levels of engagement. Overall, the descriptive statistics of the remaining variables align with the patterns observed in prior studies and reflect the realities of the Chinese market (e.g., Yuan and Wen, 2018). This consistency reinforces the reliability and representativeness of the dataset.

Table 2: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
RDI	27,163	0.041	0.046	0.000	0.256
RDI2	27,163	0.020	0.020	0.000	0.101
DT	27,163	1.415	1.388	0.000	5.056
Size	27,163	22.267	1.293	19.814	26.153
FirmAge	27,163	2.921	0.319	1.609	3.497
Lev	27,163	0.422	0.203	0.050	0.893
ROA	27,163	0.041	0.063	-0.239	0.222
Growth	27,163	0.170	0.390	-0.544	2.445
Board	27,163	2.122	0.198	1.609	2.708
TOP1	27,163	34.206	14.810	8.630	74.180

4.2 Correlation and Variance Inflation Factor Analysis

Table 3 presents the Pearson correlation matrix for the key variables in this study. Overall, the correlation coefficients between the independent and control variables are mostly below 0.50, suggesting that multicollinearity is not a significant concern. To further verify this, I performed a multicollinearity diagnostic test on the continuous variables. The results show that the variation inflation factors (VIF) for all control variables are below 2, indicating that there is no serious multicollinearity issue in the model. This ensures that the regression estimates are not distorted by high inter-variable correlations, allowing for reliable interpretation of the results.

4.3 Univariate Analysis

In conducting the univariate analysis, this study introduced a binary variable labeled as Dummy (DT). This variable takes a value of 1 if a firm's annual report contains keywords or their related synonyms that signal digital transformation, identified through machine learning techniques, and a value of 0 if not.

Table 3: Pearson Correlation

	RDI	DT	Size	FirmAge	Lev	ROA	Growth	Board	TOP1	VIF
RDI	1.000									-
DT	0.338***	1.000								1.04
Size	-0.262***	0.026***	1.000							1.61
FirmAge	-0.152***	0.019***	0.170***	1.000						1.06
Lev	-0.338***	-0.071***	0.518***	0.164***	1.000					1.71
ROA	0.028***	0.020***	-0.002	-0.080***	-0.358***	1.000				1.34
Growth	-0.003	0.043***	0.037***	-0.043***	0.021***	0.260***	1.000			1.10
Board	-0.134***	-0.079***	0.273***	0.054***	0.153***	-0.003	-0.024***	1.000		1.09
TOP1	-0.185***	-0.116***	0.197***	-0.089***	0.060***	0.126***	-0.010*	0.026***	1.000	1.10

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 4: Univariate Analysis

	Dummy (DT)=1		Dummy (DT)=0		Differences T-value
	N	Mean	N	Mean	
RDI	18006	2.13527	9157	0.00000	2.13527***

Notes: This table presents the results of univariate analysis on the mean difference of RDI between firms having digital transformation and firms having no digital transformation. The t-values for mean differences are based on t-tests. ***denotes significance at the 1% level.

Table 4 reports the results of the t-test comparing RDI between firms that have adopted DT and those that have not. On average, companies that underwent DT exhibit a significantly higher mean RDI than their non-DT counterparts, and this difference is statistically significant at the 1% level. These findings indicate that firms engaging in DT display markedly greater RDI than firms not employing this technology. Moreover, as shown in Figure 1, the positive slope of the linear regression line further reinforces the positive association between DT and RDI. Overall, these results support the conclusion that firms implementing digital transformation tend to allocate more resources to R&D compared to those that have not undergone such a transformation.

Table 5: Baseline regression results

	(1)	(2)
	RDI	RDI
DT	0.00078** (2.46)	0.00078*** (2.63)
Size		0.00279*** (3.00)
FirmAge		-0.01721*** (-3.31)
Lev		-0.02638*** (-9.24)
ROA		-0.06552*** (-11.73)
Growth		-0.00511*** (-11.50)
Board		0.00189 (0.88)
TOP1		-0.00010* (-1.84)
_cons	0.03521*** (63.97)	0.03428 (1.52)
Year FE	Y	Y
Firm FE	Y	Y
N	27163	27163
adj. R2	0.029	0.088

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; robust standard errors are clustered at the firm level, and the related t-statistics are reported in brackets.

4.4 Multivariate Results

Table 5 presents the baseline regression results testing the relationship between DT and RDI to assess H1. All regression models include year-firm two way fixed effects to account for firm-specific heterogeneity and the effects of time-related shocks or policy changes. The estimates utilize clustered standard errors at the firm level to ensure robust statistical inference.

Column (1) displays the regression results for the effect of DT on RDI without any control variables. The coefficient of DT is positive and statistically significant at the 5% level ($\alpha_1 = 0.00078$, $t = 2.46$), suggesting a meaningful positive impact of digital transformation on R&D intensity when no additional factors are considered. This indicates that firms engaged in digital transformation tend to have higher R&D intensity. Column (2) incorporates a set of control variables, refining the analysis by accounting for firm-specific factors. Even after controlling for these variables, the positive relationship between DT and RDI remains robust and statistically significant at the 1% level ($\alpha_1 = 0.00078$, $t = 2.63$). The economic significance of this effect is also noteworthy. A one standard deviation increase in DT is associated with a 2.35% increase in the standard deviation of RDI, highlighting the substantial impact of DT on firms' R&D intensity.

The control variables in Column (2) provide several key insights. Size has a positive and significant impact on R&D intensity ($\alpha_2 = 0.00279$, $t = 3.00$), indicating that larger firms invest more in R&D. In contrast, FirmAge shows a negative association ($\alpha_2 = -0.01721$, $t = -3.31$), suggesting older firms may invest less. Lev and ROA both negatively affect RDI (Lev: $\alpha_2 = -0.02638$, $t = -9.24$; ROA: $\alpha_2 = -0.06552$, $t = -11.73$), implying firms with higher debts or profitability may allocate fewer resources to R&D. Growth also negatively impacts RDI ($\alpha_2 = -0.00511$, $t = -11.50$). Board and TOP1 show marginal or insignificant effects.

The findings from this regression analysis highlight the critical role of digital transformation in enhancing R&D intensity across firms. The robust positive relationship between DT and RDI, even after controlling for firm size, age, leverage, and profitability, underscores the transformative potential of digital technologies in fostering corporate innovation. Moreover, the control variables reveal nuanced insights into firm behavior, where larger firms tend to be more R&D-focused, while older, highly leveraged, and more profitable firms may show less emphasis on innovation. These findings align with prior research, which suggests that digital technology drives firms to invest more intensively in R&D, enhancing their long-term competitiveness and growth in rapidly evolving technological landscapes (e.g., Appio *et al.*, 2021; Gao *et al.*, 2022).

Table 6: Robustness check

	Alter DV	Subsample	Cluster (Firm & Year)
	(1)	(2)	(3)
	RDI2	RDI	RDI
DT	0.00063*** (5.08)	0.00070** (2.14)	0.00078*** (2.80)
Controls	Y	Y	Y
Year FE	Y	Y	Y
Firm FE	Y	Y	Y
N	27163	19919	27163
adj. R2	0.061	0.071	0.088

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; robust standard errors are clustered at the firm or firm & year level, and the related t-statistics are reported in brackets.

4.5 Robustness Check

The preceding analysis indicates a robust positive correlation between DT and RDI. To further validate the robustness of these findings, I conducted several additional tests:

(I) Alternative Dependent Variables. I reassessed the primary model by employing an alternative measure of R&D intensity, denoted as RDI2. The regression outcomes, detailed in Column (1) of Table 6, reveal that the coefficients for the key variables retain similar magnitudes and directions as observed in the original analysis. This consistency reinforces the reliability of our baseline results.

(II) Subsample Regression Test. The COVID-19 pandemic significantly accelerated the adoption of remote work and necessitated rapid digital tool integration to support R&D efforts (Bachmann and Frutos-Bencze, 2022). To account for the potential impact of this unique period, I excluded data from 2020 and 2021. The results, shown in Column (2) of Table 6, continue to support the previous findings, affirming that the positive relationship between DT and RDI remains robust even when accounting for pandemic-specific influences.

(III) Adjusting Standard Errors to Firm-Year Level. To account for potential correlations within firm-year observations that might be overlooked by standard firm-level adjustments, I performed a clustered standard error analysis at both the firm and year levels. This dual-clustering approach addresses heteroscedasticity and autocorrelation issues more comprehensively, providing a more accurate assessment of the standard errors. The results,

detailed in Column (3) of Table 6, reveal that the magnitude and significance of the coefficients remain consistent, confirming the robustness of our conclusions.

Table 7: Endogenous treatment

	PSM	DID	GMM
	(1)	(2)	(3)
	RDI	RDI	RDI
DT	0.00080*** (2.70)		0.00150*** (3.61)
DT-did		0.00136** (2.08)	
L.RDI			0.80758*** (23.33)
Controls	Y	Y	Y
Year FE	Y	Y	Y
Firm FE	Y	Y	Y
AR(1)			0.000
AR(2)			0.701
Hansen J			0.761
N	27069	27163	22910
adj. R2	0.088	0.087	

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; in columns (1) and (2), robust standard errors are clustered at the firm level, with the corresponding t-statistics reported in parentheses, whereas in column (3), the parentheses contain z-statistics calculated using the robust standard error specification.

4.6 Endogenous Treatment

Addressing potential endogeneity in the relationship between DT and RDI is crucial to avoid biased estimates. Endogeneity can arise from multiple sources—such as sample selection, unobserved heterogeneity, and reverse causality—each of which may distort ordinary least squares (OLS) results. To mitigate these issues, I adopt three complementary econometric methods: Propensity Score Matching, Difference-in-Differences, and Generalized Method of Moments. Each technique targets different dimensions of endogeneity, thereby enhancing the robustness of our conclusions.

(I) Propensity Score Matching (PSM). Firms that choose to implement DT (the treatment group) may systematically differ from those that do not (the control group), creating selection bias if these differences also affect RDI. To address this concern, I employ a 1:1 nearest-neighbor PSM approach, using firm-level variables as covariates to generate propensity scores for treatment assignment. By matching each DT-adopting firm to a non-adopting firm with a similar probability of adopting DT, I ensure that the two groups are comparable on observable characteristics, substantially reducing selection bias. The re-estimated results, shown in the first column of Table 7, indicate that after matching, the impact of DT on RDI remains consistently positive and significant, reaffirming the robustness of our baseline findings in a setting where observable firm-level differences are balanced.

(II) Difference-in-Differences (DID). Even when observable attributes are balanced, unobserved time-invariant characteristics and time-varying external factors may still bias the estimated effect of DT. Given that digital transformation is often implemented incrementally, I exploit this quasi-experimental setting to apply a DID framework. This method compares the change in RDI over time between the treatment group (firms implementing DT) and the control group (firms not implementing DT), thereby controlling for both firm-specific unobserved heterogeneity and any common time trends. Following Wang *et al.* (2023), I specify our DID model as:

$$RDI_{i,t} = \chi_0 + \chi_1 DT - did_t + \chi_2 Controls_{i,t} + \chi_3 Year_{i,t} + \chi_4 Firm_{i,t} + \varepsilon \quad (2)$$

The regression results, reported in Column (2) of Table 7, indicate that the DID coefficient for DT is 0.00136 ($p < 0.05$). This finding suggests that, even after accounting for potential time-invariant confounders and time-varying shocks, DT continues to have a positive and statistically significant effect on RDI, reinforcing the credibility of our baseline estimations.

(III) Generalized Method of Moments (GMM). Although DID and PSM alleviate many biases, reverse causality—where firms with higher RDI may be more inclined (or better positioned) to adopt DT—could still lead to biased estimates. To address this possibility, I employ a two-step system GMM estimator (Bond, 2002), which uses internally generated instruments (lagged values of the dependent and independent variables) to correct for simultaneity and omitted variable bias in dynamic panel data settings. By leveraging appropriate lags as instruments, GMM isolates the exogenous variation in DT and provides unbiased estimates of its effect on RDI. The results in Column (3) of Table 7 show that DT remains a statistically significant predictor of RDI, even after accounting for dynamic feedback effects, and the diagnostic tests (Arellano-Bond autocorrelation and Hansen over-identification) confirm the validity of the instruments. These findings further strengthen our conclusion that digital transformation exerts a robust, positive impact on firms' R&D intensity.

By deploying these three methods in tandem, I address different dimensions of endogeneity—selection bias (PSM), time-invariant and time-varying confounders (DID), and reverse causality (GMM). Collectively, they provide a rigorous framework for inferring the causal effect of DT on RDI, offering a multifaceted assurance that our results are not driven by a single methodological assumption.

Table 8: Dimensions of digital transformation and R&D intensity

	(1) RDI	(2) RDI	(3) RDI	(4) RDI	(5) RDI
AI	0.00198*** (3.61)				
BC		0.00072 (0.24)			
CC			0.00129** (2.53)		
BD				0.00078* (1.66)	
DTA					0.00012 (0.35)
Controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
N	27163	27163	27163	27163	27163
adj. R ²	0.089	0.087	0.088	0.088	0.087

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; robust standard errors are clustered at the firm level, and the related t-statistics are reported in brackets.

4.7 Further Analysis

4.7.1 Driving Dimensions of Digital Transformation

Prior findings of this paper indicate that DT has the potential to enhance firms' R&D intensity. A pertinent question arises: which dimensions of DT serve as the key drivers of this effect?

The results in Columns (1) through (5) of Table 8 elucidate the varying impacts of distinct dimensions of DT on RDI. Column (1) highlights a significant positive correlation

between artificial intelligence (AI) and RDI ($\alpha_1 = 0.00198$, $t = 3.61$), suggesting that firms leveraging AI technology are more likely to increase their R&D investment. In contrast, Column (2) presents a positive but statistically insignificant relationship between blockchain (BC) and RDI ($\alpha_1 = 0.00072$, $t = 0.24$), implying that its effect on R&D may not be as direct or immediate. Column (3) reports a significant positive effect of cloud computing (CC) on RDI ($\alpha_1 = 0.00129$, $t = 2.53$), indicating that the adoption of cloud infrastructure facilitates greater intensity in R&D activities. Column (4) shows a marginally significant effect of big data (BD) on RDI ($\alpha_1 = 0.00078$, $t = 1.66$), while the relationship between digital technology applications (DTA) and RDI in Column (5) is statistically insignificant ($\alpha_1 = 0.00012$, $t = 0.35$), suggesting that DTA may not play a substantial role in enhancing R&D intensity.

Overall, these results underscore the differential impact of various DT dimensions on RDI. The strong influence of AI in Column (1) reflects its transformative potential in optimizing business operations and driving strategic innovation, aligning with previous research that underscores the role of AI in enhancing decision-making and innovation capacity (Guo *et al.*, 2022; Şimşek *et al.*, 2019). Meanwhile, the weaker effect of blockchain in Column (2) may be attributed to the fact that while many firms accumulate large datasets using blockchain technology, they may not yet fully utilize its analytical capabilities for R&D purposes, delaying its tangible impact on innovation. The positive contribution of cloud computing (CC) in Column (3) suggests that the flexibility and scalability provided by cloud infrastructure enable firms to streamline resource access, thus improving the efficiency of their R&D activities. The marginal effect of big data in Column (4) might be due to the varying degrees of digital adoption across firms, where initial phases of digitalization provide limited but promising boosts to R&D investment. Lastly, the insignificant coefficient of digital technology applications in Column (5) implies that while multiple digital technologies hold individual value, they may influence R&D independently rather than synergistically, contrasting with studies such as Wang *et al.* (2022), which emphasize the potential of technological integration in amplifying R&D outcomes.

Table 9: Further analysis (1)

	SOE=1	SOE=0	LOSS=1	LOSS=0
	(1)	(2)	(3)	(4)
	RDI	RDI	RDI	RDI
DT	0.00121*** (2.97)	0.00041 (1.05)	0.00179 (1.48)	0.00063** (2.19)
Controls	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
N	9611	17552	2856	24307
adj. R2	0.088	0.093	0.065	0.075

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; robust standard errors are clustered at the firm level, and the related t-statistics are reported in brackets.

4.7.2 Heterogeneity Test

To further elucidate the differential impacts of DT on RDI, this study conducts cross-sectional tests based on ownership, profitability, industry technology characteristics, and geographic location.

4.7.2(a) State-Owned Enterprises VS Non-State-Owned Enterprises

The impact of DT on RDI may vary due to ownership differences among firms. Many state-owned enterprises (SOEs) in emerging markets have intricate connections with bureaucratic systems and maintain close relationships with politicians, business elites, and government officials (Mansha *et al.*, 2022; Tee *et al.*, 2022; Wong and Hooy, 2018, 2024). These

political ties enable SOEs to access more resources for digital infrastructure development, as governments often enact policies that support digital transformation within these entities (Liu *et al.*, 2024). This combination of political backing and resource availability positions SOEs to leverage DT more effectively in fostering R&D activities.

Conversely, non-state-owned enterprises (No-SOEs) often face financial constraints, which may limit their investments in digital transformation efforts. No-SOEs are more reliant on their own profitability to finance the DT process (Li and Xia, 2008), potentially reducing the effectiveness of these efforts in enhancing R&D activities. Moreover, No-SOEs may encounter competitive market pressures that force them to prioritize short-term profits and operational survival over long-term investments in R&D (Sun *et al.*, 2022). Thus, compared to SOEs, No-SOEs face greater challenges in pursuing DT initiatives.

To test this hypothesis, the study conducted subsample analyses for SOEs and No-SOEs, as shown in Columns (1) and (2) of Table 9. The results indicate that DT has a statistically significant effect at the 1% level for SOEs, while the DT coefficient for No-SOEs is not significant. Therefore, the impact of DT on RDI is more pronounced in SOEs.

Table 10: Further analysis (2)

	HIGH-TECH=1	HIGH-TECH=0	EAST	MIDDLE	WEST
	(1)	(2)	(3)	(4)	(5)
	RDI	RDI	RDI	RDI	RDI
DT	0.00085** (2.18)	0.00047 (1.34)	0.00064* (1.77)	0.00248*** (2.88)	-0.00020 (-0.35)
Controls	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y
N	15987	11176	19260	3569	4330
adj. R2	0.124	0.048	0.095	0.074	0.087

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels respectively; robust standard errors are clustered at the firm level, and the related t-statistics are reported in brackets.

4.7.2(b) Loss-Making Enterprises VS Profit-Making Enterprises

Profit-making firms generally have more robust financial resources to invest in R&D as part of their digital transformation progress. DT often requires significant upfront investments in new technologies, such as artificial intelligence, cloud computing, and big data (Saarikko *et al.*, 2020). Given their stronger financial health, profit-making firms are better positioned to absorb these high costs and reap the benefits of R&D. In contrast, loss-making firms, constrained by limited capital, are more likely to adopt a conservative approach to cash management, reducing their investments in high-risk, high-cost R&D projects (Tang *et al.*, 2021). Therefore, this study anticipates that the impact of DT on RDI will be less pronounced for loss-making enterprises compared to profit-making ones.

To verify this assumption, the study conducted cross-sectional analyses for loss-making and profit-making enterprises, as presented in Columns (3) and (4) of Table 9. The results demonstrate that the DT coefficient is statistically significant at the 5% level for profit-making firms, while it is not significant for loss-making firms. Hence, DT has a stronger impact on RDI in profit-making enterprises.

4.7.2(c) High-Tech Enterprises VS Non-High-Tech Enterprises

High-tech firms often view R&D as a core strategic focus, with R&D expenditures constituting a significant portion of their total investments (Blank, 2024). Through the adoption of automation, intelligent systems, and data-driven decision-making tools, DT may enable these firms to more effectively identify technological opportunities and market demands, thereby increasing RDI. In addition to optimizing the R&D process, DT enhances the innovative capacities of high-tech firms, potentially yielding greater returns on R&D

investments. In contrast, for non-high-tech firms, R&D is not typically a primary strategic goal. These firms are more likely to emphasize cost control, production efficiency, and service quality over technological innovation (Nunes *et al.*, 2012). As a result, while DT may improve R&D efficiency, its impact on increasing RDI is less significant in non-high-tech sectors.

To validate this analysis, the study conducted cross-sectional analyses for high-tech and non-high-tech enterprises, as shown in Columns (1) and (2) of Table 10. The findings reveal that the DT coefficient is statistically significant at the 5% level for high-tech firms, whereas it is not significant for non-high-tech firms. Therefore, the influence of DT on RDI is more evident in high-tech enterprises.

4.7.2(d) Eastern Region VS Central Region VS Western Region

Previous studies suggest that firms in China's eastern and central regions, where the digital economy is more advanced, benefit from greater access to digital resources, superior infrastructure, and higher levels of capital investment (Zhang *et al.*, 2021). These regions boast more developed digital infrastructure, higher internet penetration rates, and broader adoption of digital technologies and IT services. In contrast, the western region of China lags behind economically, with less advanced infrastructure and technical capabilities. Although firms in the Western region have started implementing DT, limitations in technology (Li *et al.*, 2022), funding, and human capital prevent them from fully capitalizing on DT to enhance R&D efforts. Consequently, the impact of DT on RDI may be less significant in the western region.

To test this hypothesis, the study conducted cross-sectional analyses based on geographic location, dividing firms into eastern, central, and western regions, as displayed in Columns (3) to (5) of Table 10. The results show that the DT coefficients for firms in the eastern and central regions are positive and statistically significant, while the coefficient for firms in the western region is not significant. Therefore, the effect of DT on RDI is more pronounced in firms located in the eastern and central regions.

5. Conclusion

Drawing upon a dataset of Chinese publicly listed firms from 2012 to 2021, this study employs machine learning techniques to construct a comprehensive measure of digital transformation and investigate its impact on R&D intensity, as well as the underlying dimensions that drive this effect. The key findings are as follows: First, there is a significant positive relationship between digital transformation and R&D intensity. Firms that engage in DT see a notable increase in their R&D efforts, enhancing their innovation willingness. The robustness of these results was confirmed through alternative dependent variables, subsample testing, and adjustments to clustering levels. The results also remained consistent when addressing potential endogeneity using propensity score matching, difference-in-differences, and the generalized method of moments. Second, by analyzing various dimensions of DT, the study finds that the application of artificial intelligence, cloud computing, and big data are the primary drivers of enhanced R&D intensity through digital transformation. Third, the impact of DT on R&D intensity exhibits heterogeneity. Specifically, this positive effect is more pronounced in state-owned enterprises, profit-making enterprises, high-tech enterprises, and firms located in the eastern and central regions of China.

This study makes several notable contributions. First, it constructs a comprehensive DT index using machine learning-based text mining techniques combined with information extracted from annual reports. This index provides a valuable reference for assessing the level of digital transformation and its impact on R&D. Second, while prior literature has

offered limited insights into the overall impact of DT on innovation, this study enriches the understanding of DT's multifaceted effects on innovation by focusing on R&D intensity and its dynamic changes. Third, the study fills a critical gap in the literature by analyzing the sub-dimensions of DT and their distinct impacts on R&D. Prior research has generally treated DT as a monolithic factor, without exploring the nuanced differences across its dimensions. By doing so, this study offers deeper insights into the mechanisms through which DT affects R&D. Finally, this research provides a more comprehensive understanding of the economic benefits of DT by examining its heterogeneous effects based on ownership, profitability, technological characteristics, and geographical location, thereby informing the formulation of more targeted and effective policies.

Based on the findings, several important policy implications can be drawn. First, promote targeted development of AI, Big Data, and Cloud Computing. Since the sub-dimension analysis indicates that AI, Big Data, and Cloud Computing each contribute significantly to R&D intensity, policymakers should prioritize support for these three domains in national and regional digital strategies. For instance, governments can establish specialized funding programs or innovation hubs to advance cutting-edge research and practical applications in these areas. By doing so, firms can better integrate key digital tools into their innovation processes, ultimately boosting R&D output.

Second, leverage state-owned enterprises' advantages in digital transformation. The finding that digital transformation exerts a stronger influence on R&D intensity among SOEs suggests that these firms may already possess structural advantages—such as preferential financing or policy backing—that help them capitalize on digital initiatives. To further amplify this effect, policymakers could encourage technology transfers or partnerships between SOEs and private-sector firms, ensuring that smaller or less-resourced companies also benefit from the SOEs' expertise. Such collaboration may accelerate the overall diffusion of advanced digital technologies across the market.

Third, strengthen support for profitable enterprises to scale digital transformation. Because profitable firms exhibit a more pronounced positive relationship between digital transformation and R&D intensity, policymakers must reinforce incentives that enable these companies to expand their digital initiatives further. For example, tax incentives or matching grants could be offered to firms that demonstrate a commitment to using AI, Big Data, and Cloud Computing in ways that directly enhance innovation. In turn, highly profitable firms—already equipped with more robust financial resources—would find it even more attractive to invest in deeper, long-term R&D projects that can yield substantial economic benefits.

Fourth, encourage high-tech enterprises to maximize digital synergies. The stronger DT–R&D link observed in high-tech enterprises highlights how digitization can feed into existing innovation-driven business models. Policymakers can scale up sector-specific programs that reward technological breakthroughs, particularly in AI and Big Data solutions. Additionally, closer collaboration between government agencies, research institutions, and high-tech firms can help create an ecosystem that fosters rapid experimentation and advanced product development, further magnifying the returns on digital transformation.

Fifth, address regional disparities by capitalizing on eastern and central advantages. Given that the positive effect of digital transformation on R&D intensity is more pronounced in eastern and central regions of China, local governments in these areas should continue investing in digital infrastructure, talent development, and innovation ecosystems to maintain their competitive edge. At the same time, national-level policies can aim to transfer know-how and best practices from these more developed regions to western provinces—reducing digital and innovation gaps across the country. This strategy balances

further growth in established industrial centers with inclusive development in less advanced areas, helping all regions benefit from the economic gains brought about by digital transformation.

While this study underscores the role of digital transformation in enhancing R&D intensity, it is important to recognize certain limitations in the sample used. Future research could deepen our understanding of digital transformation's impact on R&D by expanding the sample to include firms from other transitioning economies, offering a broader and more nuanced perspective on this critical issue.

References

- Agnihotri, R., & Gabler, C. B. (2024). Pursuing competitive advantages in the interactive digital marketplace and resource-advantage (R-A) theory of competition: a research agenda. *Journal of Marketing Management*, 1–19.
- Aksoy, M., Yilmaz, M. K., Golgeci, I., Tatoglu, E., Canci, M., & Hızarcı, A. E. (2024). Untangling the influence of corporate sustainability on export intensity: The moderating role of R&D intensity. *Journal of International Marketing*, 32(4), 38–57.
- Ancillai, C., Sabatini, A., Gatti, M., & Perna, A. (2023). Digital technology and business model innovation: A systematic literature review and future research agenda. *Technological Forecasting and Social Change*, 188, 122307.
- Antony, J., Swarnakar, V., Sony, M., McDermott, O., & Jayaraman, R. (2024). How do organizational performances vary between early adopters and late adopters of Quality 4.0? An exploratory qualitative study. *The TQM Journal*, 36(8), 2338–2359.
- Appio, F. P., Frattini, F., Petruzzelli, A. M., & Neirotti, P. (2021). Digital transformation and innovation management: A synthesis of existing research and an agenda for future studies. *Journal of Product Innovation Management*, 38(1), 4–20.
- Ardito, L., Filieri, R., Raguseo, E., & Vitari, C. (2024). Artificial intelligence adoption and revenue growth in European SMEs: synergies with IoT and big data analytics. *Internet Research*, Vol. ahead-of-print No. ahead-of-print.
- Attaran, M., & Woods, J. (2018). Cloud computing technology: improving small business performance using the Internet. *Journal of Small Business & Entrepreneurship*, 31(6), 495–519.
- Babalola, E. O., Wu, B., Fosu, E., & Shakeel, N. (2025). Exploring whether digital technologies ameliorate firms' servitization: the moderating role of organizational slack and research and development intensity. *Journal of Manufacturing Technology Management*, 36(2), 334–357.
- Bachmann, P., & Frutos-Bencze, D. (2022). R&D and innovation efforts during the COVID-19 pandemic: The role of universities. *Journal of Innovation & Knowledge*, 7(4), 100238.
- Bejlegaard, M., Sarivan, I., & Waehrens, B. V. (2021). The influence of digital technologies on supply chain coordination strategies. *Journal of Global Operations and Strategic Sourcing*, 14(4), 636–658.
- Besiroglu, T., Emery-Xu, N., & Thompson, N. (2024). Economic impacts of AI-augmented R&D. *Research Policy*, 53(7), 105037.
- Blank, T. H. (2024). Developing incremental innovation in the high-tech industry: the effect of age and tenure in research and development teams. *Cross Cultural & Strategic Management*, 31(2), 289–308.
- Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal*, 1(2), 141–162.
- Bresciani, S., Ciampi, F., Meli, F., & Ferraris, A. (2021). Using big data for co-innovation processes: Mapping the field of data-driven innovation, proposing theoretical developments and providing a research agenda. *International Journal of Information Management*, 60, 102347.

- Butschan, J., Heidenreich, S., Weber, B., & Kraemer, T. (2019). Tackling hurdles to digital transformation—The role of competencies for successful industrial internet of things (IoT) implementation. *International Journal of Innovation Management*, 23(04), 1950036.
- Chang, Y., Iakovou, E., & Shi, W. (2019). Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities. *International Journal of Production Research*, 58(7), 2082–2099.
- Chaudhuri, R., Chatterjee, S., Vrontis, D., & Chaudhuri, S. (2022). Innovation in SMES, AI dynamism, and Sustainability: the current situation and way forward. *Sustainability*, 14(19), 12760.
- Chen, Y., & Wang, Y. (2024). Governance path of the digital innovation ecosystem in advanced manufacturing cluster: a case study of China's Tai-Xin Integrated Economic Zone. *Kybernetes*, Vol. ahead-of-print No. ahead-of-print.
- Das, T. K., & Teng, B. (2000). A Resource-Based Theory of strategic alliances. *Journal of Management*, 26(1), 31–61.
- De Propriis, L., & Bailey, D. (2021). Pathways of regional transformation and Industry 4.0. *Regional Studies*, 55(10–11), 1617–1629.
- Dearing, J. W. (2009). Applying diffusion of innovation theory to intervention development. *Research on Social Work Practice*, 19(5), 503–518.
- Eisenhardt, K. M. (1989). Agency Theory: An Assessment and review. *Academy of Management Review*, 14(1), 57–74.
- Frattini, F., Bianchi, M., De Massis, A., & Sikimic, U. (2013). The role of early adopters in the diffusion of new products: Differences between platform and nonplatform innovations. *Journal of Product Innovation Management*, 31(3), 466–488.
- Gao, P., Wu, W., & Yang, Y. (2022). Discovering themes and trends in digital transformation and innovation research. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(3), 1162–1184.
- Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., ... & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8, 100118.
- Giustiziero, G., Kretschmer, T., Somaya, D., & Wu, B. (2021). Hyperspecialization and hyperscaling: A resource-based theory of the digital firm. *Strategic Management Journal*, 44(6), 1391–1424.
- Guo, X., Song, X., Dou, B., Wang, A., & Hu, H. (2022). Can digital transformation of the enterprise break the monopoly?. *Personal and Ubiquitous Computing*, 27(4), 1629–1642.
- Gupta, S., Drave, V. A., Bag, S., & Luo, Z. (2019). Leveraging smart supply chain and information system agility for supply chain flexibility. *Information Systems Frontiers*, 21(3), 547–564.
- Hao, X., Wen, S., Xue, Y., Wu, H., & Hao, Y. (2023). How to improve environment, resources and economic efficiency in the digital era?. *Resources Policy*, 80, 103198.
- Hong, P. W., Dorasamy, M., Hong, L. J., & Malarvizhi, C. A. N. (2024). Exploring Cloud Enterprise Resource Planning and Open Innovation for Small and Medium Enterprises: Insights from Practitioners. *Journal of Open Innovation Technology Market and Complexity*, 10(4), 100418.
- Hooy, G., Hooy, C., & Chee, H. (2019). Ultimate Ownership, Control Mechanism, and Firm Performance: Evidence from Malaysian Firms. *Emerging Markets Finance and Trade*, 56(15), 3805–3828.
- Kindermann, B., Beutel, S., De Lomana, G. G., Strese, S., Bendig, D., & Brettel, M. (2020). Digital orientation: Conceptualization and operationalization of a new strategic orientation. *European Management Journal*, 39(5), 645–657.
- Kruger, S., & Steyn, A. A. (2023). Developing breakthrough innovation capabilities in university ecosystems: A case study from South Africa. *Technological Forecasting and Social Change*, 198, 123002.

- Kwok, A. O. J., & Treiblmaier, H. (2023). Blockchain technology as a driver of economic development in small economies: a dynamic capabilities framework. *Journal of Decision System*, 33(3), 413–438.
- Li, G., Jin, Y., & Gao, X. (2022). Digital transformation and pollution emission of enterprises: Evidence from China's micro-enterprises. *Energy Reports*, 9, 552–567.
- Li, J., Li, J., & Wu, Y. (2024). Product R&D strategy of new energy vehicle enterprises embedded with technical services under the subsidy threshold. *Journal of the Knowledge Economy*.
- Li, S., & Xia, J. (2008). The roles and performance of state firms and non-state firms in China's economic transition. *World Development*, 36(1), 39–54.
- Lin, W. L., Ho, J. A., Sambasivan, M., Yip, N., & Mohamed, A. B. (2021). Influence of green innovation strategy on brand value: The role of marketing capability and R&D intensity. *Technological Forecasting and Social Change*, 171, 120946.
- Liu, G., Liu, J., Gao, P., Yu, J., & Pu, Z. (2024). Understanding mechanisms of digital transformation in state-owned enterprises in China: An institutional perspective. *Technological Forecasting and Social Change*, 202, 123288.
- Liu, T., Xue, D., Fang, Y., & Zhang, K. (2023). The impact of differentiated development of the digital economy on employment quality—An empirical analysis based on provincial data from China. *Sustainability*, 15(19), 14176.
- Liu, Y., Ni, Z., Karlsson, M., & Gong, S. (2021). Methodology for digital transformation with internet of things and cloud computing: A practical guideline for innovation in small- and medium-sized enterprises. *Sensors*, 21(16), 5355.
- Mahoney, J. T. (2001). A resource-based theory of sustainable rents. *Journal of Management*, 27(6), 651–660.
- Manroop, L., Malik, A., & Milner, M. (2024). The ethical implications of big data in human resource management. *Human Resource Management Review*, 34(2), 101012.
- Mansha, S., Bhutta, A. I., Antonucci, G., & Hooy, C. (2022). Do political connections matter for firm trade credit?. *Emerging Markets Finance and Trade*, 58(14), 4014–4032.
- Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. *Technovation*, 122, 102623.
- Marion, T. J., & Fixson, S. K. (2020). The transformation of the innovation process: How digital tools are changing work, collaboration, and organizations in new product development. *Journal of Product Innovation Management*, 38(1), 192–215.
- Matheus, R., Janssen, M., & Janowski, T. (2021). Design principles for creating digital transparency in government. *Government Information Quarterly*, 38(1), 101550.
- Muldoon, J., Cant, C., Wu, B., & Graham, M. (2024). A typology of artificial intelligence data work. *Big Data & Society*, 11(1).
- Niebel, T., Rasel, F., & Viète, S. (2018). BIG data – BIG gains? Understanding the link between big data analytics and innovation. *Economics of Innovation and New Technology*, 28(3), 296–316.
- Nunes, P. M., Serrasqueiro, Z., & Leitão, J. (2012). Is there a linear relationship between R&D intensity and growth? Empirical evidence of non-high-tech vs. high-tech SMEs. *Research Policy*, 41(1), 36–53.
- Onjewu, A. E., Walton, N., & Koliousis, I. (2023). Blockchain agency theory. *Technological Forecasting and Social Change*, 191, 122482.
- Padgett, R. C., & Galan, J. I. (2009). The effect of R&D intensity on corporate social responsibility. *Journal of Business Ethics*, 93(3), 407–418.
- Park, J., & Kim, Y. B. (2019). Factors activating big data adoption by Korean firms. *Journal of Computer Information Systems*, 61(3), 285–293.

- Pu, T., & Zulkafli, A. H. (2024). Managerial ownership and corporate innovation: Evidence of patenting activity from Chinese listed manufacturing firms. *Cogent Business & Management*, 11(1), 2289202.
- Ranjan, J., & Foropon, C. (2020). Big data analytics in building the competitive intelligence of organizations. *International Journal of Information Management*, 56, 102231.
- Rong, X. (2014). word2vec Parameter Learning Explained. *arXiv (Cornell University)*.
- Saarikko, T., Westergren, U. H., & Blomquist, T. (2020). Digital transformation: Five recommendations for the digitally conscious firm. *Business Horizons*, 63(6), 825–839.
- Schneckenberg, D., Benitez, J., Klos, C., Velamuri, V. K., & Spieth, P. (2021). Value creation and appropriation of software vendors: A digital innovation model for cloud computing. *Information & Management*, 58(4), 103463.
- Schniederjans, D. G., & Hales, D. N. (2016). Cloud computing and its impact on economic and environmental performance: A transaction cost economics perspective. *Decision Support Systems*, 86, 73–82.
- Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of things and big data as enablers for business digitalization strategies. *Technovation*, 98, 102173.
- Şimşek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. D. (2019). New ways of seeing big data. *Academy of Management Journal*, 62(4), 971–978.
- Steffen, J. H., Gaskin, J. E., Meservy, T. O., Jenkins, J. L., & Wolman, I. (2019). Framework of affordances for virtual reality and augmented reality. *Journal of Management Information Systems*, 36(3), 683–729.
- Sun, C., Lin, Z., Vochozka, M., & Vincúrová, Z. (2022). Digital transformation and corporate cash holdings in China's A-share listed companies. *Oeconomia Copernicana*, 13(4), 1081–1116.
- Sun, Y., Tan, C., Lim, K. H., Liang, T., & Yeh, Y. (2024). Strategic contexts, strategic orientations and organisational technology adoption: A configurational approach. *Information Systems Journal*, 34(4), 1355–1401.
- Tang, Q., Guo, J., & Huang, Z. (2021). The role of accounting conservatism in M&A target selection. *China Journal of Accounting Research*, 14(4), 100204.
- Tee, C. M., Teoh, T. M., & Hooy, C. W. (2022). Political Connection Types and Corporate Tax Avoidance: Evidence from Malaysia. *Malaysian Journal of Economic Studies*, 59(2), 199–220.
- Tu, W., & He, J. (2022). Can Digital Transformation Facilitate Firms' M&A: Empirical Discovery Based on Machine Learning. *Emerging Markets Finance and Trade*, 59(1), 113–128.
- Urbinati, A., Chiaroni, D., Chiesa, V., & Frattini, F. (2020). The role of digital technologies in open innovation processes: an exploratory multiple case study analysis. *R&D Management*, 50(1), 136–160.
- Vassakis, K., Petrakis, E., & Kopanakis, I. (2017). Big Data Analytics: applications, prospects and challenges. In *Lecture notes on data engineering and communications technologies* (pp. 3–20).
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144.
- Wang, H., & Li, B. (2024). Research on the influence mechanism of digital innovation in high-tech enterprises — An inter-organizational perspective. *Managerial and Decision Economics*, 45(5), 2988–3001.
- Wang, H., Qi, S., Zhou, C., Zhou, J., & Huang, X. (2022). Green credit policy, government behavior and green innovation quality of enterprises. *Journal of Cleaner Production*, 331, 129834.
- Wang, L., Liu, Z., Liu, A., & Tao, F. (2021). Artificial intelligence in product lifecycle management. *The International Journal of Advanced Manufacturing Technology*, 114(3–4), 771–796.
- Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5(2), 171–180.

- Wong, W., & Hooy, C. (2018). Do types of political connection affect firm performance differently? *Pacific-Basin Finance Journal*, 51, 297–317.
- Wong, W., & Hooy, C. (2024). Is political connection associated with corporate risk-taking? Malaysian evidence. *The Singapore Economic Review*, 1–30.
- Wu, Y., Li, H., Luo, R., & Yu, Y. (2024). How digital transformation helps enterprises achieve high-quality development? Empirical evidence from Chinese listed companies. *European Journal of Innovation Management*, 27(8), 2753-2779.
- Xu, R., & Guan, E. (2022). Can blockchain innovation promote total factor productivity? Evidence from Chinese-listed firms. *Applied Economics*, 55(6), 653–670.
- Yuan, R., & Wen, W. (2018). Managerial foreign experience and corporate innovation. *Journal of Corporate Finance*, 48, 752–770.
- Yuan, S., & Pan, X. (2023). Inherent mechanism of digital technology application empowered corporate green innovation: Based on resource allocation perspective. *Journal of Environmental Management*, 345, 118841.
- Zhang, W., Zhao, S., Wan, X., & Yao, Y. (2021). Study on the effect of digital economy on high-quality economic development in China. *PLoS ONE*, 16(9), e0257365.
- Zheng, X., Wang, F., Liu, S., Wang, H., & Zhang, D. (2023). Outward foreign direct investment, dynamic capabilities and radical innovation performance: empirical evidence from Chinese high-tech companies. *Chinese Management Studies*, 18(4), 921–953.