Simen Owen Academic Proceedings Series

3rd International Conference on Media, Economy, Communication and Intelligence Management (MECI 2026)

Article

Knowledge Management Capability and Innovation Performance in Enterprise Digital Transformation: Evidence from Text Mining

Zheng Wu[1,*]

[1] Ningbo University Business School, Ningbo, China
* Correspondence: Zheng Wu, Ningbo University Business School, Ningbo, China

Abstract: This study examines how knowledge management capability affects corporate innovation performance in the context of enterprise digital transformation. Drawing on knowledge-based theory and digital transformation research, the study constructs a firm-year panel dataset of U.S. listed companies from 2015 to 2024 by integrating open-source data from SEC EDGAR 10-K filings, SEC CompanyFacts, and USPTO PatentsView. To capture knowledge management capability, this study applies natural language processing and text mining methods to corporate annual report disclosures. Specifically, a text-based index is developed from four dimensions: knowledge acquisition, knowledge sharing, knowledge integration, and knowledge application. Digital transformation is also measured through textual indicators related to digital technologies and digital business practices. Patent-based indicators are used to measure corporate innovation performance. The empirical results show that knowledge management capability has a positive effect on innovation performance, and this effect is strengthened by digital transformation. Further analysis suggests that the effect is more pronounced in high-tech and R&D-intensive firms. This study contributes to the literature by linking knowledge management theory with natural language processing methods, providing evidence on how data-driven measurement can advance management research. The findings also suggest that firms should build digital and knowledge-based mechanisms to improve innovation outcomes.

Keywords: knowledge management; digital transformation; innovation performance; text mining; natural language processing

Received: 12 May 2026
Revised: 23 June 2026
Accepted: 07 July 2026
Published: 10 July 2026

1. Introduction

Enterprise digital transformation has become a key driver of organizational change and competitive advantage. With the development of artificial intelligence, cloud computing, big data analytics, digital platforms, and automation technologies, firms increasingly use digital tools to restructure business processes, improve decision-making, and create new value. However, digital transformation is not merely technology adoption [1]. Its effectiveness depends on whether firms can identify, absorb, integrate, and apply knowledge systematically. Thus, knowledge management capability has become a critical organizational capability for explaining why some firms achieve stronger innovation outcomes than others in the digital economy.

Innovation performance is closely related to how firms process and utilize knowledge. Firms with stronger knowledge management capability are more likely to acquire external knowledge from customers, suppliers, competitors, and technological communities, while integrating internal knowledge across departments and business units. This capability helps transform fragmented information into useful insights, reduce uncertainty in innovation activities, and support new products, services, and processes. Under digital transformation, the value of knowledge management becomes more prominent because digital technologies increase the volume, speed, and complexity of information flows [2]. Firms lacking effective knowledge management may face information overload, weak knowledge sharing, and inefficient innovation conversion, even with substantial digital investment.

Although existing studies have examined the relationship between knowledge management and innovation, several limitations remain. First, many studies measure knowledge management capability through questionnaires or case evidence. These approaches are useful for understanding organizational mechanisms but often suffer from limited sample size, subjective bias, and weak comparability across firms and years. Second, prior research focuses mainly on the direct effect of digital transformation or R&D investment on innovation, while paying less attention to the knowledge-based capability that converts digital resources into innovation outputs. Third, traditional measurement methods are less suitable for capturing the dynamic and textual nature of enterprise knowledge practices, especially as firms disclose strategic information in annual reports, regulatory filings, and public documents.

Recent advances in natural language processing provide a new opportunity to address these limitations [1]. Corporate textual disclosures contain rich information about knowledge acquisition, knowledge sharing, organizational learning, technological upgrading, digital strategy, and innovation activities. By applying text mining methods to open-source corporate documents, researchers can construct scalable and comparable indicators of knowledge management capability. Compared with manually coded indicators, text-based measures can cover larger samples and longer periods, while capturing changes in firms' strategic emphasis and organizational capabilities. This methodological shift also reflects the integration of management research and data science.

Based on this background, this study investigates whether knowledge management capability improves corporate innovation performance in the context of enterprise digital transformation [1]. Specifically, it constructs a text mining-based measure of knowledge management capability from corporate textual disclosures and examines its effect on patent-based innovation performance. The study further tests whether digital transformation strengthens this relationship and whether the effect is more pronounced in high-tech and R&D-intensive firms.

This research makes three contributions. First, it enriches the literature on knowledge management and innovation by providing empirical evidence based on open-source textual and innovation data. Second, it offers a data-driven measurement approach for knowledge management capability, reducing reliance on subjective survey indicators. Third, it contributes to digital transformation research by highlighting knowledge management capability as an important mechanism through which digitalization supports enterprise innovation. Its findings also provide practical implications for firms seeking to improve innovation through more effective knowledge acquisition, sharing, integration, and application [3].

2. Literature Review and Hypothesis Development

2.1. Digital Transformation and Enterprise Innovation

Digital transformation refers to the integration of digital technologies—including artificial intelligence, big data analytics, cloud computing, digital platforms, and automation systems—to fundamentally reshape business processes, organizational structures, and value creation mechanisms [3]. It entails more than mere technology adoption; it involves a comprehensive reconfiguration of resources, operational routines, and strategic decision-making frameworks.

Enterprise innovation constitutes a central outcome of digital transformation. Digital technologies can enhance innovation performance by broadening access to external knowledge sources, accelerating product development cycles, lowering communication barriers, and enabling data-informed decision-making. However, this enhancement is not guaranteed [4]. Organizations may undertake substantial digital investments without achieving commensurate innovation gains if they lack the capacity to convert digital information into actionable knowledge. Moreover, digital transformation intensifies both the volume and complexity of information flows, potentially leading to information overload and coordination challenges. Consequently, its impact on innovation depends critically not only on technological deployment but also on organizational capabilities for knowledge management, integration, and application.

2.2. Knowledge Management Capability and Innovation Performance

Knowledge management capability refers to a firm's ability to acquire, share, integrate, and apply knowledge in organizational activities. It reflects how effectively a firm transforms dispersed information and experience into valuable organizational knowledge. In innovation activities, knowledge serves as a critical strategic resource, as new products, services, and processes frequently emerge from knowledge recombination and the absorption of external knowledge [3].

Firms with strong knowledge management capability can acquire external knowledge from customers, suppliers, competitors, research institutions, and technological communities [5]. They also promote internal knowledge sharing across departments, reduce knowledge silos, and enhance organizational learning. Knowledge integration enables firms to combine technological, market, and managerial knowledge, while knowledge application converts accumulated knowledge into patents, new products, process improvements, and business model innovations.

From a resource-based perspective, knowledge management capability constitutes an intangible organizational capability that underpins competitive advantage [6]. From a dynamic capability perspective, it supports firms in sensing technological opportunities, seizing innovation opportunities, and reconfiguring resources amid environmental change. Consequently, firms possessing stronger knowledge management capability tend to achieve superior innovation performance.

2.2.1. H1: Knowledge Management Capability Has a Positive Effect on Corporate Innovation Performance.

Digital transformation may further strengthen this effect. Digital technologies provide richer data sources, faster information processing tools, and broader collaboration channels. However, these digital resources generate innovation value only when firms can interpret, integrate, and apply them effectively. In firms with more advanced digital transformation, knowledge management capability tends to be more effective, as digital systems enhance both the speed and scope of knowledge flow [6].

2.2.2. H2: Digital Transformation Strengthens the Positive Relationship between Knowledge Management Capability and Corporate Innovation Performance.

The impact of knowledge management capability can differ across firms. High-tech and R&D-intensive enterprises typically experience more rapid technological change, greater knowledge complexity, and heightened innovation pressure [7]. Their innovation efforts depend more critically on specialized knowledge, cross-functional collaboration, and continuous learning, making them more capable of converting knowledge management capability into tangible innovation outcomes.

2.2.3. H3: The Positive Effect of Knowledge Management Capability on Corporate Innovation Performance Is Stronger in High-Tech and R&D-intensive Firms.

2.3. Text Mining in Management Research

Traditional management research often measures organizational capabilities through surveys, interviews, or manually collected indicators [8]. Although useful for capturing managerial perceptions and internal mechanisms, these methods may face limitations including small sample sizes, subjective bias, and low comparability across firms and over time.

Corporate annual reports, regulatory filings, and public disclosures contain rich information about firms' strategies, capabilities, technologies, risks, and innovation activities. Text mining methods can extract structured indicators from these unstructured texts. Keyword dictionaries enable measurement of concepts such as knowledge management or digital transformation, while topic models uncover latent thematic structures. Advanced natural language processing techniques—including word embeddings and sentence embeddings—further capture semantic similarity and contextual meaning beyond simple word frequency counts.

In this study, text mining is employed to construct indicators of knowledge management capability and digital transformation from corporate textual disclosures. This approach enhances measurement scalability and enables analysis of a large, longitudinal sample of firms [8]. It also reflects the integration of management theory and data science, offering a more objective and dynamic perspective on how organizational capabilities influence innovation performance.

3. Data, Variables, and Text Mining Method

3.1. Data Sources and Sample Construction

This study constructs a firm-year panel dataset by integrating multiple open-source data sources, including SEC EDGAR 10-K filings, SEC CompanyFacts, USPTO PatentsView, and, optionally, OpenAlex. The sample covers U.S. listed companies from 2015 to 2024, a period during which digital transformation advanced significantly and textual, financial, and patent data are relatively comprehensive.

SEC EDGAR 10-K filings provide annual report texts containing information on business models, risk factors, management discussion and analysis, technological strategies, and innovation activities [9]. SEC CompanyFacts supplies XBRL-based financial data for constructing firm-level control variables. USPTO PatentsView serves to quantify patent-based innovation performance. OpenAlex is used solely as an optional supplementary source for characterizing the external knowledge environment. Table 1 summarizes the data sources employed in this study and their corresponding research purposes.

Table 1. Data Sources and Research Use

Data source	Data type	Main use
SEC EDGAR 10-K filings	Annual report text	Construct knowledge management capability and digital transformation indices
SEC CompanyFacts	Financial statement data	Construct control variables
USPTO PatentsView	Patent data	Measure innovation performance
OpenAlex	Scholarly publication data	Optional measure of external knowledge environment

The initial sample comprises U.S. listed firms with available 10-K filings between 2015 and 2024. Financial and utility firms are excluded due to distinctive regulatory frameworks and financial reporting structures. Firm-year observations with missing core textual, financial, or patent data are omitted. Continuous variables are winsorized at the 1st and 99th percentiles [10]. Datasets are matched using company names, ticker symbols, CIK identifiers, and patent assignee names. Firm names are standardized by removing punctuation, legal suffixes, and common abbreviations.

3.2. Variable Measurement

The dependent variable is corporate innovation performance. The primary measure is the number of patent applications filed by firm i in year t, as patent applications reflect contemporaneous innovation activities and are less subject to delays associated with examination and grant processes. Patent grants and forward patent citations serve as alternative measures. Given the typical right-skewness of patent count distributions, innovation performance is calculated as:

$$\mathrm{Innovation}_{i,t} = \ln(1 + \mathrm{PatentCount}_{i,t}) \quad (1)$$

where $\mathrm{PatentCount}_{i,t}$ represents patent applications, grants, or citations [11].

The core independent variable is knowledge management capability, denoted as $\mathrm{KMC}_{i,t}$. This refers to a firm's capacity to acquire, share, integrate, and apply knowledge. As this capability is not directly observable, it is operationalized using textual disclosures in corporate reports. The index comprises four dimensions: knowledge acquisition, knowledge sharing, knowledge integration, and knowledge application.

The moderating variable is digital transformation, denoted as $\mathrm{DT}_{i,t}$. It reflects the degree to which a firm emphasizes digital technologies and digital business practices in its annual reports, including artificial intelligence, big data, cloud computing, automation, digital platforms, blockchain, data analytics, machine learning, and intelligent manufacturing.

Control variables include firm size, leverage, profitability, R&D intensity, firm growth, and capital intensity. Industry and year fixed effects are incorporated to account for sector-specific characteristics and macroeconomic fluctuations. Table 2 presents the definitions and measurement approaches for the main variables.

Table 2. Variable Definitions

Variable	Symbol	Measurement
Innovation performance	$Innovation_{i,t}$	Natural logarithm of one plus patent applications, grants, or citations
Knowledge management capability	$KMC_{i,t}$	Text-based index of knowledge acquisition, sharing, integration, and application
Digital transformation	$DT_{i,t}$	Text-based index of digital technology and digital business terms
Firm size	$Size_{i,t}$	Natural logarithm of total assets
Leverage	$Lev_{i,t}$	Total liabilities divided by total assets
Profitability	$ROA_{i,t}$	Net income divided by total assets
R&D intensity	$RD_{i,t}$	R&D expenditure divided by total assets or revenue
Growth	$Growth_{i,t}$	Annual revenue growth rate
Capital intensity	$Capital_{i,t}$	Fixed assets divided by total assets

3.3. Text Preprocessing

Before constructing text-based indicators, 10-K reports undergo cleaning and preprocessing. Raw filings frequently contain HTML tags, tables, page headers, legal notices, and repetitive formatting elements, all of which are removed to extract clean textual content. The analysis prioritizes business descriptions, management discussion and analysis, and risk-related sections, as these contain information pertinent to corporate strategy, knowledge activities, technology adoption, and innovation orientation.

The preprocessing procedure includes lowercasing, stop-word removal, elimination of irrelevant symbols, and lemmatization [5]. Boilerplate expressions and common legal terms are filtered out when they contribute no substantive analytical value. Each firm-year document is then converted into a structured text corpus. Figure 1 illustrates the overall text mining framework used to construct firm-level indicators of knowledge management capability and digital transformation from 10-K filings.

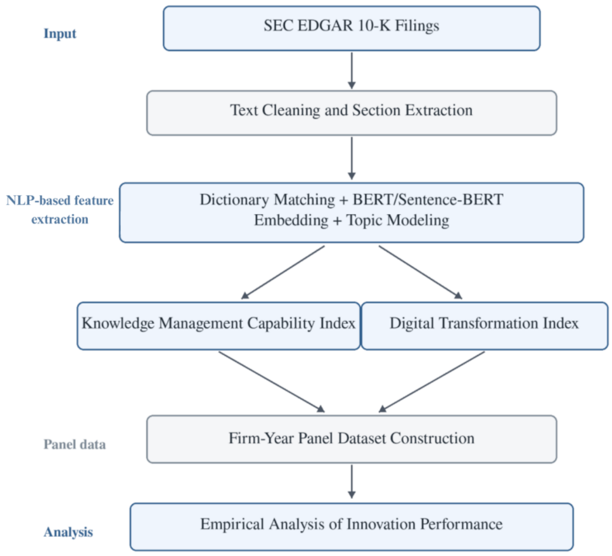

Figure 1. Text Mining Framework for Constructing Firm-Level Indices

3.4. Construction of the Knowledge Management Capability Index

This study constructs the knowledge management capability index through an integrated approach combining dictionary-based text mining, semantic embedding, and topic modeling [4]. A domain-specific dictionary is developed across four core dimensions—knowledge acquisition, sharing, integration, and application—incorporating terms such as "learning," "collaboration," "knowledge sharing," "training," "expertise," "technology transfer," "integration," "R&D cooperation," and "best practices." Term frequencies are normalized by document length to ensure comparability.

Second, BERT or Sentence-BERT embeddings are employed to capture contextual and semantic meaning beyond lexical matching [3]. For each dimension, a standardized descriptive statement is formulated; the semantic similarity between each firm-year textual document and the corresponding dimension description is computed as:

$$\mathrm{Semantic}_{i,t}^{k} = \cos(\mathrm{E}(\mathrm{Text}_{i,t}), \mathrm{E}(\mathrm{Description}^{k})) \qquad (2)$$

where $\mathrm{E}(\cdot)$ represents the text vector generated by BERT or Sentence-BERT, and $\mathrm{Description}^{k}$ represents the standard description of the k-th dimension.

Third, unsupervised topic modeling—using either LDA or BERTopic—is applied to extract latent themes associated with organizational learning, digital collaboration, technology development, and innovation management. Topic relevance scores for these domains are incorporated into the index [5]. The dimension-level composite score is calculated as:

$$\mathrm{Score}_{i,t}^{k} = \alpha \mathrm{Dictionary}_{i,t}^{k} + \beta \mathrm{Semantic}_{i,t}^{k} + \gamma \mathrm{Topic}_{i,t}^{k} \qquad (3)$$

where $\mathrm{Dictionary}_{i,t}^{k}$, $\mathrm{Semantic}_{i,t}^{k}$, and $\mathrm{Topic}_{i,t}^{k}$ denote the dictionary-based score, semantic similarity score, and topic relevance score, respectively. In the baseline specification, equal weighting is assigned to each component. The final index aggregates these dimension-level scores.

$$\mathrm{KMC}_{i,t} = \frac{1}{4}\sum_{k=1}^{4} \mathrm{Score}_{i,t}^{k} \qquad (4)$$

3.5. Construction of the Digital Transformation Index

The digital transformation index is constructed following a comparable methodology [2]. A domain-specific dictionary is developed using terms including "artificial intelligence," "big data," "cloud computing," "digital platform," "automation," "machine learning," "data analytics," "blockchain," "Internet of Things," and "digital ecosystem." Term frequencies are normalized by document length to ensure comparability across texts.

To enhance measurement precision, semantic embeddings—specifically BERT or Sentence-BERT—are employed to compute the similarity between corporate disclosures and standardized digital transformation descriptors [12]. The final index is:

$DT_{i,t} = \lambda DigitalDictionary_{i,t} + (1-\lambda)DigitalSemantic_{i,t}$(5)

where $DigitalDictionary_{i,t}$ represents digital keyword intensity, $DigitalSemantic_{i,t}$ represents semantic similarity, and λ is the weighting parameter [13]. In the baseline analysis, $\lambda = 0.5$, while alternative weights are applied in robustness checks.

3.6. Optional External Knowledge Environment Measure

OpenAlex serves as an optional source for measuring the external knowledge environment. Drawing on publication data, an industry-level knowledge intensity variable can be constructed by quantifying the volume of scientific publications associated with each industry's technological domain. This variable enables control for cross-industry variation in external knowledge availability and facilitates examination of whether firms derive greater innovation benefits from knowledge management capability in more knowledge-intensive contexts.

This chapter establishes a reproducible framework for constructing firm-level knowledge management capability and digital transformation indices from open corporate textual data. Integrating textual, patent, and financial data, the study develops a data-driven foundation to investigate how organizational knowledge capability shapes innovation performance amid enterprise digital transformation.

4. Empirical Model

To examine the effect of knowledge management capability on corporate innovation performance, this study constructs a firm-year panel regression model. The baseline model is specified as follows:

$Innovation_{i,t} = \beta_0 + \beta_1 KMC_{i,t-1} + \sum\beta_j Controls_{i,t-1} + \mu_i + \lambda_t + \varepsilon_{i,t}$ (6)

where $Innovation_{i,t}$ represents the innovation performance of firm i in year t, measured by the logarithm of one plus patent applications, patent grants, or patent citations. $KMC_{i,t-1}$ denotes the lagged knowledge management capability index. The one-year lag mitigates potential reverse causality. $Controls_{i,t-1}$ includes lagged firm-level control variables, such as firm size, leverage, profitability, R&D intensity, firm growth, and capital intensity. μ_i and λ_t represent firm and year fixed effects, respectively, and $\varepsilon_{i,t}$ is the error term.

The coefficient of interest is β_1. A significantly positive β_1 indicates that stronger knowledge management capability is associated with higher innovation performance, supporting H1.

To test whether digital transformation strengthens this relationship, the interaction term between knowledge management capability and digital transformation is added:

$Innovation_{i,t} = \beta_0 + \beta_1 KMC_{i,t-1} + \beta_2 DT_{i,t-1} + \beta_3 KMC_{i,t-1} \times DT_{i,t-1} + \sum\beta_j Controls_{i,t-1} + \mu_i + \lambda_t + \varepsilon_{i,t}$ (7)

where $DT_{i,t-1}$ represents the lagged digital transformation index. The interaction term $KMC_{i,t-1} \times DT_{i,t-1}$ tests whether digital transformation modifies the marginal effect of knowledge management capability on innovation performance. A significantly positive β_3 indicates that digital transformation enhances this positive association, supporting H2.

To test H3, the sample is divided into high-tech and non-high-tech firms, or high- and low-R&D-intensive firms. The baseline model is estimated separately for each subsample [14]. If the coefficient of $KMC_{i,t-1}$ is larger and more statistically significant in high-tech or R&D-intensive firms, H3 is supported.

5. Empirical Results and Robustness Checks

5.1. Descriptive Statistics and Correlation Analysis

Table 3 reports the descriptive statistics and correlations of the main variables. The mean value of Innovation is 0.684, with a standard deviation of 1.104, indicating considerable variation in patent output across firms. KMC and DT also exhibit substantial dispersion, reflecting heterogeneity in firms' knowledge management capability and digital transformation levels.

Table 3. Descriptive Statistics and Correlations

Variable	Mean	SD	Min	Max	1	2	3	4	5
1. Innovation	0.684	1.104	0.000	6.237	1.000				
2. KMC	0.317	0.116	0.072	0.681	0.236	1.000			
3. DT	0.244	0.138	0.031	0.754	0.189	0.284	1.000		
4. Size	8.625	1.741	4.286	13.947	0.311	0.173	0.205	1.000	
5. RD	0.047	0.064	0.000	0.328	0.267	0.214	0.158	0.096	1.000

The correlation results offer preliminary evidence for the hypothesized relationships. KMC is positively associated with $\mathrm{Innovation}$, suggesting that stronger knowledge management capability tends to coincide with higher innovation output [15]. DT is likewise positively associated with both $\mathrm{Innovation}$ and KMC. Correlations among explanatory variables remain moderate, alleviating concerns about severe multicollinearity.

5.2. Baseline Regression and Moderating Effect

Table 4 presents the baseline regression and moderating effect results. In Model 1, the coefficient of $\mathrm{KMC_{i,t-1}}$ is positive and statistically significant, indicating that firms with stronger knowledge management capability exhibit higher innovation performance in the subsequent year. This finding supports Hypothesis 1.

Table 4. Baseline Regression and Moderating Effect Results

Variable	Model 1	Model 2	Model 3
$KMC_{i,t-1}$	0.418**	0.392**	0.337**
	(2.54)	(2.41)	(2.18)
$DT_{i,t-1}$		0.286*	0.214
		(1.87)	(1.42)
$KMC_{i,t-1} \times DT_{i,t-1}$			0.531**
			(2.36)
Size	0.119***	0.113***	0.110***
Leverage	-0.082	-0.076	-0.079
ROA	0.204*	0.196*	0.188*
RD intensity	1.427***	1.389***	1.361***
Growth	0.057	0.061	0.058
Capital intensity	-0.143*	-0.137*	-0.132
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	18,426	18,426	18,426
Adjusted (R^2)	0.412	0.417	0.423

Notes: t-statistics are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Model 2 introduces the digital transformation index, yielding a positive coefficient. Model 3 incorporates the interaction term $\mathrm{KMC_{i,t-1}} \times \mathrm{DT_{i,t-1}}$; its coefficient is positive and statistically significant, suggesting that digital transformation enhances the positive association between knowledge management capability and innovation performance. Thus, Hypothesis 2 is supported.

These findings imply complementarity between digital transformation and knowledge management capability [4]. Digital technologies supply data resources, analytical tools, and collaborative platforms; however, their contribution to innovation hinges on firms' capacity to interpret, synthesize, and deploy knowledge effectively. As illustrated in Figure 2, the relationship between knowledge management capability and innovation performance becomes more pronounced under conditions of high digital transformation.

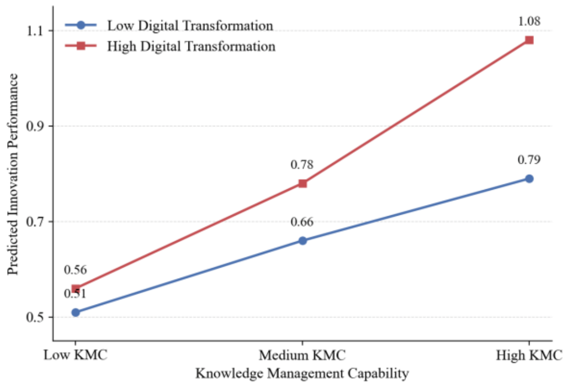

Figure 2. Moderating Effect of Digital Transformation

5.3. Heterogeneity Analysis and Robustness Checks

Table 5 reports the heterogeneity analysis and robustness checks. The coefficient of KMC is larger and more statistically significant in high-tech firms compared to non-high-tech firms. A similar pattern emerges in firms with high R&D intensity, whereas the effect is weaker and statistically insignificant in low R&D-intensive firms. These findings support Hypothesis 3 and indicate that knowledge management capability yields greater innovation benefits under conditions of higher technological complexity and innovation pressure.

Table 5. Heterogeneity and Robustness Checks

Test	Dependent variable / sample	Coefficient of $\boldsymbol{KMC}$
High-tech firms	Patent applications	0.516***
Non-high-tech firms	Patent applications	0.247*
High R&D intensity	Patent applications	0.489**
Low R&D intensity	Patent applications	0.221
Alternative measure	Patent grants	0.364**
Alternative measure	Invention patents	0.401**
Alternative measure	Patent citations	0.295*
Dictionary-only index	Patent applications	0.318**
Topic-model index	Patent applications	0.287*
With industry FE	Patent applications	0.351**
Excluding extremes	Patent applications	0.376**
Two-year lag	Patent applications	0.304*

The robustness checks confirm the stability of the results. The positive effect of KMC persists when innovation performance is measured using alternative indicators—namely, patent grants, utility patents, and patent citations. Results remain consistent when the knowledge management capability index is reconstructed using either dictionary-based or topic-modeling approaches. The main conclusion holds unchanged after incorporating industry fixed effects, excluding outliers, and applying a two-year lag structure.

Overall, the findings demonstrate that knowledge management capability enhances corporate innovation performance, digital transformation amplifies this effect, and the enhancement is more pronounced in high-tech and R&D-intensive firms [8].

6. Conclusion and Implications

This study examines the impact of knowledge management capability on corporate innovation performance under enterprise digital transformation. By integrating open-source data from SEC EDGAR 10-K filings, SEC CompanyFacts, and USPTO PatentsView, it constructs a firm-year panel dataset of U.S. listed companies from 2015 to 2024. Knowledge management capability and digital transformation are measured through text mining of annual report disclosures, while innovation performance is measured by patent-related indicators.

The empirical results show that knowledge management capability positively affects corporate innovation performance. Firms with stronger capabilities in knowledge acquisition, sharing, integration, and application are more likely to generate higher innovation output. Digital transformation strengthens this relationship. Digital technologies provide richer data resources, faster information processing, and broader collaboration channels, but their innovation value depends on whether firms can transform digital information into organizational knowledge. The effect of knowledge management capability is also stronger in high-tech and R&D-intensive firms, suggesting greater relevance under conditions of heightened technological complexity and innovation pressure.

This study makes three theoretical contributions. First, it enriches the literature on knowledge management and innovation by providing empirical evidence derived from open-source corporate disclosures. Second, it extends digital transformation research by demonstrating that digitalization does not automatically promote innovation; its effectiveness hinges on underlying knowledge-based organizational capabilities. Third, it introduces a text mining–based approach to measure knowledge management capability, bridging management theory with natural language processing methodologies.

The findings offer practical implications for managers. Digital transformation should not be approached solely as a technological investment. To enhance innovation performance, organizations must develop data-driven knowledge management mechanisms—including systematic external knowledge acquisition, structured internal knowledge sharing, cross-functional knowledge integration, and deliberate application of knowledge in R&D and business processes. Enterprises should also strengthen digital platforms and data infrastructure to support collaboration and continuous organizational learning.

This study has several limitations. First, the sample comprises U.S. listed companies and relies exclusively on English-language 10-K reports, potentially limiting generalizability. Second, patent-based indicators capture only certain dimensions of innovation and may not reflect service, organizational, or business model innovations. Third, text-based measures may be subject to semantic ambiguity, as corporate disclosures can reflect both actual capabilities and strategic impression management. Future research could address these constraints by incorporating multilingual data, alternative innovation metrics, and more sophisticated natural language processing techniques.

References

1. M. H. Nguyen and V. H. Dinh, "Digital sustainability empowerment and B2B customer co-creation: how CSR value creation drives green innovation performance in emerging market SMEs," *Journal of Business & Industrial Marketing*, pp. 1–20, 2026.
2. F. J. G. Silva et al., "Digital Transformation and Innovation: The Influence of Digital Technologies on Turnover from Innovation Activities and Types of Innovation," *Systems*, vol. 12, no. 9, p. 359, 2024.
3. T. A. Arsanti, N. Rupidara, and T. Bondarouk, "Managing knowledge flows within open innovation: knowledge sharing and absorption mechanisms in collaborative innovation," *Cogent Business & Management*, vol. 11, no. 1, p. 2351832, 2024.
4. F. Sonmez Cakir, A. A. Jahanshahi, Z. Adiguzel, and O. Kalaycioglu, "Knowledge capabilities and AI firm performance: the mediating roles of innovation culture and technology strategy," *Business Process Management Journal*, pp. 1–34, 2026.
5. A. D. Stajkovic and K. S. Stajkovic, *Human Sustainability and Cognitive Overload at Work: The Psychological Cost of Working*, Taylor & Francis, 2025, p. 183.
6. L. Danko and Z. Crhová, "Rethinking the role of knowledge sharing on organizational performance in knowledge-intensive business services," *Journal of the Knowledge Economy*, vol. 16, no. 4, pp. 13873–13893, 2025.
7. S. Li et al., "Exploring the effect of digital transformation on Firms' innovation performance," *Journal of Innovation & Knowledge*, vol. 8, no. 1, p. 100317, 2023.
8. S. Iqbal et al., "Impact of knowledge sharing and innovation on performance of organization: evidence from the pharmaceutical industry of Pakistan," *Cogent Business & Management*, vol. 12, no. 1, p. 2471536, 2025.
9. T. Dong, S. Yin, and N. Zhang, "The interaction mechanism and dynamic evolution of digital green innovation in the integrated green building supply chain," *Systems*, vol. 11, no. 3, p. 122, 2023.

10. P. J. Jordan, A. C. Troth, and H. Yan, "Objective and subjective measurement in applied business settings: Improving research in organizations," *Australian Journal of Management*, vol. 50, no. 1, pp. 8–31, 2025.
11. S. Hafeez, K. Shahzad, P. Helo, and M. F. Mubarak, "Knowledge management and SMEs' digital transformation: A systematic literature review and future research agenda," *Journal of Innovation & Knowledge*, vol. 10, no. 3, p. 100728, 2025.
12. M. Hassanzadeh, "Transformational Knowledge Management: A New Generation of Knowledge Management to Facilitate Digital Transformation," *Sciences & Techniques of Information Management*, vol. 7, no. 4, p. 1, 2021.
13. A. Fitriani, I. Irwandi, and I. N. Marani, "Digital transformation and organizational agility: The mediating role of knowledge management in enhancing firm performance," *Asian Journal of Applied Business and Management*, vol. 5, no. 1, pp. 61–74, 2026.
14. A. Di Vaio, R. Palladino, A. Pezzi, and D. E. Kalisz, "The role of digital innovation in knowledge management systems: A systematic literature review," *Journal of Business Research*, vol. 123, pp. 220–231, 2021.
15. Q. D. Ngo and T. L. Ngo, "Knowledge ecosystem integration in digital transformation: a configurational analysis of anticipatory governance, creative behaviour and digital culture in cultural industries," *Journal of Knowledge Management*, vol. 30, no. 2, pp. 824–845, 2026.