Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 24, No. 2, 2026

A Study on the Impact of Natural Language Processing Technology on Enterprises' New-Quality Productive Forces

Fan He, Qingqing Tan * and Xi Cao

School of Accounting, Chongqing Technology and Business University, 400067, China

* **Corresponding author**: (Email: tanqingqing@ctbu.edu.cn)

Abstract: This paper uses Chinese A-share listed companies from 2015 to 2023 as a sample to examine the impact of natural language processing technology on enterprises' new-quality productive forces. The study finds that: 1) the application of this technology by enterprises can significantly enhance the level of new-quality productive forces, with data as a factor and innovation-driven growth playing a mediating role in this process; 2) heterogeneity analysis indicates that the positive effect of this technology on new-quality productive forces is more pronounced in non-labor-intensive firms and in firms with a higher level of digital transformation. The study's findings provide empirical evidence for understanding the economic consequences of natural language processing technology and offer decision-making guidance for stakeholders.

Keywords: Natural Language Processing Technology; New-Quality Productive Forces; Data as a Factor of Production; Innovation-Driven Development.

1. Introduction

As a new-generation digital technology, the economic impact of natural language processing (NLP) on businesses has attracted significant attention from the academic community. In 2024, the Chinese government launched the "AI+" initiative, in which natural language processing technology is regarded as a key component aimed at driving the enhancement of new-quality productive forces. Against this backdrop, Chinese enterprises are actively embracing and applying natural language processing technology. According to data cited by the Cyberspace Administration of China, the enterprise adoption rate of generative AI in China reached 15% in 2023, with an estimated market scale of about RMB 14.4 trillion. Given that NLP is a core technological foundation of generative AI, this trend highlights the growing relevance of NLP-related technologies in corporate practice.

However, while existing research has largely explored the technology's effects in areas such as language understanding, generation, and processing efficiency, there has been little systematic examination of its economic consequences—particularly its impact on new-quality productive forces. Accordingly, this paper focuses on exploring the impact and mechanisms of enterprises' application of natural language intelligence processing technology on new-quality productivity. Its main marginal contributions are as follows: First, it expands the research boundaries of this technology and extends to a systematic examination of productivity outcomes, thereby enriching the empirical evidence; second, it validates potential impact pathways from the dual perspectives of data factor allocation and innovation-driven development, revealing the "black box" of how the application of emerging technologies influences new-quality productivity.

2. Literature Review

2.1. Natural Language Processing Technologies and Language Effects

The application of natural language processing (NLP) technology in research on corporate economic effects continues to deepen. Early studies primarily employed dictionary-based methods for text sentiment analysis. With the advancement of deep learning, pre-trained language models—such as BERT—have enhanced semantic understanding through bidirectional encoding mechanisms, providing new technical tools for the analysis of unstructured corporate text (Devlin et al., 2019) [1]. Liu et al. (2022) [2] utilized BERT to perform sentiment analysis on key audit matters in audit reports and found that textual sentiment information can explain a company's current and future performance, indicating that deep learning methods can effectively extract economic information from financial texts.

In recent years, generative AI technologies such as large language models (LLMs) have driven the further development of natural language processing, with their applications gradually expanding from text analysis to the assessment of corporate economic performance. Pastor-Merino et al. (2025) [3] combined web crawling with LLMs to identify AI adoption among Spanish firms, finding that firms adopting AI had higher sales and value added, and that AI adoption was significantly and positively correlated with firms' annual sales growth rates. Relevant studies indicate that natural language processing (NLP) technology provide a new research perspective for identifying and evaluating the economic effects of AI applications.

2.2. Natural Language Processing Technologies and New-Quality Productive Forces in Enterprises

Existing research on the economic effects of natural language processing technology is relatively limited, with relevant empirical evidence primarily drawn from the fields of digital transformation and artificial intelligence applications. Ma and Zhang (2026) [4] found that improvements in a company's level of digitalization help promote the development of new-quality productive forces; Rong and Wu (2026) [5] argue that the application of artificial intelligence can enhance a company's new-quality productive forces by optimizing resource allocation and strengthening internal controls; Yan and Sun (2026) [6] point out at the industry level that the large language model (LLM) industry can drive the development of new-quality productive forces through pathways such as optimizing capital allocation. Existing research provides important insights into the relationship between intelligent technologies and new-quality productive forces; however, most studies focus on broad digitalization or AI applications, while attention to the economic effects of natural language processing (NLP) technology itself remains relatively limited. Based on this, this paper takes NLP as the core explanatory variable to systematically examine its impact on firms' new-quality productive forces and its underlying mechanisms.

3. Research Design

3.1. Sample Selection and Data Sources

This paper selects Chinese A-share listed companies from 2015 to 2023 as the research sample and applies the following screening criteria: (1) Exclude financial listed companies; (2) Exclude ST, *ST, and delisted companies; (3) Exclude samples with missing key variables. To eliminate the impact of outliers, continuous variables were trimmed at the top and bottom 1%, resulting in a final sample of 11,815 observations. Data for natural language processing (NLP) were sourced from text extracted from listed companies' annual reports, while data for the remaining variables were obtained from the CSMAR database.

3.2. Definition of Main Variables

3.2.1. Dependent Variable

New-Quality Productivity (Nqp) in Enterprises: Drawing on the research by Zhang Xiu'e et al. (2025) [7], this paper constructs an evaluation indicator system based on three dimensions—new-quality laborers, new-quality objects of labor, and new-quality means of labor. Specifically, the entropy method is used to determine the weights of indicators at each level, and a comprehensive score is employed to measure the level of new-quality productivity in enterprises. Specifically, "New-Quality Workers" are measured by the proportion of highly qualified employees, the proportion of R&D personnel, the digital background of management, and the breadth of the CEO's functional experience; "New-Quality Objects of Labor" are measured by environmental performance, the proportion of fixed assets, and robot penetration rate; and "New-Quality Means of Labor" are measured by the enterprise's innovation level, green technology level, proportion of green patents, level of intelligence, and proportion of digital assets.

3.2.2. Independent Variable

Applications of Natural Language Processing (NLP) Technologies: Drawing on the approach proposed by Wu Fei et al. (2021) [8], this study extracts keywords from the "Management Discussion and Analysis" section of listed companies' annual reports, counts the frequency of NLP-related keywords, adds 1 to each frequency, and calculates the natural logarithm of the result to serve as a proxy variable for the extent of a company's NLP technology adoption. These keywords primarily cover foundational technologies (such as natural language processing, large language models, pre-trained models, multimodal large models, Transformer models, and neural language models), model and application names (ChatGPT, Claude, Gemini, Llama, Wenxin Yiyan, Tongyi Qianwen, Doubao, iFlytek Xinghuo, KIMI, DeepSeek, Zhipu Qingyan, Tencent Hunyuan, Yun Pangu, Mistral-Large-Instruct, etc.), and technical capabilities (semantic understanding, contextual reasoning, sample learning, prompt engineering, model fine-tuning, machine learning, deep learning, etc.).

3.3. Control Variables

In accordance with academic conventions, the control variables selected include: firm size (Size), debt-to-equity ratio (Lev), board size (Board), cash flow level (CashFlow), firm age (FirmAge), dual role (Dual), and largest shareholder's stake (Top1). The definitions of the main variables are shown in Table 1.

Table 1. Variable Definition

Variable Type	Variable Name	Variable Symbol	Variable Definition
Dependent Variable	New-Quality Productive Forces in Enterprises	Nqp	Comprehensive Measurement Using the Entropy Method
Independent Variable	Applications of Natural Language Processing Technology	NLP	Natural logarithm of the keyword frequency in the annual report text plus 1
Control Variables	Enterprise Size	Size	Natural logarithm of total assets
	Financial Leverage	Lev	Total liabilities / Total assets
	Board Size	Board	Natural logarithm of the number of board members
	Cash Flow Levels	CashFlow	Net cash flow from operating activities / Total assets
	Age of the Enterprise	FirmAge	Natural logarithm of the number of years listed
	Combined Roles	Dual	Whether the Chairman and General Manager hold both positions
	Largest Shareholder's Stake	Top1	Shareholding percentage of the largest shareholder

3.4. Model Setting

To examine the impact of natural language processing technology on enterprises' new-quality productive forces and the underlying mechanisms, a multiple linear regression model was employed, controlling for fixed effects of year (Year) and industry (Industry), with clustering adjustments made at the industry level for listed companies during the regression analysis.

4. Empirical Analysis

4.1. Descriptive Statistical Analysis

Table 2 presents the descriptive statistics for the main variables. The mean value of enterprises' new-quality productivity (Nqp) was 1.480, with a standard deviation of 0.720, a minimum value of 0.470, and a maximum value of 3.400, indicating that there are certain differences in the levels of new-quality productivity among the sample enterprises. The mean for natural language processing (NLP) technology was 0.150, with a standard deviation of 0.460, suggesting that some enterprises have already adopted NLP technology to a certain extent. The distribution characteristics of the remaining control variables were generally consistent with those reported in previous studies.

Table 2. Descriptive Statistics

Variable	N	Mean	SD	P50	Min	Max
Nqp	11815	1.480	0.720	1.310	0.470	3.400
NLP	11815	0.150	0.460	0.000	0.000	2.560
Size	11815	22.20	1.090	22.060	20.190	25.220
Lev	11815	0.390	0.180	0.380	0.060	0.850
Board	11815	2.100	0.190	2.200	1.610	2.560
CashFlow	11815	0.050	0.060	0.050	-0.120	0.240
FirmAge	11815	3.000	0.280	3.000	2.300	3.610
Dual	11815	0.330	0.470	0.000	0.000	1.000
Top1	11815	0.320	0.140	0.290	0.080	0.700

4.2. Regression Analysis

Table 3. Baseline Regression Results

	(1)	(2)	(3)
	Nqp	Nqp	Nqp
NLP	0.232***	0.209***	0.126***
	(16.263)	(14.723)	(11.398)
Size		0.092***	0.119***
		(11.897)	(7.065)
Lev		-0.012	-0.021
		(-0.282)	(-0.274)
Board		0.045	0.126**
		(1.278)	(2.103)
FirmAge		-0.195***	-0.117***
		(-18.233)	(-7.376)
CashFlow		-0.576***	-0.105
		(-5.477)	(-0.795)
Dual		-0.008	-0.007
		(-0.594)	(-0.368)
Top1		-0.451***	-0.105*
		(-9.542)	(-1.723)
_cons	1.442***	-0.103	-1.140***
	(209.954)	(-0.644)	(-3.077)
N	11815	11815	11815
ar2	0.022	0.056	0.288
Year	No	No	Yes
Industry	No	No	Yes

Note: *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively; the values in parentheses are two-tailed t-values adjusted for standard robustness at the cluster industry level; the same applies below.

Table 3 reports the results of the baseline regression. In Column (1), which includes only the core explanatory variables, the estimated coefficient for NLP is 0.232, which is significant at the 1% level. In Column (2), after controlling for covariates, the coefficient for NLP is 0.209, which remains significant at the 1% level. Column (3) further controls for industry and year fixed effects; the coefficient for NLP is 0.126, which is significant at the 1% level. The above results indicate that natural language processing technology plays a significant role in promoting enterprises' new-quality productive forces.

4.3. Robustness Tests

To ensure the reliability of the benchmark regression results, this paper conducts robustness tests in three aspects: changing the sample period, excluding extreme events, and excluding samples from specific regions. Column (1) of Table 4 considers the impact of the COVID-19 pandemic by shortening the sample period to 2015–2020; the estimated coefficient for NLP is 0.164, which is significant at the 1% level. Taking into account abnormal volatility in the A-share market, escalating U.S.-China trade friction, and major public health events, columns (2) through (4) exclude the 2015, 2018, and 2020 samples, respectively. The estimated coefficients for NLP are 0.131, 0.126, and 0.115, respectively, all of which are significant at least at the 1% level, indicating that the baseline results are not driven by extreme events in specific years. Column (5) excludes samples from the four first-tier cities of Beijing, Shanghai, Guangzhou, and Shenzhen, while Column (6) excludes samples from all municipalities directly under the central government. The NLP estimated coefficients are 0.139 and 0.143, respectively, and remain significant at the 1% level. The above tests indicate that the core conclusions remain robust after adjusting the sample range and excluding samples related to extreme events and specific regions.

Table 4. Robustness Tests Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Nqp	Nqp	Nqp	Nqp	Nqp	Nqp
NLP	0.164***	0.131***	0.126***	0.115***	0.139***	0.143***
	(8.274)	(11.361)	(9.796)	(10.516)	(8.309)	(7.540)
Size	0.112***	0.120***	0.115***	0.119***	0.104***	0.112***
	(6.340)	(7.197)	(6.790)	(7.190)	(5.511)	(6.214)
Lev	0.036	-0.027	-0.013	-0.047	0.046	0.004
	(0.421)	(-0.356)	(-0.171)	(-0.576)	(0.519)	(0.054)
Board	0.110	0.130**	0.128**	0.131**	0.103*	0.139***
	(1.410)	(2.170)	(2.185)	(2.280)	(1.794)	(2.697)
FirmAge	-0.098***	-0.116***	-0.124***	-0.115***	-0.107***	-0.099***
	(-5.031)	(-6.988)	(-7.488)	(-7.635)	(-6.364)	(-6.124)
CashFlow	-0.005	-0.118	-0.107	-0.172	-0.053	-0.036
	(-0.030)	(-0.847)	(-0.842)	(-1.299)	(-0.328)	(-0.336)
Dual	-0.003	0.000	-0.012	-0.012	-0.021	-0.007
	(-0.134)	(0.012)	(-0.623)	(-0.597)	(-0.793)	(-0.303)
Top1	-0.093	-0.118*	-0.103*	-0.090	-0.034	-0.090
	(-1.149)	(-1.821)	(-1.781)	(-1.501)	(-0.502)	(-1.535)
_cons	-0.992**	-1.177***	-1.043***	-1.178***	-0.893**	-1.114***
	(-2.576)	(-3.226)	(-2.811)	(-3.163)	(-2.396)	(-2.939)
N	6943	10955	10319	10107	8537	9764
ar2	0.278	0.289	0.290	0.290	0.271	0.280
Year	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes

4.4. Mechanism Testing

Natural language processing technology can transform vast amounts of unstructured text into structured data assets, enhancing the accumulation of data elements within enterprises and, in turn, empowering new-quality productive forces by optimizing decision-making quality and resource allocation efficiency. To this end, drawing on the research by Saunders and Tambe (2013) [9], this paper uses the natural logarithm of the frequency of data-element-related keywords in listed companies' annual reports plus one (lndata) to measure a firm's level of data elements. The results in Table 5, Column (1), show that the estimated coefficient for NLP on lndata is 0.117, which is significant at the 1% level, indicating that the application of natural language processing technology significantly enhances a firm's level of data element accumulation. The results in Column (2) show that the estimated coefficient for lndata is 0.192, which is significant at the 1% level; furthermore, the estimated coefficient for NLP decreases from 0.126 in the baseline regression to 0.104, indicating that natural language processing technology drives the enhancement of new-quality productive forces by stimulating the allocation of data elements.

Natural language processing technology may also promote the enhancement of new-quality productive forces by strengthening innovation-driven development. Following Adhikari et al. (2016) [10], who measured innovation-driven growth in terms of innovation input intensity, we use the ratio of R&D expenditure to operating revenue (RD) as an indicator. The results in Column (3) show that the estimated coefficient for NLP on RD is 0.014, which is significant at the 1% level, indicating that the application of natural language processing technology has significantly promoted corporate R&D investment. The results in Column (4) show that the estimated coefficient for RD is 2.532, which is significant at the 1% level, and that the estimated coefficient for NLP has decreased from 0.126 in the baseline regression to 0.090. This indicates that innovation-driven growth is a key transmission mechanism through which natural language processing technology empowers new-quality productive forces.

Table 5. Mechanism Test Results

	(1)	(2)	(3)	(4)
	lndata	Nqp	RD	Nqp
NLP	0.117***	0.104***	0.014***	0.090***
	(3.975)	(8.873)	(6.032)	(6.381)
lndata		0.192***		
		(5.010)		
RD				2.532***
				(5.002)
Size	0.036***	0.112***	0.001	0.115***
	(4.350)	(7.079)	(1.222)	(7.067)
Lev	0.002	-0.021	-0.057***	0.123*
	(0.053)	(-0.278)	(-4.390)	(1.849)
Board	0.025	0.122**	0.003	0.118**
	(1.118)	(2.077)	(0.926)	(2.133)
CashFlow	-0.025***	-0.113***	-0.007***	-0.100***
	(-3.140)	(-7.327)	(-3.279)	(-7.242)
FirmAge	-0.311***	-0.045	-0.081***	0.101
	(-2.866)	(-0.376)	(-3.293)	(0.783)
Dual	0.024***	-0.012	0.004*	-0.016
	(3.151)	(-0.634)	(1.798)	(-0.831)
Top1	-0.159***	-0.075	-0.028***	-0.036
	(-4.492)	(-1.231)	(-5.073)	(-0.660)
_cons	3.508***	-1.814***	0.069***	-1.314***
	(20.471)	(-4.134)	(3.461)	(-3.567)
N	11815	11815	11815	11815
ar2	0.515	0.295	0.399	0.312
Year	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes

4.5. Heterogeneity Test

Differences in factor intensity across different industry types may affect the effectiveness of emerging technology adoption. Following the approach of Cheng Pengfei et al. (2025) [11], the sample firms were classified into three categories: labor-intensive, technology-intensive, and capital-intensive. The results of the group-specific regression are presented in Table 6, columns (1)– (3). The NLP-estimated coefficients are significantly positive only in columns (2) and (3), indicating that natural language processing technology has a more pronounced positive effect on technology-intensive and capital-intensive firms, while its impact on labor-intensive firms is relatively limited. A possible reason is that the former two categories of firms typically possess strong technology absorption capabilities or ample financial support, whereas the latter rely primarily on standardized manual operations, where the application scenarios for technology and its role in promoting productivity gains are limited.

Furthermore, following the digital transformation index developed by Wu Fei et al. (2021) [8], the sample was divided into high- and low-transformation groups based on the median. The regression results are presented in Table 6, columns (4)–(5). The results show that the NLP coefficient is significantly positive in the high-transformation group, whereas it fails the significance test in the low-transformation group, indicating that enterprises with a stronger digital foundation are better able to effectively leverage the enabling effects of natural language processing technology.

Table 6. Results of Heterogeneity Test

	(1)	(2)	(3)	(4)	(5)
	Labor-Intensive Group	Technology-Intensive Group	Capital-Intensive Group	High-Digitalization Group	Low-Digitalization Group
	Nqp	Nqp	Nqp	Nqp	Nqp
NLP	0.032	0.140***	0.157***	0.111***	0.086
	(0.688)	(11.641)	(2.861)	(6.832)	(1.645)
Size	0.093***	0.139***	0.098***	0.128***	0.090***
	(3.443)	(5.722)	(4.461)	(5.957)	(4.498)
Lev	-0.051	0.040	-0.123*	-0.048	0.021
	(-0.284)	(0.329)	(-2.008)	(-0.436)	(0.290)
Board	0.063	0.184**	0.061	0.144	0.147**
	(0.451)	(2.127)	(0.939)	(1.663)	(2.256)
CashFlow	-0.045**	-0.159***	-0.079***	-0.122***	-0.112***
	(-2.217)	(-8.314)	(-3.716)	(-5.355)	(-6.425)
FirmAge	0.368**	-0.282	-0.101	-0.028	-0.162
	(2.414)	(-1.379)	(-0.605)	(-0.136)	(-1.015)
Dual	0.003	0.015	-0.079**	0.000	-0.030
	(0.066)	(0.599)	(-2.800)	(0.016)	(-1.040)
Top1	-0.081	-0.151*	0.003	-0.171	-0.037
	(-0.518)	(-1.951)	(0.025)	(-1.344)	(-0.528)
_cons	-0.992*	-1.447**	-0.777*	-1.191**	-0.721*
	(-1.716)	(-2.629)	(-1.962)	(-2.215)	(-1.733)
N	2074	6849	2877	5674	6134
ar2	0.187	0.232	0.191	0.306	0.194
Year	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes

5. Conclusions and Recommendations

Using a sample of Chinese A-share listed companies from 2015 to 2023, this study empirically examines the impact of natural language processing (NLP) technology on firms' new-quality productive forces. The results indicate that the application of NLP technology by firms significantly enhances their new-quality productive forces by stimulating the allocation of data resources and strengthening innovation-driven development. Furthermore, these effects are more pronounced in non-labor-intensive firms with a solid digital foundation.

Based on these findings, the following recommendations are proposed: (1) The government should encourage and support enterprises in adopting natural language processing technologies. It can employ a combination of policy tools, such as fiscal and tax measures, to reduce the cost of adoption for enterprises. At the same time, the government should focus on the precision of policy support, paying particular attention to supporting non-labor-intensive enterprises with a strong digital foundation. (2) Enterprises should recognize the positive effects of digital empowerment on productivity, accelerate the adoption of natural language processing technologies, leverage their role in activating data elements and strengthening innovation-driven development, and unleash the vitality of new-quality productivity.

Acknowledgements

This study was supported by the 2024 General Project of Chongqing Language and Script Research Program "Research on the Impact of Natural Language Intelligent Processing Technology on Enterprise New Quality Productive Forces" (yyk24228), and the 2023 Youth Project of Chongqing Educational Science Planning Program

"Research on the Cultivation System of Top-notch Accounting Talents in Finance and Economics Colleges from the Perspective of Digitalization" (K23YY2080011).

References

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT (pp. 4171–4186).

[2] Liu, W. P., Yen, M. F., & Wu, T. Y. (2022). Report users' perceived sentiments of key audit matters and firm performance: Evidence from a deep learning-based natural language processing approach. Journal of Information Systems, 36(3), 191–209.

[3] Pastor-Merino, A., Martínez-Barbero, X., Vicente, M. R., & Domenech, J. (2025). Does AI boost firm productivity? A web scraping and LLMs approach. Telecommunications Policy, 50(2), 103138.

[4] Ma, Y. Z., & Zhang, H. L. (2026). Enterprise digitalization and new quality productive forces: A textual analysis. Finance Research Letters, 106, 110272.

[5] Rong, Y., & Wu, H. (2026). Artificial intelligence applications and enterprise new quality productivity: Empirical evidence from China's high-tech industry. Finance Research Letters, 97, 109777.

[6] Yan, Z., & Sun, L. (2026). Research on the efficacy and mechanisms of the large language models industry in driving the development of new quality productivity. Journal of Industrial Technological Economics, 45(4), 133–145.

[7] Zhang, X., Wang, W., & Yu, Y. B. (2025). Research on the influence of digital intelligence transformation on the new quality productivity of enterprises. Studies in Science of Science, 43(5), 943. (In Chinese)

[8] Wu, F., Hu, H. Z., Lin, H. Y., & Ren, X. Y. (2021). Enterprise digital transformation and capital market performance: Empirical evidence from stock liquidity. Journal of Management World, 37(7), 130–144. (In Chinese)

[9] Saunders, A., & Tambe, P. (2013). A measure of firms' information practices based on textual analysis of 10-K filings. Working paper.

[10] Adhikari, B. K., & Agrawal, A. (2016). Religion, gambling attitudes and corporate innovation. Journal of Corporate Finance, 37, 229–248.

[11] Cheng, P. F., Li, S. L., Xie, F. T., & Liu, M. (2025). Mechanism and empirical study of how the digital economy empowers the optimization and upgrading of the manufacturing industry structure. Systems Engineering, 43(5), 56–69. (In Chinese)