Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O):3005-7507
ISSN(P):3005-7493
Vol. 2 No.4 (2025)

Public Data Access and AI Adoption for Sustainable Digital Transformation: Evidence from China

Wei Zhao[1,2]*

1.IEBIS, Department of High-Tech Business and Entrepreneurship, Faculty of Behavioral, Management and Social Sciences, University of Twente, 7522 NB Enschede, The Netherlands

2.Faculty of Business Administration, Turiba University, LV-1058 Riga, Latvia

Corresponding author: Wei Zhao, zhaowei11208@gmail.com

Abstract: The correlation between public data accessibility and the adoption intensity of artificial intelligence (AI) in Chinese enterprises remains systematically understudied. By leveraging the rollout of municipal public data platforms as a quasi-natural experiment, this study demonstrates that enterprises in cities with such platforms exhibit significantly stronger AI adoption than those in non-platform regions. Mechanistically, this effect operates through dual pathways: significant reductions in operational expenditures and structural upgrades in specialized AI workforce allocation. This study elucidates the action pathway of China's data platform opening up in facilitating the application of artificial intelligence within enterprises. Furthermore, it offers a universal analytical framework for examining the coupling mechanism between data elements and industrial digital transformation during the technological transition in developing countries. These findings also suggest that improving public data accessibility contributes to sustainable digital transformation by aligning technological diffusion with efficient and inclusive resource utilization.

Keywords: Public Data Access; Artificial Intelligence; Information-Acquisition Costs; Talent Structure; Sustainability

Published: Nov 13, 2025

DOI: https://doi.org/10.62177/jaet.v2i4.848

1.Introduction

As a disruptive technology at the forefront of a new wave of technological revolution, artificial intelligence is a crucial driver of economic growth. According to the Chinese State Council's "Notice on the Issuance of the New Generation Artificial Intelligence Development Plan," it is projected that by 2030, the scale of China's core industries related to AI will exceed 10 trillion yuan, with related industries surpassing 100 trillion yuan. Existing literature has also shown that AI technologies contribute to enhancing enterprise innovation capabilities and product quality (Xu and Tian, 2023; Babina et al., 2024). However, Accenture's "2024 Digital Transformation Index Report for Chinese Enterprises" indicates that most companies in China are still taking a wait-and-see approach regarding AI technologies. Therefore, how to enhance the level of AI technology adoption among Chinese enterprises is a crucial issue.

In addition to driving economic growth, AI adoption also plays a pivotal role in achieving sustainable digital transformation, where technological innovation serves environmental, social, and governance objectives. Recent research has shown that data accessibility and digital intelligence can strengthen the sustainability of industrial systems by improving energy efficiency

and promoting green innovation(Zhao, Chen, & Bulis, 2025). Likewise, the integration of AI and Industry 4.0 technologies supports a sustainable society across sectors by optimizing resource use and reducing operational waste(Zhao, Chen, Yazan, et al., 2025). Hence, understanding the mechanisms through which public data access fosters AI adoption provides not only an economic but also a sustainability-oriented perspective on how data infrastructure enables responsible innovation in developing economies.

The technology adoption theory posits that the adoption of new technologies is influenced by various factors, with technology adoption costs and human resources being particularly significant (Straub, 2009; Rojas-Berrio et al., 2022). The application of AI technology highly relies on the collection and utilization of information; however, the high transformation costs associated with gathering quality data and the demand for skilled personnel significantly dampen enterprises' enthusiasm for adopting artificial intelligence technologies. Fortunately, the public data access offers a potential solution to this challenge.

From the standpoint of sustainable digital governance, the establishment of public data platforms can be regarded as a long-term institutional innovation that enhances both technological inclusiveness and social welfare. By lowering information-acquisition costs, these platforms contribute to equitable technology diffusion and inclusive growth—two cornerstones of the United Nations' Sustainable Development Goals (SDGs 9 and 16).

Public data platform accessibility facilitates enterprise adoption of artificial intelligence (AI) technologies via two primary pathways. First, public platforms reduce operational costs by overcoming institutional barriers to data circulation through their authoritative governance and cross-domain interoperability. Compared with the monopolistic nature of proprietary data and privacy limitations of individual data, standardized public datasets with high credibility lower enterprises' comprehensive costs—including data procurement, preprocessing, and regulatory compliance. These efficiency gains in technology assimilation strengthen enterprises' economic incentives for AI adoption (Farboodi & Veldkamp, 2021; Du et al., 2024). Second, platform-driven technological spillovers promote human capital restructuring by attracting specialized talent clusters—a phenomenon particularly evident in AI-intensive industries (Zhao & Fan, 2021). To maximize public data utility, enterprises optimize talent deployment by constructing skill hierarchies that integrate AI research capabilities with domain-specific implementation expertise. This restructuring elevates the proportion of technical positions and fosters cross-disciplinary competency networks, creating a self-reinforcing cycle for AI diffusion.

As previously noted, while data accessibility theoretically enhances enterprise AI adoption, most developing economies struggle to translate public data into productive resources due to inadequate data governance capabilities and institutional constraints. Weak institutional willingness to share data and limited technical capacity exacerbate cost barriers for technology adoption (Du et al., 2024). In this context, following the U.S. launch of the world's first national public data platform in 2009, countries like the UK, Singapore, and Russia have implemented similar government data portals. Among developing nations, China pioneered systematic efforts in public data accessibility, initiating its Shanghai Municipal Data Service Platform in 2012 and subsequently establishing localized public data infrastructures. By 2021, China had operationalized 173 municipal-level platforms, creating a robust quasi-experimental setting for empirically identifying causal relationships between data accessibility and enterprise AI adoption.

Although data accessibility can enhance enterprise AI adoption, developing economies face challenges in operationalizing public data due to insufficient governance frameworks and structural barriers. Institutional reluctance toward data sharing and technical capacity gaps intensifies preprocessing and compliance costs (Du et al., 2024). Since the United States inaugurated the first national public data platform in 2009, industrialized nations including the UK and Singapore have established comparable national platforms. As a developing economy pioneer, China launched the Shanghai Municipal Data Service Platform in 2012, followed by 173 municipal-level public data infrastructures nationwide. This multi-tiered infrastructure creates natural experiment conditions through staggered implementation timelines, enabling rigorous analysis of data accessibility's causal impact on AI adoption via difference-in-differences methodology.

Using the establishment of public data platforms as a quasi-natural experiment, our research findings indicate that companies located in cities with public data platforms exhibit a higher degree of AI technology application compared to those in cities without public data platforms. This effect operates through dual pathways: significant reductions in operational expenditures

and structural upgrades in specialized AI workforce allocation.

The contributions of our research are as follows: First, we enhance the understanding of the economic consequences of public data access. The academic community generally agrees that public data access can create significant socio-economic value (Zhao and Fank, 2021); It also can help overcome managerial limitations and optimize the external information landscape (Farboodi and Veldkamp, 2021; Du et al., 2024). Our research further supplements the literature on the economic implications of public data access from the perspective of AI technology application. Second, our study has practical significance. Historically, governments have often relied on high-tech subsidies to encourage enterprises to adopt new technologies. However, such measures can lead to issues like the emergence of "zombie" companies (Chang et al., 2021). Comparing with the subsidy strategies, the establishment of public data platforms reduces intermediate costs for enterprises, promotes technological advancement while mitigating resource misallocation issues. This practical value is also beneficial for other developing countries.

2.Research design

2.1 Sample selection

The sample period of our study spans from 2010 to 2022. Our study manually collected information on cities with public data access. Annual report data was sourced from the Giant Tide Information Network, while other financial indicators were obtained from the CSMAR databases. Furthermore, we excluded ST companies, those in the financial sector, and firms with missing or abnormal data, applying a 1% Winsorization to the dataset.

2.2 Variable definition

2.2.1 Dependent variable

Artificial Intelligence (AI): Drawing on the AI dimension dictionary provided by Chen and Srinivasan (2024), our study further integrates insights from multiple industry research reports and the AI vocabulary supplied by the World Intellectual Property Organization (WIPO) to identify 52 seed terms, including "artificial intelligence," "machine learning," and "internet of things." Utilizing Word2Vec technology, we employed the Skip-gram model to train on the vocabulary extracted from annual reports and patent texts. After manually removing duplicate keywords generated by the machine learning process, we finalized the AI dictionary. We then calculated the natural logarithm of the number of AI-related keywords plus one from the MD&A sections to serve as our indicator of corporate AI adoption.

2.2.2 Independent variable

Government Data Openness (Open): Following Du et al. (2024), we assign a value of 1 to Open if the location of the company has a public data platform, 0 otherwise.

2.2.3 Control variables

Our study employs common firm-level characteristics as control variables; Table 1 provides detailed definitions.

Table 1 Definitions of the main variables

Variables	**Definition**
Open	An indicator variable equals one if the public data platform opens a year before in the city where the firm is headquartered and zero otherwise.
AI	The natural logarithm of the sum of artificial intelligence terms mentions in the MD&A plus one.
Size	The natural logarithm of the total assets.
Lev	Total liabilities/Total assets.
Roa	Net profit/Total assets.
ListAge	The natural logarithm of one plus listing age.
Cashflow	Net cash flow from operating activities/Ending current liabilities
Growth	Increase in current year revenue/Previous year total revenue.

Variables	Definition
Top1	The share of top 1 shareholders.
Board	The natural logarithm of the board count.
Dual	Whether the chairman and general manager concurrently serve.

2.3 Model construction

To examine the impact of public data access on the application degree of AI technology, we employ a difference-in-differences (DID) regression model:

$$AI_{i,t} = \alpha_1 + \alpha_2 Open_{i,t} + \alpha_3 Controls_{i,t} + \delta_i + \eta_t + \varepsilon_{i,t} \quad (1)$$

Where i represents a firm and t indicates the year. The dependent variable, AI, measures the application degree of artificial intelligence technology. The independent variable, Open, is an indicator that equals 1 if the the public data platform has been implemented in the city where firm i is in year t, and 0 otherwise. Additionally, our study controlled for firm fixed effects (δi) and year fixed effects (ηt). The sample is clustered at the firm level.

3.Empirical methodology

From a policy perspective, public data infrastructures can also be interpreted as catalysts of sustainable digital governance. By reducing informational asymmetries and fostering transparent, data-driven decision-making, these infrastructures contribute to responsible innovation and long-term socio-economic resilience.

3.1 Summary statistics

Table 2 presents the descriptive statistics of the main variables. The mean of AI is 0.062 with a standard deviation of 0.185. This indicates that the level of application of AI among Chinese listed companies is relatively low, which indirectly highlights the necessity of our research.

Table 2 Summary statistics

Variable	N	Mean	SD	Min	p50	Max
AI	24,365	0.062	0.185	0.000	0.000	1.000
Open	24,365	0.364	0.481	0.000	0.000	1.000
Size	24,365	22.350	1.318	19.960	22.180	26.110
Lev	24,365	0.437	0.209	0.053	0.435	0.889
Roa	24,365	0.036	0.063	-0.219	0.035	0.207
ListAge	24,365	2.393	0.761	0.000	2.565	3.367
Cashflow	24,365	0.044	0.068	-0.149	0.042	0.230
Growth	24,365	0.163	0.367	-0.511	0.106	1.916
Top1	24,365	0.333	0.147	0.088	0.308	0.725
Board	24,365	2.135	0.197	1.609	2.197	2.639
Dual	24,365	0.242	0.429	0.000	0.000	1.000

3.2 Baseline regression

Table 3 presents the baseline regression results. As shown in the table, the coefficient of Open are all significantly positive at the 1% level, regardless of whether control variables and fixed effects are included. This indicates that the establishment of public data platforms significantly enhances the application of AI technologies by enterprises.

Table 3. Baseline results.

	(1)	(2)	(3)
		AI	
Open	0.019***	0.078***	0.018***
	(3.86)	(15.92)	(3.61)
Size		0.016***	0.034***
		(7.81)	(7.90)
Lev		-0.097***	-0.015
		(-7.43)	(-0.90)
ROA		-0.178***	-0.124***
		(-4.90)	(-3.81)
ListAge		0.008***	0.035***
		(3.34)	(5.29)
Cashflow		-0.065***	-0.044**
		(-3.07)	(-2.43)
Growth		-0.001	-0.012***
		(-0.44)	(-3.76)
Top1		-0.136***	-0.091***
		(-8.80)	(-3.46)
Board		-0.034***	0.019
		(-3.17)	(1.34)
Dual		0.019***	0.004
		(3.83)	(0.88)
Constant	0.055***	-0.188***	-0.788***
	(30.39)	(-4.51)	(-8.21)

3.3 Mechanism test

3.3.1 Reduce costs

Cost is a significant factor that influences whether firms adopt AI technologies (Straub, 2009). To assess whether the establishment of public data platforms contributes to reducing corporate costs and thereby encourages the adoption of AI technologies, we selected two closely related indicators: (1) Management Expense Ratio (Mfee), defined as the ratio of a company's management expenses to its revenue from core operations; (2) Operating Cost Ratio (OpCostR), defined as the ratio of a company's operating costs to its revenue from core operations. The relevant regression results are presented in Table 4 columns (1)-(2), where the coefficient of Open are all significantly negative. This suggests that the public data access contributes to lowering corporate costs, thereby facilitating the adoption of AI technologies by enterprises.Beyond cost efficiency, these reductions may indirectly promote corporate sustainability by minimizing redundant resource use and energy consumption. As firms leverage open data for process optimization, their digital transformation also aligns more closely with sustainable production and environmental responsibility.

3.3.2 Optimize the talent structure

Talent is also a crucial factor that limits the application of new technologies by enterprises (Rojas-Berrio et al., 2022). To

evaluate whether the establishment of public data platforms aids in optimizing the talent structure of companies, we used the proportion of employees with postgraduate degrees as the measuring indicator (Gra). The relevant regression results are presented in Table 4 column (3), where the coefficient of Open is significantly positive. This indicates that the public data access facilitates the aggregation effect of talent, thereby promoting the application of AI technologies through the optimization of the company's talent structure.

Table 4. Mechanism test.

	(1)	(2)	(3)
	Mfee	**OpcostR**	**Gra**
Open	-0.003**	-0.005*	0.336***
	(-2.12)	(-1.78)	(4.17)
Size	-0.013***	-0.020***	0.733***
	(-8.91)	(-7.14)	(7.19)
Lev	-0.033***	0.107***	-1.095***
	(-5.70)	(8.93)	(-2.77)
ROA	-0.189***	-1.478***	-0.844
	(-17.16)	(-45.21)	(-1.48)
ListAge	-0.006***	0.015***	-0.369***
	(-3.22)	(4.22)	(-2.79)
Cashflow	-0.024***	-0.102***	-0.505
	(-3.47)	(-6.01)	(-1.30)
Growth	-0.019***	-0.032***	0.054
	(-16.29)	(-11.19)	(0.74)
Top1	-0.005	-0.029*	0.406
	(-0.67)	(-1.82)	(0.60)
Board	0.005	0.001	0.220
	(1.31)	(0.13)	(0.84)
Dual	-0.001	0.001	-0.055
	(-0.35)	(0.37)	(-0.64)
Constant	0.405***	1.380***	-12.594***
	(12.35)	(21.88)	(-5.68)
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	24,365	24,365	24,365
adj. R2	0.716	0.677	0.750

Notes: (1) Standard errors in parentheses are clustered at the firm level; (2) *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

3.4 Robustness test

To address potential endogeneity issues, our study includes the following tests. First, a key assumption of the difference-in-differences (DID) approach is the parallel trends assumption. Following the methodology of Beck et al. (2010), we conducted a parallel trends test, with the relevant results presented in Figure 1. The estimated coefficients for each period prior to the establishment of the public data platform are not statistically significant, indicating that the research sample satisfies the parallel trends assumption. Second, to mitigate endogeneity issues in the construction of indicators, we constructed a dummy variable (Iiai) as a substitute indicator based on whether the level of AI technology adoption in the firm exceeds the median. Additionally, we broadened the scope of term frequency collection by expanding the dictionary from the MD&A section to encompass the entire annual report (AI_ar). The relevant results are presented in Table 5 columns (1)-(2), and our conclusions remain robust. Third, macroeconomic uncertainty may influence corporate strategy. To avoid any interference from the COVID-19 pandemic on our findings, we limited the sample period to years prior to 2020, with the relevant results shown in Table 5 column (3), again confirming the robustness of our conclusions.

Figure 1-parallel trends test

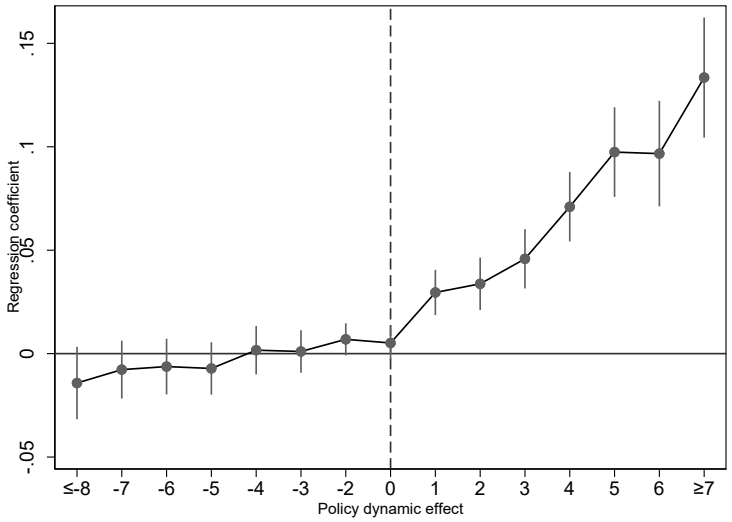

Table 5. Robustness test.

	(1)	(2)	(3)
	Iiai	**AI_ar**	**AI**
Open	0.005**	0.029***	0.019***
	(2.15)	(4.13)	(3.74)
Size	-0.002	0.057***	0.030***
	(-0.83)	(8.31)	(6.40)
Lev	-0.005	-0.012	-0.016
	(-0.39)	(-0.48)	(-0.96)
ROA	0.004	-0.205***	-0.080**
	(0.17)	(-4.51)	(-2.24)
ListAge	0.006	0.053***	0.020***
	(1.36)	(5.42)	(3.51)
Cashflow	-0.027**	-0.063**	-0.027
	(-2.22)	(-2.45)	(-1.55)
Growth	-0.001	-0.014***	-0.007**
	(-0.65)	(-3.32)	(-2.50)
Top1	0.017	-0.137***	-0.058**
	(0.98)	(-3.64)	(-2.35)
Board	-0.007	0.038*	0.028**
	(-0.85)	(1.94)	(2.20)
Dual	0.003	0.008	0.004
	(0.98)	(1.07)	(0.94)
Constant	1.035***	-1.336***	-0.708***
	(16.47)	(-8.64)	(-6.80)
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
N	24,365	24,365	18,185
adj. R2	0.554	0.490	0.334

Notes: (1) Standard errors in parentheses are clustered at the firm level; (2) *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

4.Conclusion

The public data access is a crucial method for realizing the value creation of data as a factor. Our research confirms that access to public data significantly enhances the level of AI technology application, with cost reduction and talent optimization identified as potential mechanisms. This study elucidates the "data-to-technology" transmission mechanism, yielding two critical policy implications for AI diffusion in developing economies: First, institutional innovation should prioritize establishing data ownership frameworks and cross-jurisdictional sharing protocols, positioning public data platforms as core digital infrastructure components—exemplified by phased deployment models. Second, implementation strategies must adapt to regional industrial ecosystems through differentiated policy portfolios. Integrating such open-data frameworks into sustainable industrial policy can therefore serve as a dual lever for digital innovation and environmental responsibility. In this way, data accessibility supports not only economic modernization but also the broader transition toward a more sustainable and inclusive digital economy. For instance, manufacturing hubs should prioritize industrial data interface standardization, while talent-concentrated regions ought to reform skill certification systems to align with AI workforce demands.

Integrating such open-data frameworks into sustainable industrial policy can therefore serve as a dual lever for digital innovation and environmental responsibility. In this way, data accessibility supports not only economic modernization but also the broader transition toward a more sustainable and inclusive digital economy.

Beyond the economic dimension, the findings highlight that expanding open-data ecosystems can enhance long-term sustainability by fostering responsible digital transformation and inclusive growth. Public data infrastructures serve not only as economic catalysts but also as enablers of greener, more resilient innovation systems. By aligning data accessibility with sustainable development objectives, policymakers can ensure that technological progress advances environmental responsibility and social well-being in parallel with industrial upgrading.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

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

[2] Beck, T., Levine, R., & Levkov, A. (2010). Big bad banks? The winners and losers from bank deregulation in the United States. The Journal of Finance, 65(5), 1637-1667.

Xu, X., & Tian, C. (2023). Does artificial intelligence improve the quality of export products? Evidence from China. Applied Economics Letters, 1-5.

[3] Chang, Q., Zhou, Y., Liu, G., Wang, D., & Zhang, X. (2021). How does government intervention affect the formation of zombie firms? Economic Modelling, 94, 768-779.

[4] Chen, W., & Srinivasan, S. (2024). Going digital: Implications for firm value and performance. Review of Accounting Studies, 29(2), 1619-1665.

[5] Du, J., Gao, H., Wen, H., & Ye, Y. (2024). Public data access and stock price synchronicity: Evidence from China. Economic Modelling, 130, 106591.

[6] Farboodi, M., & Veldkamp, L. (2021). A model of the data economy (No. w28427). Cambridge, MA, USA: National Bureau of Economic Research.

[7] Magalhaes, G., & Roseira, C. (2020). Open government data and the private sector: An empirical view on business models and value creation. Government Information Quarterly, 37(3), 101248.

[8] Rojas-Berrio, S., Rincon-Novoa, J., Sánchez-Monrroy, M., Ascúa, R., & Montoya-Restrepo, L. A. (2022). Factors influencing 4.0 technology adoption in manufacturing SMEs in an emerging country. Journal of Small Business

Strategy, 32(3), 67-83.

[9] Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. Review of Educational Research, 79(2), 625-649.

[10] Zhao, W., Chen, H., & Bulis, A. (2025). How are Industry 4.0 technologies transforming a sustainable society across industries? Digital Transformation and Society, 4(3), 363–380. https://doi.org/10.1108/DTS-11-2024-0225

[11] Zhao, W., Chen, H., Yazan, D. M., Taghavifar, H., Lyu, Y., & Bulis, A. (2025). Few-shot learning and deep predictive models for cost optimization and carbon emission reduction in energy-water management. Journal of Environmental Management, 389, 126077. https://doi.org/10.1016/j.jenvman.2025.126077