Proceedings of ICMRED 2026 Symposium: Financial Innovation, Risk Governance, and the Dynamics of Global Capital Flows
DOI: 10.54254/2754-1169/2026.BL34783

Public Data and Digital Technology Innovation — Empirical Evidence from the Establishment of Chinese Government Data Platforms

Yingjie Wang

Shanghai University, Shanghai, China
EasonWang05@outlook.com

Abstract. Against the background of the continuous in-depth development of the digital economy, whether the opening of public data can effectively drive enterprises to carry out digital technology innovation has become an important issue that needs urgent exploration in the cultivation of new-quality productive forces. This paper regards the successive launch of local government data open platforms in China as a quasi-natural experiment, selects the panel data of A-share listed enterprises from 2010 to 2022 as research samples, and uses a multi-period difference-in-differences model to systematically investigate the actual impact, mechanism paths and heterogeneous characteristics of public data opening on corporate digital technology innovation. The empirical results show that public data opening can significantly increase the output scale of enterprises' digital technology patents, and has tangible promoting value for industrial digital transformation. Further analysis at the mechanism level reveals that this positive enabling effect is mainly realized through two paths: first, public data opening can enrich digital application scenarios, expand the scope of IPC classifications covered by enterprises' digital technology patents, and facilitate the deeper integrated development of digital technology and the real economy; second, public data opening can provide authoritative and multi-dimensional macro information, effectively mitigating the uncertain risks faced by enterprises in market operation and policy judgment, and jointly promoting the implementation of corporate digital technology innovation. This paper quantitatively clarifies the enabling value of public data opening for digital technology innovation from the micro-enterprise level, which not only provides empirical evidence for the government to optimize the allocation of data factors and implement differentiated data opening policies in the digital era, but also offers feasible policy ideas and practical references for the cultivation of new-quality productive forces and the advancement of industrial digital transformation.

Keywords: public data opening, digital technology, difference-in-differences method, digital-real integration, digitalization

DOI: 10.54254/2754-1169/2026.BL34783

1. Introduction

Technological innovation is the core support for cultivating new-quality productive forces and the core driving force for China's high-quality economic development. With the in-depth evolution of the digital economy, data has become a core production factor, completely reshaping enterprise innovation models and industrial organizational forms. As the core driving force for industrial digital transformation, digital technology innovation relies heavily on the continuous supply of high-quality data resources for its R&D and implementation. The report to the 20th National Congress of the Communist Party of China clearly proposed accelerating the building of a "Digital China", emphasizing the release of data element value and the strengthening of development momentum for the digital economy. The subsequent *Opinions of the Central Committee of the Communist Party of China and the State Council on Building a Basic Data System to Better Leverage the Role of Data Elements* completed top-level institutional design from the dimensions of data property rights, circulation and trading, and income distribution, with the core goal of improving the efficiency of data resource allocation and actual utilization.

At present, prominent contradictions still exist in China's data resource allocation. A large amount of high-value public data has long been retained within various government departments, with business systems fragmented from each other, forming a closed and scattered pattern of "data silos". High-quality public data is difficult to achieve cross-domain sharing and market-oriented circulation, directly restricting the release of the overall value of data elements [1].

In response to this practical pain point, since 2012, national and local governments have successively launched public data opening pilot projects and made every effort to build exclusive government data opening platforms. According to the *Report on China's Local Government Data Opening* released by the Digital and Mobile Governance Laboratory of Fudan University, in recent years, local governments have continuously integrated data in livelihood and industrial fields such as economy, transportation, education, and environmental protection and opened them to the public. Public data resources have gradually broken the closed and fragmented state and transformed into an interconnected and shared "data continent". The government continues to release public data dividends through open platforms, laying a solid data foundation and application environment for enterprises to carry out digital technology R&D and scenario-based innovation.

Against this background, it is necessary to conduct in-depth exploration: Can public data opening effectively elevate the level of enterprise digital technology innovation? From the perspective of practical value, open and shared public data can expand data sources for enterprises, enrich application scenarios, and provide solid support for digital technology R&D. However, from another perspective, public data has strong sharing and non-exclusive characteristics, which may easily lead to problems such as imitation among market entities and homogeneous competition in the industry, thereby weakening the endogenous motivation of enterprises for long-term innovation R&D. Therefore, clarifying the internal connection between public data opening and enterprise digital technology innovation not only has important theoretical research value but also provides reliable references for policy formulation in practice.

Public data opening is a key link in digital government construction and market-oriented allocation of data elements, and its economic and social effects have attracted extensive academic attention in recent years. Existing research mostly focuses on three major directions: economic efficiency, urban entrepreneurial vitality, and regional coordinated development. From a macro perspective, Peng Yuanhuai [2] conducted an empirical study using a multi-period difference-in-differences model and confirmed that government data opening can significantly improve enterprise

DOI: 10.54254/2754-1169/2026.BL34783

total factor productivity, with core effects reflected in optimizing the information environment and improving resource allocation efficiency. Fang Jincheng et al. [3] further pointed out that public data opening can break information barriers between regions and promote regional coordinated development, and this driving effect is more pronounced in areas with prominent problems of traffic segmentation and administrative segmentation.

At the micro-subject level, the impact of public data opening on entrepreneurial behavior and enterprise innovation is also a hot topic in academic research. Cai Yunkun et al. [4] found that public data opening can lower industry market access thresholds, alleviate market information asymmetry, and significantly activate overall urban entrepreneurial vitality. Meanwhile, research on the definition, quantitative measurement, and influencing factors of digital technology innovation, as the core starting point for industrial digital transformation, has formed a mature analytical framework. In terms of indicator construction and quantitative identification, domestic scholars generally use patent data to represent the level of digital technology innovation. Hu Zengxi and Ma Shuzhong [5], referring to the *Correspondence Table between Digital Economy Core Industry Classification and International Patent Classification* issued by the National Bureau of Statistics, constructed quantitative indicators for enterprise digital innovation by matching enterprise patent IPC classification numbers. Huang Bo et al. [6] also verified the positive boosting effect of digital technology innovation on high-quality enterprise development based on digital technology patent data, providing methodological references for the construction of enterprise digital technology innovation indicators in this paper.

From the perspective of innovation driving mechanisms, existing research mainly unfolds from three dimensions: digital technology application, production factor input, and institutional environment constraints. Tao Feng et al. [7] confirmed that public data opening can alleviate market information asymmetry and provide diversified application scenarios, serving as an important external condition for empowering digital technology innovation. Sufficient R&D capital investment and the reserve of high-quality professional and technical talents are also indispensable basic guarantees for enterprises to implement digital technology innovation.

By sorting out existing literature, it can be found that most existing studies focus on the impact of public data opening on conventional economic behaviors such as enterprise operating performance and traditional R&D innovation, and there is still a lack of systematic and complete discussion on the deep-seated mechanism of its role in cutting-edge strategic fields such as digital technology innovation. Meanwhile, in relevant research on the driving factors of digital technology innovation, academia mostly only regards data elements as background resources, and there are still obvious gaps in research on the micro-logic and internal paths of how government-led public data, as an independent production factor, drives enterprise digital technology innovation. Relevant reports from the World Intellectual Property Organization clearly point out that data is the core raw material for technological innovation in the digital era, and data availability and openness directly determine the iteration direction of digital technology. Based on this, this paper starts from the perspective of data element supply and systematically explores the actual impact of public data opening, an exogenous policy shock, on enterprise digital technology innovation, which can effectively make up for the shortcomings of existing research and has important theoretical supplementary significance.

To address the above research issues, this paper regards the launch of local government data opening platforms as an exogenous policy shock of public data opening, selects panel data of China's A-share listed enterprises from 2010 to 2022, constructs a multi-period difference-in-differences model, and empirically tests the policy effects and transmission mechanisms of public data opening on enterprise digital technology innovation. Meanwhile, combined with enterprise

DOI: 10.54254/2754-1169/2026.BL34783

industry attributes and R&D investment intensity characteristics, it further explores the heterogeneous differences in policy effects. The research results show that public data opening can significantly improve enterprise digital technology innovation output. Compared with enterprises in cities without data opening platforms, the number of digital technology patents of policy-covered enterprises has increased significantly. Mechanism tests show that, on the one hand, public data opening promotes the deep integration of digital technology and the real economy by expanding application scenarios and reducing the cost of cross-border data acquisition, stretching the application boundary of digital technology innovation. On the other hand, public data opening outputs authoritative and multi-dimensional macro information, effectively alleviating market and policy uncertainties faced by enterprises and boosting their willingness to invest in innovation R&D. Extended analysis also finds that this positive promotion effect is more prominent in technology-intensive industries and enterprises with high R&D investment, showing significant enterprise heterogeneity.

The marginal contributions of this paper are mainly reflected in three aspects: First, innovation in research perspective. Existing literature mostly focuses on analyzing the impact of public data opening on conventional economic performance such as enterprise total factor productivity and urban entrepreneurial vitality [2], or exploring its role in enterprise traditional R&D activities [8], with little attention paid to the cutting-edge field of digital technology innovation. This paper directly takes digital technology innovation as the research core, selects enterprise digital technology patent output as the core measurement indicator, accurately identifies the empowering value of public data opening for industrial digital transformation and new-quality productive force cultivation, and fills the gap in research on the connection between data elements and enterprise core technology innovation. Second, innovation in action mechanism. This paper constructs a dual-path analysis framework of "digital-real integration + uncertainty mitigation", dismantling the internal logic of public data opening driving digital technology innovation from the two dimensions of alleviating external information friction and fostering diversified application scenarios. Different from the existing research idea of explaining public data value solely from the perspective of resource allocation optimization, the dual-path framework constructed in this paper can more comprehensively interpret the internal logic of data elements empowering technological innovation and enrich the relevant theoretical system. Third, innovation in research design. This paper accurately matches micro-enterprise panel data with the launch time of local government data opening platforms and uses a multi-period difference-in-differences model to identify policy net effects. This setting can not only control enterprise individual fixed effects and year time fixed effects to weaken the interference of endogeneity problems but also conduct grouped regression based on enterprise characteristics to depict the differentiated boundaries of policy effects. Compared with existing research that mostly uses provincial and city-level macro data, this paper conducts empirical analysis at the micro-enterprise level, which not only improves the robustness of research conclusions but also provides accurate micro-empirical evidence for local governments to implement differentiated public data opening policies.

2. Theoretical analysis and research hypotheses

The opening of public data is a typical exogenous policy shock. Its impact on corporate digital technology innovation is not a simple linear effect, and the effect can only be gradually released through specific transmission paths. Based on this logic, this paper systematically sorts out the internal mechanism through which public data opening influences enterprise digital technology innovation.

DOI: 10.54254/2754-1169/2026.BL34783

From the perspective of the attributes of data elements, public data is collected, sorted and generated by the government in the process of performing public management functions, and is naturally endowed with the attributes of public goods [3]. Different from traditional production factors such as land, capital and labor, the core value of data elements lies not in their physical form, but in the information content they carry. By mining and analyzing effective information in public data, enterprises can effectively reduce uncertain risks in operational decision-making and R&D layout [9]. The opening of public data enables enterprises to obtain massive high-quality data resources at low or even zero cost. Relying on the effective information contained in the data, enterprises can mitigate unknown risks in the process of technological R&D, thus becoming more willing to increase innovation investment and promote the implementation of technological innovation [10].

From the perspective of the characteristics of digital technology innovation itself, as the core force supporting industrial digital transformation, digital technology R&D relies heavily on the supply of high-quality data. First, public data can be directly used as core input materials for enterprise digital technology innovation. After public data is opened, enterprises can directly access available data resources for technological R&D and scenario-based application development, greatly reducing innovation costs and lowering R&D entry barriers. For example, auto companies developing autonomous driving systems rely on government-released data on urban traffic flow, road networks and geographic elevation; drone enterprises optimize flight control algorithms based on officially issued no-fly zone boundaries and meteorological monitoring data [11]. If such cross-domain, high-value data were collected by enterprises themselves, it would not only incur huge financial and time costs but also face legal compliance and data security restrictions [12]. The opening of public data effectively breaks such constraints, helping enterprises quickly convert scarce data resources into R&D inputs, reducing R&D risks and entry barriers, and directly driving the growth of digital technology patent output [13].

Second, the opening of public data can foster brand-new digital application scenarios and force the iterative upgrading of technological innovation. The data opened by the government is not scattered information fragments, but comprehensive social data covering multiple dimensions such as economic operation, urban governance, public services and ecological environment. The centralized opening of such data essentially generates a large number of new data-intensive application scenarios. For instance, the opening of geospatial and public transport data directly gives rise to market demands such as intelligent travel route planning and real-time congestion early warning, thereby promoting the R&D and implementation of digital technologies including navigation algorithms, vehicle dispatching systems and smart parking solutions [14]. The opening of economic census and enterprise registration data provides data support for market trend analysis, supply chain optimization and business site selection modeling, further stimulating innovation demand for business intelligence analysis tools and data visualization platforms. To adapt to the new scenarios and new demands fostered by public data, enterprises take the initiative to carry out targeted R&D and continuously expand the development space of digital technology innovation. Based on the above theoretical deduction, this paper proposes the core research hypothesis:

H1: The opening of public data can significantly promote enterprise digital technology innovation.

After clarifying the basic positive effect of public data opening on enterprise digital technology innovation, it is necessary to further dismantle the specific transmission mechanisms. As an exogenous policy shock, public data opening does not directly determine enterprise innovation output, but releases empowering effects through specific paths. Combining the technical and

DOI: 10.54254/2754-1169/2026.BL34783

economic attributes of data elements with innovation theory, this paper holds that public data opening drives enterprise digital technology innovation mainly through two paths: "promoting the deep integration of digital technology and the real economy" and "alleviating external environmental uncertainty".

Digital technology is a typical general-purpose technology whose value is not limited to technological breakthroughs in a single field, but more reflected in its diffusion and implementation across industries and scenarios. The diversified and scenario-based data resources released by public data opening serve as an important foundation for promoting the deep integration of digital technology and the real economy. On the one hand, public data opening can build diverse application scenarios for digital technology innovation. Government data open platforms bring together data resources in economy, transportation, education, environmental protection and other fields. After opening to the public, they derive a large number of new application scenarios, generate corresponding technological R&D demands, and force enterprises to tackle key digital technology problems. Research by Jiang Jing et al. [9] shows that public data opening can significantly improve the level of digital-real technology integration of enterprises by creating exclusive application scenarios. Typical examples include the development of intelligent dispatching algorithms based on traffic data and the construction of precision agriculture models based on meteorological data. This logic of "scenarios generating demand and demand driving R&D" encourages enterprises to arrange targeted R&D around public data scenarios and continuously broaden the boundaries of digital technology innovation.

On the other hand, public data opening reduces the cost and threshold for enterprises to obtain cross-domain data, encouraging enterprises to carry out cross-industry and cross-domain technological integration. Enterprises can integrate public data with their own business data and industry-specific data, explore new entry points for technological integration, and extend the application coverage of digital technology. Shen Kunrong and Lin Jianwei [1] point out that public data opening can promote more patents to embed digital technology elements and accelerate the penetration and diffusion of digital technology into various traditional industries. Drawing on this idea, this paper uses the number of IPC classification numbers covered by enterprise digital technology patents to quantitatively characterize the level of digital-real integration. In addition, public data is naturally shared, enabling upstream and downstream enterprises in the industrial chain to carry out collaborative R&D and joint scenario exploration, and build an innovation ecosystem centered on digital-real integration. This not only improves the innovation capacity of individual enterprises but also accelerates the digital transformation of the entire industry. In summary, public data opening effectively broadens the coverage of digital technology innovation by expanding application scenarios and promoting cross-border collaboration.

Innovation and R&D activities are highly vulnerable to various uncertainties. When deploying digital technology R&D, enterprises generally face multiple unknown risks such as market demand fluctuations, technological route iteration and policy environment adjustment. By providing massive, authoritative and multi-dimensional public information, public data opening can effectively alleviate the external environmental uncertainty faced by enterprises and boost their willingness to invest in innovation [15]. First, public data opening reduces market information uncertainty. By analyzing government-released data on population statistics, economic operation, consumption trends and industry supervision, enterprises can accurately judge changes in market demand, identify development opportunities and potential risks in advance, formulate more robust R&D plans, and reduce the uncertainty of the market-oriented implementation of innovation achievements. Second, public data opening improves the predictability of the policy environment. The full disclosure of

DOI: 10.54254/2754-1169/2026.BL34783

policy documents, industry development plans, technical standards and other information can reduce ambiguity caused by policy changes, help enterprises accurately align their innovation strategies with macro policy orientations, and avoid the risk of strategic layout deviations caused by policy adjustments.

When overall external uncertainty declines, enterprises will have more stable income expectations for innovation investment, thus becoming more willing to deploy high-risk, high-return digital technology R&D projects. Notably, the mechanism by which public data mitigates uncertainty is distinctly different from that of enterprise-owned data or individual user data. Personal data mainly helps enterprises portray consumer behavioral characteristics and alleviate terminal market demand uncertainty; internal enterprise data is mainly used to optimize R&D and production processes and reduce internal operational risks; while public data focuses on outputting macro information on market operation and public policies, reducing information barriers between enterprises and the external market as well as public institutions, and specifically acting to alleviate external environmental uncertainty. Based on the above analysis, this paper proposes the following research hypotheses:

H2: The opening of public data promotes enterprise digital technology innovation by promoting the deep integration of digital technology and the real economy.

H3: The opening of public data promotes enterprise digital technology innovation by reducing external environmental uncertainty.

3. Research design

3.1. Data sources and sample selection

The research data in this paper are mainly obtained from four authoritative databases and public materials, covering the enterprise micro-level, policy event level, patent innovation level and control variable level, so as to ensure the reliability and comparability of empirical results. The sources, time intervals and sample characteristics of each type of data are specified as follows:

First, enterprise-level data. Basic enterprise information and financial indicators are derived from the CSMAR database, including core variables such as stock codes of listed enterprises, registered cities, industry classifications and annual operating financial data. This database fully covers A-share listed enterprises on the Shanghai and Shenzhen stock exchanges. This paper selects the time interval from 2010 to 2022 to ensure the continuity and integrity of panel data, and the sample size can objectively reflect the overall characteristics of digital transformation of domestic enterprises.

Second, policy-level data. The core policy variables are collected from the annual report of the "Open Data Forest Index" issued by Fudan University, which mainly counts the construction time of public data open platforms in cities across China, with a data span from 2015 to 2021. This paper revises the problems of missing launch time and vague expression of platforms in some cities by checking official websites of local governments and official news announcements one by one, and accurately matches the timing of policy shocks, laying a foundation for subsequent empirical analysis.

Third, patent-level data. Raw patent data are obtained from the patent database of the China National Intellectual Property Administration, with focus on extracting patent application and grant records of listed enterprises in the field of digital technology. In strict accordance with the *Correspondence Table between Digital Economy Core Industry Classification and International Patent Classification (2023)* and *Patent Classification System for Key Digital Technologies (2023)*issued by the National Bureau of Statistics, this paper matches the main and sub IPC

DOI: 10.54254/2754-1169/2026.BL34783

classification numbers of patents to accurately identify digital technology-related patents, with the time interval also set from 2010 to 2022 to maintain a consistent caliber with enterprise data.

To ensure data quality and improve the credibility of empirical conclusions, this paper conducts strict screening and preprocessing on the initial sample. In terms of time dimension, only observation samples from 2014 and subsequent years are retained. This treatment is adopted because the construction of public data open platforms in China entered an accelerated implementation stage after 2014, the national big data strategy was fully promoted, platforms in various regions were launched intensively, and policy effects gradually emerged. Meanwhile, the availability of various statistical data has been greatly improved since 2014, enabling the complete identification of the dynamic impact of policy shocks.

The sample construction takes "enterprise–year" as the basic observation unit, and integrated balanced panel data are formed through three steps: first, many-to-one matching is conducted between basic information of listed enterprises and the launch year of public data platforms in each city, retaining successfully matched samples; second, the matched data are matched with annual enterprise patent data year by year to retain valid observations; third, all enterprise-level control variables are incorporated. For samples without matched patent records, missing patent values are uniformly assigned 0, indicating that the enterprise had no digital technology patent output in that year; a small number of missing values of control variables are supplemented by linear interpolation. At the same time, to avoid extreme values interfering with regression results, all continuous variables are winsorized at the 1% and 99% percentiles. After a series of preprocessing, a high-quality annual panel dataset of listed enterprises from 2014 to 2022 is finally formed, providing data support for subsequent benchmark regression, mechanism testing and heterogeneity analysis.

3.2. Model specification and variable definition

As an exogenous policy shock, the impact of public data opening on enterprise digital technology innovation includes dual paths: direct R&D empowerment and indirect mitigation of information friction. Before empirical analysis, clarifying model specification rules, variable definition standards and quantitative calibers is the key to ensuring the rigor of policy effect identification. Since public data open platforms in different cities of China were not launched at the same time, the traditional two-period difference-in-differences model cannot adapt to the scenario of batch policy shocks. Therefore, this paper selects a multi-period difference-in-differences model, incorporates enterprise fixed effects and year fixed effects, and adopts cluster-robust standard errors at the enterprise level to weaken estimation bias caused by intra-enterprise time series correlation. The benchmark model is specified as follows:

$$\mathrm{Patent}_{it}=\alpha_0+\alpha_1\mathrm{DID}_{it}+\gamma X_{it}+\mu_i+\lambda_t+\epsilon_{it} \tag{1}$$

Dependent variable: Enterprise digital technology innovation level (Patent). This paper adopts the annual total number of granted digital technology patents of enterprises as the measurement indicator. Compared with patent applications, granted patents have undergone substantive examination, which can more truly reflect the quality of innovation achievements and the actual R&D capabilities of enterprises, and can objectively characterize the level of enterprise digital technology innovation. The specific calculation method of the variable is: screen digital technology-related patents according to IPC classification rules, summarize the number of grants by enterprise and year, denoted as Patentit, where i represents the individual enterprise and t represents the corresponding year.

DOI: 10.54254/2754-1169/2026.BL34783

Core explanatory variable: Public data opening policy dummy variable (DID). Constructed in accordance with the standard rules of multi-period difference-in-differences: if the city where the enterprise is registered has launched a public data open platform in and after the current year, DID is assigned 1, otherwise 0. The policy launch time is sorted based on relevant reports of Fudan University, and this variable is generated by matching the registered city of the enterprise with the platform launch year, to identify the net effect of the public data opening policy on enterprise digital technology innovation.

Meanings of other parameters in the model: α1 is the core coefficient to be estimated, which directly reflects the net impact of public data opening on enterprise digital technology innovation; a significantly positive coefficient indicates that the policy has a positive promoting effect. μi is the enterprise fixed effect, controlling inherent enterprise characteristics that do not change over time. λt is the year fixed effect, absorbing macro-level time trend shocks. Xit is the random disturbance term, capturing unobservable random factors in the model. This setting can simultaneously strip the interference of enterprise individual heterogeneity and macro time-varying factors, and accurately identify the net policy effect.

To reduce estimation bias caused by omitted variables, this paper incorporates four types of enterprise-level control variables Xit: ① total operating income, controlling the impact of enterprise scale and market volume on innovation activities; ② net profit, characterizing enterprise profitability and internal financing constraint capacity; ③ comprehensive ESG score, measuring enterprise environmental, social responsibility and corporate governance performance, and controlling the potential impact of corporate sustainable development concepts on innovation layout; ④ asset-liability ratio, reflecting enterprise financial risk and liability structure. All continuous variables are winsorized at the 1% and 99% percentiles to avoid interference from extreme values.

The core premise for the difference-in-differences model to effectively identify causal effects is that the treatment group and the control group satisfy the parallel trend assumption. This paper uses the event study method to conduct parallel trend testing and dynamic effect analysis: taking the policy implementation year as the base period, set relative time dummy variables dk with k ranging from -8 to +8, interact them with the treatment group dummy variable, and compare whether the innovation trends of the two groups of enterprises remain consistent before policy implementation. If the coefficients of interaction terms in each period before policy implementation are not significant, the parallel trend assumption is judged to hold; meanwhile, an event study graph is drawn to visually observe the temporal evolution characteristics of policy effects.

To ensure the robustness and reliability of research conclusions, this paper sets multiple robustness tests: first, adjust the level of fixed effects by adding "industry×year" interactive fixed effects to absorb unobservable industry-level factors that change over time; second, conduct a placebo test by randomly assigning policy implementation times of each city and performing 500 simulated regressions to eliminate spurious regression interference; third, replace the dependent variable by using the enterprise digital innovation comprehensive index instead of the number of granted patents for re-regression. At the same time, subsample regressions are carried out from two dimensions of industry attributes and R&D investment intensity to explore the heterogeneous boundaries of policy effects.

At the mechanism testing level, this paper sets two core mechanism variables to verify the two transmission paths H2 and H3 respectively:

First, the degree of digital-real integration (Fusion). Quantified by the number of subcategories of IPC classification numbers covered by the enterprise's digital technology patents in that year. The total number of IPC categories involved in digital patents of each enterprise each year is counted; a

DOI: 10.54254/2754-1169/2026.BL34783

larger value indicates a wider penetration scope of digital technology into the real economy and a higher degree of digital-real integration.

Second, external environmental uncertainty (Uct). First, the standard deviation of enterprise operating income in the past five years is used to calculate the unadjusted uncertainty index Uct1; then the impact of overall industry cyclical fluctuations is eliminated to calculate the industry-adjusted uncertainty index Uct2. A higher value indicates greater fluctuations in the market and policy environment faced by the enterprise.

Based on the two mechanism variables, mechanism test models are constructed respectively:

$$\mathrm{Fusion}_{it}=\alpha_0+\alpha_1\mathrm{DID}_{it}+\gamma X_{it}+\mu_i+\lambda_t+\epsilon_{it} \tag{2}$$

$$\mathrm{Uct}_{it}=\alpha_0+\alpha_1\mathrm{DID}_{it}+\gamma X_{it}+\mu_i+\lambda_t+\epsilon_{it} \tag{3}$$

The significance and sign of the core coefficient α1 are used to determine whether public data opening affects enterprise digital technology innovation through the two paths of digital-real integration and uncertainty reduction.

4. Analysis of empirical results

4.1. Benchmark regression results

To accurately identify the real impact of public data opening on enterprise digital technology innovation, this paper takes the successive launch of government data open platforms in various regions after 2014 as a quasi-natural experiment. The model controls both enterprise fixed effects and year fixed effects, and adopts cluster-robust standard errors at the enterprise level to weaken estimation bias caused by serial correlation.

The benchmark regression results are shown in Table 1. Column (1) excludes control variables and only retains two-way fixed effects. The results show that the coefficient of the core explanatory variable DID is 0.415, which is significantly positive at the 5% statistical level. This means that compared with enterprises in cities without public data platforms, the digital technology patent output of policy-covered enterprises has increased significantly. The average number of granted digital technology patents of sample enterprises is 3.745, and the increment brought by the policy accounts for 11.08% of the mean, with clear economic and practical significance, which preliminarily verifies that research hypothesis H1 holds.

Column (2) further incorporates control variables such as total operating income, net profit, comprehensive ESG score, and asset-liability ratio. The DID coefficient rises to 0.477, which is still significantly positive at the 5% level, and the positive policy effect remains stable. At the control variable level, the coefficient of total operating income is significantly negative at the 1% level, indicating a slight negative correlation between enterprise scale and digital technology innovation, most likely due to the solidified organizational structure of large enterprises and their lagged response to emerging digital technologies. The coefficient of net profit fails the significance test, showing that the current profitability of enterprises is not a core factor restricting digital technology innovation. The coefficient of comprehensive ESG score is significantly negative, reflecting that enterprises with excellent ESG performance are more prudent in cutting-edge high-risk digital technology R&D. The coefficient of asset-liability ratio is significantly negative, indicating that higher financial leverage and greater financial risks of enterprises inhibit innovation R&D investment, which is consistent with existing research conclusions. Overall, after incorporating

DOI: 10.54254/2754-1169/2026.BL34783

control variables, the sign and significance of the core explanatory variable coefficient remain unchanged, further supporting the validity of H1.

In terms of economic connotation, the DID coefficient of 0.477 represents the average treatment effect of the public data opening policy. After the launch of urban public data platforms, the number of digital technology patents of enterprises in the treatment group increases by an average of 47.7% compared with the control group, with both statistical significance and economic practical significance. The model strips out the interference of inherent enterprise characteristics and macro time shocks through two-way fixed effects, yielding reliable conclusions and providing solid micro-empirical evidence for public data opening to empower enterprise digital technology innovation.

Table 1. Benchmark regression results

Variables	(1) Patent Grants	(2) Patent Grants
DID	0.415**	0.477**
	(0.1728)	(0.1914)
Operating Revenue		-0.000***
		(0.0000)
Net Profit		0.000
		(0.0000)
ESG Score		-0.098***
		(0.0124)
Asset–Liability Ratio		-0.028***
		(0.0081)
Constant	3.745***	11.260***
	(0.1044)	(0.9067)
Observations	45,273	39,909
R^2	0.7015	0.7055
Firm Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

Note: Robust standard errors are in parentheses; *, **, *** indicate significance at the 10%, 5%, and 1% levels respectively; the same applies to subsequent tables.

4.2. Parallel trend test

The core premise for causal identification in the multi-period difference-in-differences model is that the treatment group and the control group satisfy the parallel trend assumption before policy implementation. That is, without the shock of the public data opening policy, the changing trends of digital technology innovation levels of the two groups of enterprises would remain consistent. If this assumption is violated, the policy effect in the benchmark regression will be mixed with inherent differences between the two groups before the policy, impairing the credibility of conclusions.

This paper takes the year of policy implementation (k=0) as the base period and constructs relative time dummy variables from period -8 to period +8. The model is specified as follows:

$$\mathrm{Pantent}_{it}=\alpha+\sum_{k=-8}^{k=8}\beta_k D_{it}^{k}+\gamma X_{it}+\mu_i+\lambda_t+\epsilon_{it} \tag{4}$$

DOI: 10.54254/2754-1169/2026.BL34783

The coefficient βk represents the policy effect in period k relative to the base period. This paper focuses on whether the coefficients of each period before policy implementation are significantly different from 0 and the dynamic trend of coefficients after policy implementation.

The parallel trend test results show that the confidence intervals of estimated coefficients for all periods before the launch of the policy platform (k<0) include 0, indicating no significant difference in digital technology innovation levels between the treatment group and the control group before policy implementation. The parallel trend assumption is strictly satisfied, laying a solid foundation for causal identification of the multi-period DID model. This also proves that the growth of enterprise patent output in the benchmark regression does not stem from inherent trend differences between the two groups before the policy was introduced.

In the year of policy implementation and subsequent years, the coefficients gradually turn positive and continue to rise. In the initial stage of policy implementation, the policy effect is not yet stably released due to imperfect platform functions and insufficient data resource accumulation. Starting from the second period after implementation, the coefficients become significantly positive and expand continuously, with confidence intervals no longer covering 0. This indicates that the positive empowerment of public data opening on enterprise digital technology innovation presents obvious lagging characteristics, and as platform operation matures and data resources continue to expand, policy dividends will gradually amplify and persist in the long run. The dynamic test results further confirm the robustness of the benchmark regression conclusions and strengthen the causal interpretation of policy shocks (Figure 1).

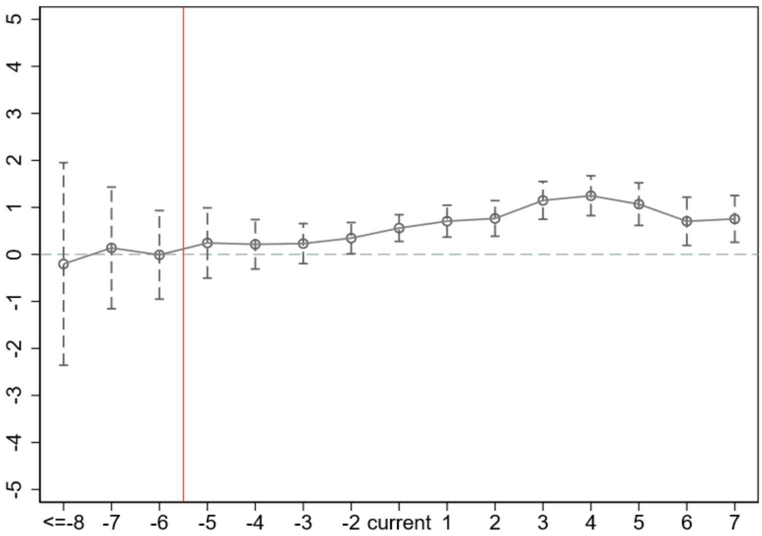

Figure 1. Parallel trend test

4.3. Robustness tests

To further verify the stability and credibility of the benchmark regression conclusions, this paper conducts robustness tests from three dimensions: adjustment of fixed effects, placebo test, and replacement of dependent variables, to avoid conclusion errors caused by model specification bias, random factor interference, and indicator measurement bias.

The benchmark model has controlled two-way fixed effects of enterprises and years, but there exist unobservable time-varying factors at the industry level, such as adjustments to industry regulatory policies and differences in technological iteration rhythms. If such factors correlate with

DOI: 10.54254/2754-1169/2026.BL34783

the public data opening policy, estimation bias may arise. Accordingly, this paper introduces industry×year interactive fixed effects on the basis of the benchmark model to fully absorb industry-level dynamic shocks.

The regression results show that after incorporating interactive fixed effects, the DID coefficient is 0.404, still significantly positive at the 5% level. Compared with the benchmark regression, the coefficient value and sign fluctuate slightly, indicating that even after controlling for industry time-varying heterogeneity, the positive promotion effect of public data opening on enterprise digital technology innovation remains robust. The benchmark model shows no obvious omitted variable bias, and the core conclusions are reliable.

Table 2. Robustness test results with replaced fixed effects

Variables	Patent Grants
DID	0.404**
	(0.1903)
Operating Revenue	-0.000***
	(0.0000)
Net Profit	0.000
	(0.0000)
ESG Score	-0.058***
	(0.0123)
Asset–Liability Ratio	0.014
	(0.0149)
Constant	8.418***
	(0.9067)
Observations	39,865
R^2	0.7261
Firm Fixed Effects	Yes
Year Fixed Effects	Yes

To eliminate spurious regression caused by unobservable random factors and interference from other concurrent policies, this paper adopts a placebo test with randomly assigned policy time points. Specifically, while keeping the treatment group samples of each city unchanged, false policy implementation years are randomly generated to construct pseudo-DID variables, and the benchmark regression is repeated 500 times. If the false coefficients are concentrated near 0, it can be proved that the real policy effect is not caused by random factors.

The kernel density results of the placebo test show that the estimated coefficients generated from 500 random simulations are normally distributed closely around 0. The absolute values of most simulated coefficients are much smaller than the real benchmark regression coefficient of 0.477, and the real estimated value lies at the right tail of the simulated distribution, far from the zero interval. This indicates that the positive effect observed in the benchmark regression entirely stems from the public data opening policy itself, rather than interference from random factors, model specification bias, or other exogenous events, further verifying the effectiveness of the model identification strategy (Figure 2).

DOI: 10.54254/2754-1169/2026.BL34783

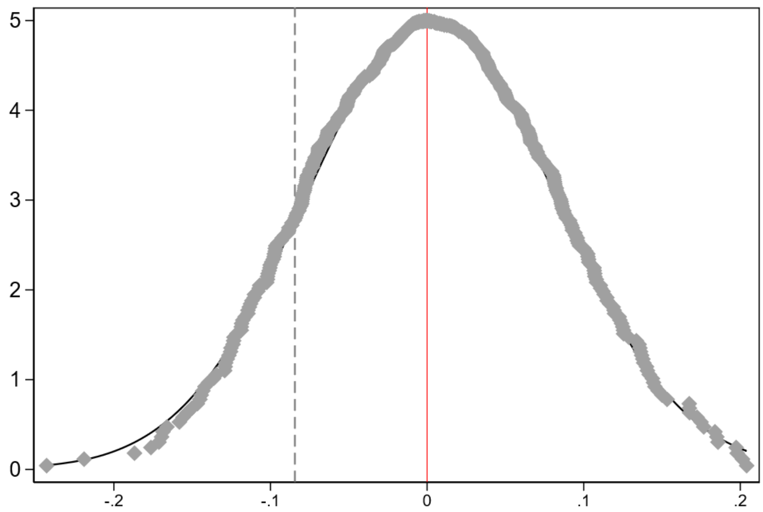

Figure 2. Placebo test results

The benchmark regression uses the number of granted digital technology patents to measure innovation level, yet a single patent indicator can hardly fully cover the comprehensive performance of enterprise digital innovation. To avoid measurement bias of the dependent variable, this paper replaces it with an enterprise digital innovation comprehensive index covering patent output, R&D investment, and digital transformation level, and re-performs regression estimation.

The results show that after controlling two-way fixed effects and enterprise control variables, the DID coefficient is 0.002, significantly positive at the 10% level. Affected by differences in dimension and calculation logic between the comprehensive index and patent quantity, the coefficient value is relatively small but still significantly positive, confirming that public data opening positively promotes comprehensive digital innovation of enterprises. Meanwhile, the regression results of each control variable are basically consistent with the benchmark model, indicating that the core conclusions do not depend on a single indicator setting and feature strong robustness.

Table 3. Robustness test results with replaced dependent variable

Variables	Corporate Digital Innovation
DID	0.002*
	(0.0011)
Operating Revenue	0.000**
	(0.0000)
Net Profit	0.000
	(0.0000)
ESG Score	0.000***
	(0.0001)
Asset–Liability Ratio	-0.000
	(0.0001)
Constant	0.043***
	(0.0054)

DOI: 10.54254/2754-1169/2026.BL34783

Table 3. (continued)

Observations	34,759
R^2	0.8891
Firm Fixed Effects	Yes
Year Fixed Effects	Yes

In summary, multiple robustness tests are consistent with the benchmark regression conclusions. The positive promotion effect of public data opening on enterprise digital technology innovation is stably established, and hypothesis H1 is fully corroborated. From the perspective of causal identification, uniform conclusions under different model specifications and variable measurement methods prove that the empirical results reflect the real effect of the public data opening policy, rather than accidental results caused by model or indicator selection.

4.4. Heterogeneity analysis

After verifying that public data opening can generally promote enterprise digital technology innovation, this paper further examines whether there are differences in policy effects among different types of enterprises. Specifically, heterogeneity analysis is conducted from two dimensions: industry attributes and R&D investment intensity.

Industries differ significantly in technology intensity, data dependence and digital foundation, leading to distinct gaps in their ability to absorb and utilize the public data opening policy. Industry-grouped regression results show that the innovation incentive effect of the policy concentrates on industries with high technology integration and strong data demand.

Among them, the DID coefficient of the manufacturing industry is significantly positive at the 10% level, with the most prominent response. The core reason is that the digital transformation of manufacturing relies heavily on data-driven production process optimization, intelligent manufacturing upgrading and product technology innovation, and public data opening can precisely match its R&D and transformation needs. The health and social work industry also presents a significant positive effect, as digital service technology R&D can be rapidly implemented based on public medical and livelihood data. In contrast, the DID coefficients of agriculture, forestry, finance, real estate and other industries are insignificant or even negative, since their digital transformation has low dependence on public data, making it difficult to effectively convert data dividends into innovation output. The overall conclusions are consistent with the research of Shen Kunrong and Lin Jianwei, namely that public data opening is more likely to release innovation value in industries with high data dependence and close digital-real integration.

Table 4. Industry heterogeneity test results

Industry Code	Industry	DID Coefficient	Standard Error	Observations	R^2
A	Agriculture, Forestry, Animal Husbandry, and Fishery	-0.786	(1.0028)	405	0.6520
B	Mining	2.301	(1.9661)	683	0.8929
C	Manufacturing	0.455*	(0.2385)	22,003	0.8172
D	Production and Supply of Electricity, Heat, Gas, and Water	-2.075*	(1.2221)	984	0.6560

DOI: 10.54254/2754-1169/2026.BL34783

Table 4. (continued)

E	Construction	0.484	(0.6495)	844	0.6779
F	Wholesale and Retail Trade	-0.043	(0.1366)	1,501	0.5606
G	Transportation, Warehousing, and Postal Services	0.250	(0.4408)	899	0.7075
H	Accommodation and Catering Services	0.033	(0.0329)	86	0.2326
I	Information Transmission, Software, and Information Technology Services	-0.028	(0.6676)	2,544	0.8129
J	Financial Services	-1.821	(1.1827)	828	0.7555
K	Real Estate	-0.155	(0.1402)	1,134	0.5643
L	Leasing and Business Services	0.016	(0.1161)	450	0.6748
M	Scientific Research and Technical Services	-1.539	(1.2601)	506	0.8681
N	Water Conservancy, Environment, and Public Facilities Management	-0.477	(0.4693)	518	0.6811
P	Education	5.878	(6.4918)	53	0.6224
Q	Health and Social Work	0.612*	(0.2948)	88	0.6432
R	Culture, Sports, and Entertainment	0.065	(0.1801)	472	0.5745
S	Public Administration, Social Security, and Social Organizations	-0.252	(0.1893)	182	0.7474
—	Other Industries	-0.149	(0.9392)	10,620	0.5279

Note: Industry codes are classified according to the *Guidelines for the Industry Classification of Listed Companies(revised in 2012)*.

Further investigation from the perspective of R&D investment reveals significant differences in the impact of public data opening on enterprises with different R&D intensities. Compared with low R&D investment enterprises, high R&D investment enterprises are subjected to a more significant policy promotion effect. Such enterprises usually have stronger data processing, technology absorption and achievement transformation capabilities, so they can more efficiently utilize public data resources to carry out digital technology innovation. Table 5 presents the regression results of R&D investment heterogeneity. Among them, the DID coefficient of the high R&D investment group is significantly positive at the 10% statistical level, while the estimated coefficient of the low R&D investment group is not only small in value but also fails the significance test. This result indicates that R&D investment level plays an important moderating role between public data opening and enterprise digital technology innovation. High R&D investment enterprises usually have more complete R&D systems, technical talent reserves and knowledge accumulation, and can more effectively identify and integrate external data resources and internalize them into innovation output [16]. On the contrary, limited by technology absorption capacity, low R&D investment enterprises struggle to effectively utilize public data resources, resulting in insignificant policy effects.

Table 5. R&D investment heterogeneity test results

Variables	High R&D Intensity GroupPatent Grants	Low R&D Intensity GroupPatent Grants
DID	4.873*	0.065
	(2.8486)	(0.2118)

DOI: 10.54254/2754-1169/2026.BL34783

Table 5. (continued)

Operating Revenue	-0.000**	0.000*
	(0.0000)	(0.0000)
Net Profit	-0.000	0.000
	(0.0000)	(0.0000)
ESG Score	-0.188**	-0.035**
	(0.0950)	(0.0153)
Asset–Liability Ratio	7.841	-0.156
	(11.3462)	(0.1010)
Constant	28.940***	4.265***
	(11.1402)	(1.2142)
Observations	13,735	13,037
R^2	0.6391	0.6716
Firm Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

5. Mechanism tests

5.1. Promoting digital-real integration

The benchmark regression has confirmed the overall positive effect of public data opening on enterprise digital technology innovation. This section specifically tests the first transmission path: promoting the deep integration of digital technology and the real economy (H2). The mechanism logic is as follows: by enriching application scenarios and lowering the threshold for cross-border data acquisition, public data opening expands the coverage of enterprises' digital technology applications and accelerates the integration and penetration of digital technology into various industries and fields. This paper takes the number of IPC classification categories covered by enterprises' digital technology patents (Fusion) as the dependent variable to conduct the mechanism regression test.

The regression results are shown in Table 6. The DID coefficient is 3.354, significantly positive at the 1% level. This indicates that after the launch of urban public data platforms, the number of IPC classifications covered by enterprises' digital technology patents increases by an average of 3.35 categories, which directly confirms that public data opening can help enterprises break through the limitations of a single technical field, carry out cross-industry and cross-category technology R&D, and accelerate the integration and penetration of digital technology into various segments of the real economy.

Public data opening boosts digital-real integration from two dimensions: scenario creation and cost reduction. On the one hand, the full opening of public data such as transportation, meteorology, geography, and livelihood data gives rise to brand-new application scenarios including intelligent transportation, precision agriculture, and smart cities, forcing enterprises to conduct targeted R&D of adaptive technologies and naturally extend patent layout to more technical fields. On the other hand, public data opening greatly reduces enterprises' cost of cross-border data collection. Enterprises can integrate public data with their own business data and industry-specific data, explore integration points between digital technology and traditional industries, and promote the full

DOI: 10.54254/2754-1169/2026.BL34783

embedding of digital technology into traditional production and operation links [17]. In summary, digital-real integration is validated as a core transmission path through which public data opening empowers digital technology innovation, providing empirical support for Hypothesis H2.

Table 6. Regression results of the digital-real integration mechanism

Variables	IPC Categories
DID	3.354***
	(1.2823)
Operating Revenue	0.000***
	(0.0000)
Net Profit	-0.000
	(0.0000)
ESG Score	-0.074
	(0.0696)
Asset–Liability Ratio	0.041
	(0.0838)
Constant	17.604***
	(4.8828)
Observations	49,704
R^2	0.6723
Firm Fixed Effects	Yes
Year Fixed Effects	Yes

5.2. Reducing uncertainty

The second mechanism hypothesis H3 of this paper holds that public data opening mitigates external uncertainties such as market demand, policy changes, and technological route selection faced by enterprises by providing authoritative and multi-dimensional macro information, thus boosting enterprises' willingness to invest in high-risk digital technology R&D. This paper tests this transmission path by adopting the environmental uncertainty index Uct1 without industry adjustment and the industry-adjusted index Uct2 as dependent variables respectively.

The regression results in Table 7 show that in Column (1) with Uct1 as the dependent variable, the DID coefficient is -0.006, significantly negative at the 5% level; in Column (2) with the industry-adjusted Uct2 as the dependent variable, the DID coefficient is -0.045, significantly negative at the 10% level. The consistent results of both groups indicate that after the implementation of the public data opening policy, the fluctuation range of enterprises' operating income narrows significantly, and the uncertainty of the external market and policy environment decreases notably.

At the control variable level, the coefficient of the comprehensive ESG score remains significantly negative, suggesting that enterprises with higher ESG governance performance have stronger capabilities in predicting and resisting external risks and face lower environmental uncertainty. The decline in external uncertainty stabilizes enterprises' expected returns on innovation, weakens risk concerns about digital technology R&D, and encourages enterprises to increase investment in cutting-edge technology R&D. This confirms that reducing external

DOI: 10.54254/2754-1169/2026.BL34783

environmental uncertainty is an important transmission path for public data opening to promote enterprise digital technology innovation, and Hypothesis H3 is supported.

Table 7. Regression results of the uncertainty reduction mechanism

Variables	(1) Uct1	(2) Uct2
DID	-0.006**	-0.045*
	(0.0027)	(0.0264)
Operating Revenue	-0.000	-0.000
	(0.0000)	(0.0000)
Net Profit	0.000	0.000**
	(0.0000)	(0.0000)
ESG Score	-0.001***	-0.010***
	(0.0002)	(0.0017)
Asset–Liability Ratio	0.001	0.011
	(0.0012)	(0.0113)
Constant	0.226***	2.113***
	(0.0131)	(0.1263)
Observations	29,914	29,914
R^2	0.4971	0.4553
Firm Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes

The two mechanisms do not function independently but form a complementary and interactive logic. Public data opening first eases enterprises' external information friction and environmental fluctuations by releasing macro-authoritative information, stabilizes expected returns on innovation, and encourages enterprises to engage in high-risk digital technology R&D. On this basis, relying on diversified application scenarios and low-cost data supply, it promotes cross-border penetration of digital technology and deep integration of digital and real economies, further unlocking the innovation value of data elements. The coordinated effect of the dual paths forms a complete logical chain for public data opening to empower enterprise digital technology innovation, and also provides empirical support for the government to optimize data opening governance and for enterprises to revitalize the value of public data.

6. Conclusions and policy implications

6.1. Research conclusions

Taking the panel data of China's A-share listed enterprises from 2010 to 2022 as samples, this paper regards the launch of local government public data open platforms as a quasi-natural experiment, and uses a multi-period difference-in-differences model to systematically investigate the policy effects, transmission mechanisms and heterogeneous characteristics of public data opening on corporate digital technology innovation. The main research conclusions are as follows:

First, public data opening has a significant positive promoting effect on enterprise digital technology innovation. After the establishment of urban public data open platforms, the granted output of enterprises' digital technology patents increases significantly. This core conclusion remains

DOI: 10.54254/2754-1169/2026.BL34783

stable after multiple robustness tests including parallel trend test, fixed effect adjustment, placebo test and replacement of core variables, proving that public data opening can effectively empower enterprise digital technology innovation and support industrial digital transformation and the cultivation of new-quality productive forces.

Second, the policy effects of public data opening show significant heterogeneity. At the industry level, industries with high data dependence and close digital-real integration such as manufacturing, health and social work release more sufficient innovation dividends from the policy; industries such as traditional agriculture, forestry, finance and real estate have insignificant policy effects due to constraints of their own transformation needs. At the enterprise characteristic level, enterprises with high R&D investment can efficiently utilize public data resources to improve innovation output by virtue of stronger technology absorption and achievement transformation capabilities; enterprises with low R&D investment are restricted by technological foundation and talent reserve, making it difficult to take over policy dividends, resulting in insignificant positive effects.

Third, public data opening drives digital technology innovation through the dual paths of deep digital-real integration and reduction of external environmental uncertainty. On the one hand, public data opening broadens the application scenarios of digital technology, reduces the cost of cross-border data acquisition, expands the coverage of IPC classifications of enterprises' digital patents, and promotes the penetration and integration of digital technology into various fields of the real economy. On the other hand, public data outputs authoritative and multi-dimensional macro information, which effectively mitigates operational uncertainty caused by market demand and policy changes, stabilizes expected returns on innovation, and boosts enterprises' willingness to invest in R&D. The two paths connect and coordinate with each other, jointly forming the internal logic of public data opening empowering innovation.

6.2. Policy implications

Based on the above research conclusions, this paper puts forward targeted policy suggestions from three levels: government governance, enterprise development and institutional construction, so as to further release the value of public data factors and strengthen their empowering effect on digital technology innovation and industrial transformation.

First, at the government level: promote high-quality hierarchical opening of public data and optimize the platform service system. Firstly, improve the standards and quality control mechanisms for public data opening, unify data formats, update frequencies and interface specifications, and give priority to the opening of high-value data such as economic operation, transportation networks, livelihood services and industry supervision, so as to ensure the authority, timeliness and practicability of data and give full play to the role of public data in alleviating enterprise uncertainty. Secondly, implement differentiated opening strategies, carry out in-depth opening pilots in data-dependent industries such as manufacturing, medical health and information technology preferentially according to industry heterogeneity, customize industry-specific data opening lists and application scenarios, and accurately match enterprises' innovation and R&D needs. Thirdly, upgrade the functions of government data open platforms, simplify data acquisition processes, provide standardized interfaces, development toolkits and application cases to lower the threshold for enterprises to use data; simultaneously improve the construction of digital infrastructure, strengthen the cultivation of digital technology professionals, and build a complete ecosystem of data supply, technical support and talent guarantee.

Second, at the enterprise level: strengthen data absorption and integration capabilities and take the initiative to deploy digital-real integrated innovation. Enterprises should actively adapt to the

DOI: 10.54254/2754-1169/2026.BL34783

trend of public data opening and improve the ability to transform external data resources. Firstly, set up professional data governance and analysis teams, deeply integrate public data with their own business data, production data and customer data, explore the application value of public data in R&D design, production optimization, market prediction, supply chain management and other links, and broaden the boundaries of digital technology R&D relying on public data. Secondly, increase R&D investment in digital technologies such as artificial intelligence, big data and Internet of Things, improve technology and talent reserve, and enhance the ability to absorb, analyze and transform public data into achievements. In particular, high-R&D enterprises should give play to their own advantages and take the lead in building benchmark scenarios for innovative application of public data. Thirdly, actively participate in government-enterprise and enterprise-enterprise collaborative innovation, jointly build a data application ecosystem with governments, scientific research institutions and industrial chain enterprises, share data application experience under the premise of compliance to achieve cross-border collaborative innovation; meanwhile, improve the internal data security and compliance system, adhere to the bottom line of data security, and maximize the innovative value of public data.

Third, at the institutional level: improve public data governance regulations and build a market-oriented operation system. Firstly, perfect laws and regulations related to data security and privacy protection, clearly define the boundaries of rights and responsibilities and the division of security responsibilities for public data opening, dispel the risk concerns of government departments in data opening, specify the mandatory opening scope of non-secret-related and non-sensitive public data, and improve the overall opening efficiency. Secondly, establish a hierarchical and classified management system for public data, divide opening levels according to data sensitivity, and balance the inclusive sharing of public data with security and confidentiality needs; build a compliance governance framework covering the whole life cycle of data collection, storage, circulation, development and application, and clarify the rights and obligations of various market entities. Thirdly, moderately explore market-oriented models such as authorized operation of public data and data trust, introduce professional market entities to participate in the deep processing and scenario-based development of public data on the premise of prioritizing public attributes and safeguarding public interests, and improve the allocation efficiency and application value of data resources; construct a data factor governance system guided by the government, participated by the market and coordinated by multiple entities, so as to provide institutional guarantee for the long-term empowerment of public data in digital technology innovation.

6.3. Research limitations and future prospects

Although this paper uses a multi-period difference-in-differences model to complete empirical tests and mechanism decomposition, there are still certain research limitations. First, the sample only covers large and medium-sized A-share listed enterprises, excluding a large number of small and medium-sized enterprises. The data acquisition capacity and innovation models of SMEs differ from those of listed enterprises, leading to certain limitations in the external applicability of the conclusions. Second, only the number of granted patents is used to measure digital technology innovation, which cannot fully reflect innovation quality, technological value and market industrialization effects. Third, only two core transmission paths are verified, without further distinguishing the differential impacts of different types of public data and different application scenarios on innovation effects.

Follow-up research can be expanded and deepened from multiple dimensions. First, broaden the sample scope by including small, medium and micro enterprises and enterprises of different ownerships to conduct heterogeneity analysis and improve the universality of conclusions. Second, construct multi-dimensional evaluation indicators of innovation quantity, innovation quality and market transformation to more comprehensively measure the level of digital technology innovation. Third, refine the dimensions of public data quality (timeliness, granularity, coverage) and platform functional characteristics to explore their moderating effects on innovation effects. Fourth, expand the research perspective to the regional innovation ecosystem and analyze the impact of public data opening on regional innovation networks and industrial collaborative innovation. Fifth, carry out cross-regional and cross-institutional comparative studies, summarize replicable and promotable public data governance experience, and provide Chinese practical evidence for global data factors empowering innovation.

References

[1] Shen, K. R., & Lin, J. W. (2025). From "data islands" to "data continents": The technological innovation effects of public data openness [in Chinese]. Management World, 41(2), 83–104. https: //doi.org/10.19744/j.cnki.11-1235/f.2025.0025
[2] Peng, Y. H. (2023). The value creation effects of government data openness: Evidence from the perspective of firms' total factor productivity [in Chinese]. The Journal of Quantitative & Technical Economics, 40(9), 50–70. https: //doi.org/10.13653/j.cnki.jqte.20230725.001
[3] Fang, J. C., Liu, Y., Gao, H. Y., et al. (2023). Can public data openness promote regional coordinated development? Evidence from a quasi-natural experiment of government data platform launches [in Chinese]. Management World, 39(9), 124–142. https: //doi.org/10.19744/j.cnki.11-1235/f.2023.0112
[4] Cai, Y. K., Zhou, J. K., & Yuan, W. P. (2024). Data factor sharing and urban entrepreneurial vitality: Empirical evidence from public data openness [in Chinese]. The Journal of Quantitative & Technical Economics, 41(8), 5–25. https: //doi.org/10.13653/j.cnki.jqte.20240619.004
[5] Hu, Z. X., & Ma, S. Z. (2023). Market integration and corporate digital innovation: With a discussion of digital innovation measurement methods [in Chinese]. Economic Research Journal, 58(6), 155–172.
[6] Huang, B., Li, H. T., Liu, J. Q., et al. (2023). Digital technological innovation and the high-quality development of Chinese enterprises: Evidence from corporate digital patents [in Chinese]. Economic Research Journal, 58(3), 97–115.
[7] Tao, F., Zhu, P., Qiu, C. Z., et al. (2023). The impact of digital technological innovation on corporate market value [in Chinese]. The Journal of Quantitative & Technical Economics, 40.
Note: The original reference lacks issue and page numbers.
[8] Chen, Y. L., & Jiang, Q. (2024). Open public data and enterprise innovation from the perspective of data as a production factor: Evidence from a quasi-natural experiment of public data development platforms [in Chinese]. Business and Management Journal, 46(1), 5–24.
[9] Jiang, J., Yin, Y. L., & Quan, L. P. (2025). How does government public data openness promote the integration of digital and real technologies? [in Chinese]. Shanghai Journal of Economics, 12, 29–40. https: //doi.org/10.19626/j.cnki.cn31-1163/f.2025.12.004
[10] Akcigit, U., & Liu, Q. (2016). The role of information in innovation and competition. Journal of the European Economic Association, 14(4), 828–870. https: //doi.org/10.1111/jeea.12185
[11] Hughes-Cromwick, E., & Coronado, J. (2019). The value of U.S. government data to U.S. business decisions. Journal of Economic Perspectives, 33(1), 131–146. https: //doi.org/10.1257/jep.33.1.131
[12] Xu, K. (2021). The legitimacy and boundaries of data scraping [in Chinese]. China Legal Science, 2, 166–188. https: //doi.org/10.14111/j.cnki.zgfx.2021.02.008
[13] Zheng, L., & Lü, W. Z. (2017). Research on the outputs and effects of public data openness: Evidence from the Shanghai Open Data Innovation Application Contest [in Chinese]. E-Government, 9, 2–10. https: //doi.org/10.16582/j.cnki.dzzw.2017.09.001
[14] Xie, K., Xia, Z. H., & Xiao, J. H. (2020). The enterprise realization mechanism of big data as a real production factor: A product innovation perspective [in Chinese]. China Industrial Economics, 5, 42–60. https: //doi.org/10.19581/j.cnki.ciejournal.2020.05.014
[15] Schumpeter, J. A. (1934). The theory of economic development. Commercial Press.
Note: Since the Chinese edition was cited, another acceptable APA format is:

DOI: 10.54254/2754-1169/2026.BL34783

Schumpeter, J. A. (1934/Chinese ed.). The theory of economic development. Commercial Press.
[16] Ge, X. (2023). The impact of smart city construction on green economic development: An empirical test based on the PSM-DID model [in Chinese]. Trade Fair Economy, 24, 36–42. https: //doi.org/10.19995/j.cnki.CN10-1617/F7.2023.24.036
[17] Yang, Z. X., & Jin, R. Y. (2025). Smart city construction and the coupled coordinated development of urban green economies [in Chinese]. Modern Urban Research, 1, 103–111.