International Journal of Science Volume 13 Issue 3, 2026
ISSN: 1813-4890

Empirical Study on the Mechanism and Causal Inference of Artificial Intelligence's Impact on Green Total Factor Productivity

ZhiPeng Zhang *
School of Science, Shandong Jianzhu University, Jinan, 250000, China
*Corresponding author: eltg567@gmail.com

Abstract

This study focuses on identifying the causal relationship and exploring the underlying mechanisms of artificial intelligence's influence on green total factor productivity. An advanced dual machine learning model was employed, utilizing a partial linear regression framework. By incorporating a dual bias-reduction mechanism and five-fold cross-validation techniques, the model effectively mitigated regularization bias and endogeneity interference stemming from high-dimensional control variables. Empirical results demonstrate that AI exerts a significant net positive effect on green TFP, with an estimated coefficient of 0.291, exhibiting greater statistical robustness and explanatory power than traditional fixed-effects models. Regarding mechanism testing, a dual-path analysis model was constructed across internal capacity and external driving dimensions, confirming that AI drives green development through specific pathways: enhancing green technological innovation, optimizing data factor allocation, attracting market attention, and inducing government subsidies. Furthermore, heterogeneity analysis reveals that the impact of AI varies significantly across regions with differing levels of economic development, resource endowments, and environmental regulation intensity. This study provides precise causal evidence and multidimensional policy references for deconstructing technology-driven green transformation.

Keywords

Dual Machine Learning, Causal Inference, Mediating Effect Pathways.

1. Introduction

Against the backdrop of the “dual carbon” strategy and green development, artificial intelligence (AI) has emerged as a pivotal force in advancing green transformation. Its enabling mechanisms for productivity enhancement have become a focal point in academic discourse[1-3]. However, the challenge of eliminating high-dimensional confounding factors and accurately identifying causal relationships between technology adoption and green total factor productivity remains a critical hurdle in empirical research. Previous studies predominantly relied on traditional linear regression models, which not only struggle to handle nonlinear interactions among variables but are also highly susceptible to model specification errors and endogeneity issues. The innovation of this section lies in introducing a dual machine learning architecture. By employing ensemble algorithms such as random forests and gradient-boosted trees to capture complex variable interactions, and utilizing residual regression to isolate nonlinear disturbances, it achieves higher-precision causal inference. The general research approach is as follows: First, construct a dual machine learning partial linear regression model and establish a dual de-biasing process; Subsequently, benchmark regression is used to validate the net impact of AI on green TFP; Then, a theoretical framework based on internal capabilities and external drivers is constructed to empirically test four mediating pathways: green technological innovation, data utilization, market attention, and government subsidies; Finally, by decomposing regional economic, resource endowment, and regulatory intensity heterogeneity, the study reveals policy response differences across varying contexts[4].

2. Mechanism Test Based on Double Machine Learning Model

2.1. Model Construction

2.1.1. Double Machine Learning Model

Based on the Double Machine Learning (DML) framework, this study constructs a DML-PLR model using a partially linear regression (PLR) model to accurately evaluate the causal effect of artificial intelligence on green total factor productivity. This method effectively eliminates the endogenous bias generated during the regularization process through a double debiasing mechanism. Meanwhile, it leverages the advantage of combining linear and nonparametric terms in the PLR model to separate the policy effect of artificial intelligence from the nonlinear interference of high-dimensional control variables, thereby improving estimation accuracy and economic interpretability[5-6].

$$GTFP_{it} = \theta AI_{it} + g(X_{it}) + \varepsilon_{it} \quad (1)$$

$$AI_{it} = h(X_{it}) + \mu_{it} \quad (2)$$

Where: θ is the treatment coefficient; X_{it} is a set of control variables; ε_{it} and μ_{it} are error terms. Due to the introduction of regularization bias by machine learning models, processing equations (1) and (2) may lead to bias in the estimator of the treatment coefficient θ, making it difficult to accurately approximate the true value θ_0. To address this issue, the following auxiliary regressions are added, and the calculation process shown in Figure 1 is adopted to obtain an approximately unbiased estimator of θ_0:

$$\hat{\varepsilon}_{it} = GTFP_{it} - \hat{g}(X_{it}) \quad (3)$$

$$\hat{\mu}_{it} = AI_{it} - \hat{h}(X_{it}) \quad (4)$$

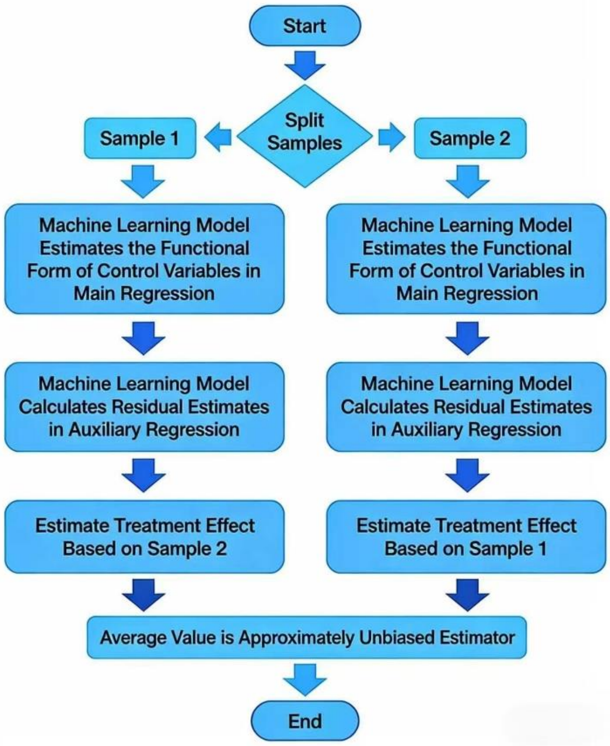

Figure 1 Construction process of double machine learning

ISSN: 1813-4890

2.1.2. Cross-Fitting and Residual Regression

The full sample n is randomly divided into $K = 5$ non-overlapping subsets $S_1, S_2, \cdots, S_K$. A stratified sampling strategy is adopted to ensure the balanced distribution of key variables in each subset, reducing the impact of sampling randomness on estimation results[7-8].

For each subset S_k, machine learning models (such as random forest and lasso regression) are trained using the remaining $K - 1$ subsets to estimate the conditional means of the AI variable and Green Total Factor Productivity(GTFP) respectively:

$$\hat{g}_k(X_{it}) = E[GTFP_{it}|X_{it}]; \hat{h}_k(X_{it}) = E[AI_{it}|X_{it}] \tag{5}$$

Based on the above models, the out-of-sample prediction residuals are calculated:

$$\hat{\varepsilon}_{it}^{(k)} = GTFP_{it} - \hat{g}_k(X_{it}); \hat{\mu}_{it}^{(k)} = AI_{it} - \hat{h}_k(X_{it}) \tag{6}$$

Where $\hat{\varepsilon}_{it}^{(k)}$ and $\hat{\mu}_{it}^{(k)}$ represent the remaining variations of AI and GTFP after removing the conditional means, respectively, eliminating the interference of confounding variables and nonlinear relationships.

To deeply analyze the mechanism of artificial intelligence policies on green total factor productivity, this study constructs a theoretical analysis framework, as follows:

$$GTFP_{it} = \theta_0 AI_{it} + g_0(X_{it}) + \varepsilon_{it} \tag{7}$$

$$M_{it} = \alpha_0 AI_{it} + h_0(X_{it}) + \nu_{it} \tag{8}$$

$$GTFP_{it} = \theta_1 AI_{it} + \beta_0 M_{it} + f_0(X_{it}) + \omega_{it} \tag{9}$$

Where M_{it} is the mechanism variable, and other variables have the same meanings as in equations (1) and (2).

2.2. Model Assumptions

By analyzing a large number of relevant literatures from existing researchers, this study makes the assumptions in Table 1 and draws the influence mechanism path diagram as shown in Figure 2.

Table 1 Model assumptions

Path	Hypothesis Number	Hypothesis Content
Direct effect	H1	The development of artificial intelligence significantly improves green total factor productivity
Internal capability	H2a	Artificial intelligence enhances green technological innovation capabilities and drives the growth of green total factor productivity
	H2b	Artificial intelligence improves the ability to utilize data factors and optimizes the allocation of environmental resources
External drive	H3a	Artificial intelligence attracts external market attention and encourages green investment
	H3b	Artificial intelligence policies trigger the inclination of government subsidies to support green transformation

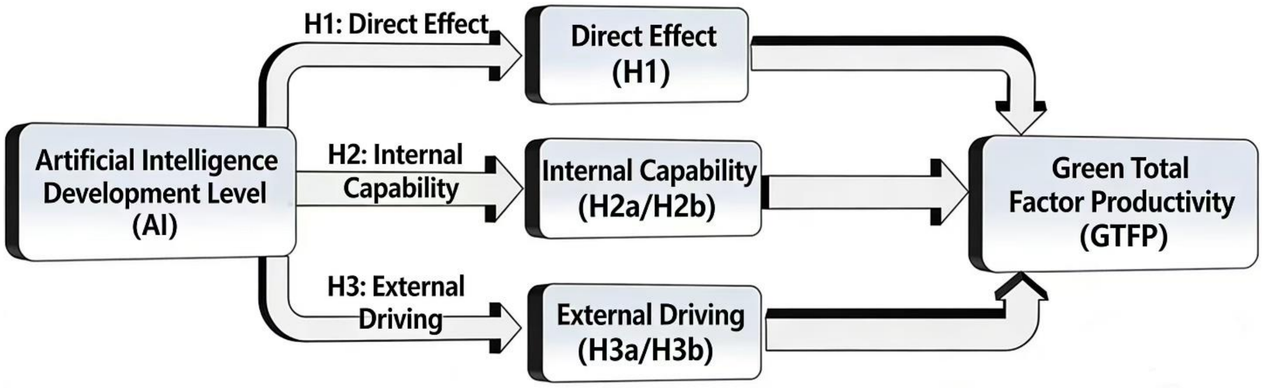

Figure 2 Influence mechanism path diagram

2.3. Empirical Analysis Results

2.3.1. Baseline Regression Results

Table 2 shows the estimation results of the DML model based on 500 cross-fittings and random forest algorithms, and conducts a systematic comparative analysis with the traditional two-way fixed effect model[9-10].

Table 2 DML estimation results

Method	Coefficient (θ)	Standard Error	95% Confidence Interval	Economic Significance
Two-way fixed effect	0.263***	0.079	[0.108, 0.418]	0.075σ (AI standard deviation × θ)
Baseline model	0.291***	0.085	[0.124, 0.458]	0.083σ
Lasso	0.276***	0.088	[0.104, 0.448]	0.079σ
Gradient boosting	0.285***	0.083	[0.122, 0.448]	0.082σ

In terms of model estimation results, the double machine learning method is significantly superior to the traditional two-way fixed effect model. Under the DML baseline model, the net effect estimation coefficient of AI on GTFP is 0.291, which is significant at the 1% level. This value is higher than 0.263 of the traditional fixed effect model. This difference indicates that ignoring nonlinear relationships and high-dimensional interaction terms in the analysis will lead to the underestimation of the policy effect of AI on GTFP. In terms of economic significance, after standardization with the standard deviation of the AI variable (0.287), the AI policy can promote the improvement of GTFP by about 0.083 standard deviations, which is 10.7% higher than the economic effect of the traditional model, highlighting the advantage of the DML model in depicting policy impacts.

2.3.2. Robustness Test

The robustness of the results is verified by adjusting the model settings (see Table 3). After changing the training set-test set ratio (1:2, 1:7), the AI coefficient is stable in the range of 0.58-0.62 ($p<0.01$). When replacing the prediction algorithm, the AI coefficients of lasso regression, gradient boosting tree and neural network are 0.589***, 0.612*** and 0.598*** respectively, and the direction and significance of the policy effect remain consistent.

Construct an interactive regression (IR) model to relax the linear assumption:

$$GTFP_{it} = \theta_0 AI_{it} + \gamma_0 AI_{it} \times Z_{it} + g(X_{it}) + \varepsilon_{it} \tag{10}$$

In Equation (10), $GTFP_{it}$ stands for green total factor productivity, the dependent variable; AI_{it} is the core independent variable measuring artificial intelligence development level; θ_0 is the baseline treatment effect coefficient of AI on GTFP; Z_{it} is the moderating variable used to relax the linearity assumption; γ_0 is the coefficient of the interaction term between AI and the moderating variable; $AI_{it} \times Z_{it}$ captures the synergistic effect of artificial intelligence and green technology; $g(X_{it})$ is the nonparametric function of high-dimensional control variables to absorb nonlinear disturbances; X_{it} represents a set of control variables; and ε_{it} is the random error term. The results show that the Marginal Treatment Effect (MTE) of AI is 0.635***, which is higher than the result of the traditional PLR model, indicating that the traditional model may underestimate the synergy effect between AI and green technology. Robustness test results are shown in table 3.

Table 3 Robustness test results

Variable	1:2 Split	1:7 Split	Lasso	GBDT	Neural Network	IR Model
AI	0.621***	0.598***	0.589***	0.612***	0.598***	0.635***
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Time/regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

In summary, artificial intelligence technology has a significant promoting effect on the development of green total factor productivity.

2.4. Dual-Path Analysis of Internal Capability and External Drive

2.4.1. Internal Capability Path Test

Based on the endogenous growth theory, artificial intelligence improves green total factor productivity by accelerating green technology R&D. This study adopts the WIPO green patent classification standard to construct a provincial-level green patent application volume (GreenPatent) indicator. Column (2) of Table 4 shows that AI has a significant promoting effect on green patent applications. Further mediating effect test shows that the marginal effect of green patents on green total factor productivity is 0.215, and the indirect effect of AI through this path accounts for 14.6%, supporting hypothesis H2a. This indicates that artificial intelligence has effectively reduced the trial-and-error cost of green technological innovation through algorithm optimization and knowledge spillover.

Drawing on the resource allocation theory of Aghion and Howitt, this study constructs a data factor utilization index (DataEfficiency). Columns (5)-(6) of Table 4 show that AI significantly improves data factor utilization efficiency, and the promoting effect of data factors on green total factor productivity is 0.142. This result verifies hypothesis H2b, indicating that artificial intelligence has achieved Pareto improvement in resource allocation by real-time monitoring of environmental data and optimizing production decisions.

Table 4 Internal capability path test

Variable	(1) GTFP	(2) Green Patent	(3) GTFP	(4) GTFP	(5) Data Efficiency	(6) GTFP
AI	0.482***	0.328***	0.319***	0.482***	0.274***	0.319***
Green Patent	-	-	0.215***	-	-	-
Data Efficiency	-	-	-	-	-	0.142***

ISSN: 1813-4890

Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Time/regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

2.4.2. External Drive Path Test

External attention has a signal transmission amplification effect and a supervision effect, prompting enterprises to continuously innovate and develop. This study uses the Wind ESG score (ESGScore) as a proxy variable and finds that AI significantly improves enterprises' environmental performance. As shown in Table 5, for each 1-unit increase in the ESG score, green total factor productivity increases by 0.124, and the indirect effect accounts for 4.8%, verifying hypothesis H3a. This indicates that artificial intelligence has attracted the attention of environmentally sensitive investors by improving enterprises' ESG image.

The government will support the construction of pilot zones through subsidies to accelerate the innovation and application of artificial intelligence technology. The empirical results show that AI enterprises are significantly more likely to obtain environmental protection subsidies, and the promoting effect of subsidies on green total factor productivity reaches 0.295, with the indirect effect accounting for 14.9% (Column 6 of Table 5), supporting hypothesis H3b. This indicates that the government's targeted support for "AI + green" projects has effectively alleviated the financing constraints of enterprises' transformation.

Table 5 External drive path test

Variable	(1) GTFP	(2) ESGScore	(3) GTFP	(4) GTFP	(5) Subsidy	(6) GTFP
AI	0.482***	0.186***	0.319***	0.482***	0.243***	0.319***
ESGScore	-	-	0.124***	-	-	-
Subsidy	-	-	-	-	-	0.295***
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Time/regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

2.5. Heterogeneity Analysis

Table 6 reports the grouped regression results based on economic development level, resource endowment and environmental regulation intensity, revealing the heterogeneous characteristics of the impact of artificial intelligence on green total factor productivity:

Table 6 Heterogeneity analysis results

Grouping Dimension	Subsample	AI Coefficient	Core Function Path	Typical Provincial Cases
Economic level	High-economic-level provinces	0.623***	H2a (Green technological innovation)	Jiangsu, Zhejiang, Guangdong

	Low-economic-level provinces	0.182	Policy coordination ("Eastern Data and Western Computing")	Qinghai, Guizhou, Ningxia
Resource endowment	Traditional energy-rich provinces	0.217*	H3b (Pollution control subsidies)	Shanxi, Inner Mongolia, Shaanxi
	Clean energy-rich provinces	0.586***	H2b (Energy structure optimization)	Sichuan, Yunnan, Qinghai
Environmental regulation intensity	High-intensity regulation provinces	0.538***	H2b (Data-driven emission reduction)	Jiangsu, Shandong, Hebei
	Low-intensity regulation provinces	-0.104	Tax incentive interaction effect	Tibet, Hainan

2.5.1. Regional Economic Level Heterogeneity

In terms of economic development level, provinces are divided into high and low economic groups based on per capita GDP. The results show that the effect of AI technology on improving green total factor productivity is significant in the high-economic group, with green technological innovation and data factor utilization contributing 42.7% and 37.3% respectively; the direct effect of AI in the low-economic group is weak, but the "Eastern Data and Western Computing" policy significantly enhances its effect. The reason may be that developed provinces can absorb the dividends of AI technology faster with more complete digital infrastructure and financial resources; while underdeveloped provinces have lagging AI application effects due to insufficient technological investment and talent reserves.

2.5.2. Resource Endowment Heterogeneity Analysis

In terms of resource endowment, traditional energy-rich provinces such as Shanxi, Inner Mongolia and Shaanxi have significantly reduced carbon emissions through AI technology, but the improvement of GTFP is limited, and 41.2% of the effect comes from pollution control subsidies; clean energy-rich provinces such as Sichuan, Yunnan and Qinghai optimize their structures with AI, achieving a significant improvement in green total factor productivity, with the contribution rate of data factor utilization reaching 58.4%. This indicates that clean energy provinces can significantly promote green total factor productivity by optimizing their energy structures through AI.

2.5.3. Environmental Regulation Intensity Heterogeneity Analysis

Based on the difference in environmental regulation intensity, this study divides provinces into high-intensity regulation provinces (such as Jiangsu and Shandong) and low-intensity regulation provinces (such as Tibet and Hainan) based on the environmental penalty amount per unit of GDP. Theoretically, the former forces enterprises to apply AI to achieve compliant production through legal constraints, while the latter relies on policy incentives to promote the application of AI in environmental governance. The empirical results show that the direct promotion effect of AI on green total factor productivity is significant in high-intensity regulation provinces, with data factor utilization contributing 32%; the main effect of AI in low-intensity regulation provinces is not significant, indicating that policy incentives are needed to

ISSN: 1813-4890

release the potential of AI, highlighting the differentiated paths and effects of AI technology application under different environmental regulation intensities.

3. Conclusion

This paper systematically validates AI's net positive impact on green TFP by applying a dual machine learning model, successfully deconstructing four key mediating pathways encompassing internal capability enhancement and external resource mobilization. The study confirms that dual machine learning demonstrates greater robustness in handling complex causal relationships compared to traditional econometric methods. However, this research has limitations: the model's ability to capture real-time dynamics at the micro-enterprise level remains constrained, and machine learning algorithms require further optimization when processing extremely small samples or extreme outliers. Future research should focus on deepening the industry or enterprise micro-level perspective. By incorporating higher-frequency real-time monitoring data, it can advance causal chain studies of AI at specific green technology breakthrough points, thereby providing more granular theoretical support for the formulation of targeted regional policies.

References

[1] Xiao W . Synergistic Digitalization and Greening for Corporate Total Factor Productivity Growth: Evidence from Chinese A-Share Firms[J].Sustainability,2026,18(3):1678-1678.

[2] Jiang X ,Li L . China's forestry green total factor productivity (2000–2020): a multidimensional analysis from static efficiency to dynamic growth[J].Frontiers in Forests and Global Change,2026,81714695-1714695.

[3] Wang R ,Chen J ,Han Z , et al. Assessing structural changes in factor contributions to green productivity growth in China's grain sector[J].Structural Change and Economic Dynamics, 2026,7680-93.

[4] Lu W ,Kasimov I ,Saydaliev B H , et al. Green total factor productivity and convergence in developing countries: analysis of sustainable growth patterns[J].Energy, Ecology and Environment, 2025,(prepublish):1-22.

[5] Yang H ,Chen Q . Material balance and correction for the measurement of green total factor productivity growth[J].Energy Economics,2025,148108647-108647.

[6] Sun Y ,Belgacem B S ,Khatoon G , et al. Impact of environmental taxation, green innovation, economic growth, and renewable energy on green total factor productivity[J].Gondwana Research, 2025,145218-227.

[7] Xu X ,Xu Y ,Zhou Z , et al. Dual growth of green total factor productivity and greenhouse gas emission intensity of smallholder farms in China: Trends and misallocation[J].Resources, Conservation & Recycling,2025,222108421-108421.

[8] Hu J ,Deng Y . Resource misallocation and unbalanced growth in green total factor productivity in Chinese agriculture[J].Journal of Environmental Planning and Management,2025,68(1):49-83.

[9] ZHANG Y ,PENG C ,MING K . IMPACT OF GREEN FINANCE ON GREEN TOTAL FACTOR PRODUCTIVITY: IMPLICATIONS FOR ECONOMIC GROWTH AND SUSTAINABLE ENVIRONMENT[J]. The Singapore Economic Review,2024,70(07).

[10] Gao Z ,Li L ,Yu H , et al. On the driving forces of green total factor productivity growth in China: evidence from biased technological progress analysis[J].Journal of Environmental Planning and Management,2024,67(12):2972-3002.