# Journal of Applied AI and Interdisciplinary Innovations

ISSN 3104-7017
eISSN 3104-7025

Volume 1, Issue 1
February 2026

9 773104 701258

HKIHER
香港高等教育研究所
HONG KONG INSTITUTE OF HIGHER EDUCATION RESEARCH

HEPM
HIGHER EDUCATION PUBLISHING MANAGEMENT
高图学术出版

# Journal of Applied AI and Interdisciplinary Innovations

**Sponsor** Hong Kong Institute of Higher Education Research

**Operations** Higher Education Publishing Management

**Editorial Publishing** Editorial Office of the Journal of Applied Artificial Intelligence and Interdisciplinary Innovation

# TABLE OF CONTENTS

# The Impact of Urban Digital Development on Corporate Environmental Information Disclosure

*Lipai Zhang[1], Jihan Wang[2*]*

*[1]Development authority of Shenzhen park of Hetao Shenzhen-Hongkong science and technology innovation cooperation Zone 518045, China; [2*]Business School, The University of Sydney 2006, Australia*

Keywords:

Urban digital development;
Environmental information disclosure;
Disclosure quality;
Market signaling;
Green finance;
Marketization

ABSTRACT

Based on a market-signaling perspective, this study explores whether and how urban digital development affects corporate environmental information disclosure (EID). Using Chinese A-share listed firms from 2011–2019 as the sample, and employing fixed-effects panel regressions and a series of robustness checks, we find that urban digital development significantly increases both (i) whether firms disclose environmental information and (ii) the quality of such disclosure. Mechanism tests show that urban digital development improves the quality of EID through (1) promoting regional marketization and intensifying market competition, and (2) strengthening green finance development, which enhances financing support and resource allocation efficiency. Heterogeneity analyses indicate that the effect is stronger for state-owned enterprises (SOEs), heavily polluting industries, and firms located in eastern and central China. The findings suggest that in the digital economy era, environmental disclosure should be viewed not only as a passive response to legitimacy pressure but also as a competitive tool that helps firms obtain differentiated resources and green finance support.

CONTACT:

[1]Author: zhanglipai1994@126.com

*Corresponding Author: wjh002212@163.com

DOI: 10.64549/jaai-ii.v1i1.53

# 1.Introduction

With environmental protection becoming a basic national policy, corporate environmental information disclosure has become an important part of their social responsibility performance. By providing corporate environmental information to different stakeholders, enterprises can meet the needs of external legitimacy evaluation and avoid potential environmental conflicts and disputes (Zeng et al., 2010; Nason et al., 2018). In May 2022, the Ministry of Ecology and Environment issued the "Reform Plan for the Legally Mandated Environmental Information Disclosure System", which further clarified the important value of corporate environmental disclosure, regarding it as a basic component of the ecological civilization system and related to the modern comprehensive management capacity of enterprises.

However, as an important corporate decision, the willingness to disclose and the quality of disclosed information depend on the comprehensive trade-off and game of managers between external pressures and internal development needs (Song Jianbo & Li Dani, 2013; Li Qiang & Feng Bo, 2015).

(1) From the perspective of external pressure, most enterprises in China lack mandatory and in-depth supervision on environmental information disclosure. Firstly, except for key pollution discharge monitoring units, environmental information disclosure is still a voluntary part, and the form and content of disclosure do not need to be subject to external auditing, which provides enterprises with great discretion in "whether to disclose, what to disclose, and how to disclose". For heavily polluting enterprises that themselves cause negative environmental externalities, "not disclosing or not reporting bad news" has become their "taken-for-granted" choice. Secondly, even if higher-level units have mandatory requirements on environmental information disclosure for some enterprises, if the expected cost of publishing false information is low, enterprises may also engage in "false green" behaviors and deliberately conceal true environmental protection information (Zhu Wei et al., 2019). Based on this, there is a potential problem of "information asymmetry" in corporate environmental information disclosure, and it is urgent to strengthen the environmental supervision system to increase their "legitimacy pressure".

(2) From the perspective of internal needs, most enterprises do not pay enough attention to disclosure work, resulting in low disclosure quality. Firstly, among enterprises participating in environmental information disclosure, the form of disclosure reports is relatively random and

fragmented, key indicators are not fully disclosed, and the disclosure is mainly vague, formalistic and difficult-to-verify qualitative disclosure, lacking effective proprietary environmental information (Song Jianbo & Li Dani, 2013); secondly, some enterprises still remain at the "pressure response" level in environmental information disclosure and have not fully realized its market strategic value (Waddock and Graves, 1997; Dowell et al., 2000). Therefore, in addition to necessary regulatory pressure, the government should consider how to drive the enthusiasm of enterprises for information disclosure, making the release of high-quality information an important means of market competition.

Regarding how to drive enterprises to take the initiative to disclose environmental information and improve the quality of information disclosure, academia and industry have mainly sought solutions from the perspectives of strengthening environmental supervision and increasing punishment intensity (Waddock and Graves, 1997; Wu Jianfeng et al., 2015), but few have conducted analysis from the perspective of market-oriented incentives. In view of the fact that previous studies and national policy documents have confirmed the important significance of digital development for the allocation of market factors, this paper will attempt to explore whether urban-level digital development can affect the local enterprise information disclosure work by influencing the local marketization process.

## 2.Literature review

### *2.1 Visual Exploration in Early Development*

According to the definition in the "Dictionary of Ecological Civilization Construction" (2016), this paper defines corporate environmental information disclosure as "enterprises disclose various environmental information related to the enterprise to stakeholders through various forms such as financial reports, notes to financial statements, prospectuses, social responsibility reports, environmental reports, sustainable development reports, announcements on major events, and corporate websites."

For environmental information disclosure, this paper will elaborate on the research progress from the perspectives of its motivation, influencing factors, and economic consequences.

Existing studies have summarized the motivations of corporate environmental information disclosure into the following two aspects.

Firstly, enterprises must meet the pressure of environmental legitimacy, and disclosing green information helps to show what they have done (Mei Xiaohong et al., 2020). With the strengthening of government environmental regulation and the improvement of public participation in environmental protection, social calls for environmental protection have been strengthened, gradually forming a social atmosphere and concept of "resource conservation" and "environmental friendliness", and the demand for corporate green information has increased accordingly (Sun Xiaohua et al., 2023). While pursuing the maximization of economic interests, enterprises must appropriately take into account the demands of external stakeholders, reduce the negative externalities caused by pollution, and take the initiative to disclose to respond to external information needs in order to achieve legitimate recognition.

Secondly, corporate environmental information disclosure helps enterprises obtain good market competitiveness and a social green image to strive for more differentiated resources. At this stage, ESG information has become an important evaluation item in domestic and foreign capital markets, and enterprises with low ratings will encounter obstacles in IPO, investment and financing, business cooperation and other links (Liliana, 2003). High-quality environmental information disclosure sends positive and responsible value signals to the market, and is more likely to obtain resource inclination from the green financial market, which will enhance consumer confidence and ease the pressure of financing costs. Mei Xiaohong et al. (2020) found that in regions with a high degree of marketization and weak factor distortion, due to the lack of relational resources and insufficient rent-seeking space, enterprises are more encouraged to disclose green information to alleviate information asymmetry and improve market attractiveness. However, objectively, disclosing environmental information requires enterprises to have certain slack resources to undertake environmental governance, and also requires external environment to provide supportive conditions, which is not fully studied in the current research.

Current research covers a wide range of influencing factors. Among internal factors, it involves carbon risk awareness, carbon emissions, carbon emission reductions, environmental protection investment, environmental system certification, manager characteristics, etc. (Yan Haizhou & Chen Baizhu, 2017; Yan Huahong et al., 2019); among external factors, it involves climate change, environmental regulation, environmental inspection, public environmental supervision and demands (He Yu et al., 2014; Yan Haizhou & Chen Baizhu, 2017; Qi Fangfang & Wu Xiaowen, 2023).

However, corporate environmental information disclosure is often the result of the joint action of multiple factors and is complex. The institutional logics generated by these factors may have incompatibility, competition, conflict and other relationships (Wu Danhong et al., 2021). For example, in environmental information disclosure, there may be a phenomenon that "enterprises tend to superficially follow and cater to the political institutional logic, but actually strive to serve the market institutional logic", which is often manifested as: managers perfunctorily fulfill environmental responsibilities and symbolically disclose environmental information in order to save environmental governance costs, leading to the failure of environmental governance work to achieve substantial results.

Based on this, according to the real motives of enterprise management, this paper considers dividing corporate environmental information disclosure into two aspects: the presence or absence of disclosure behavior and disclosure quality: disclosure behavior reflects whether the enterprise has made formal response work to meet the legitimate status; disclosure quality reflects the degree of substantial efforts of the enterprise in the field of environmental governance, and is an important indicator to determine whether the enterprise is "inconsistent in words and deeds".

Regarding the economic consequences of corporate environmental information disclosure, existing literature has found that disclosure work has a promoting value for the medium and long-term development of enterprises, but has a crowding-out effect on short-term economic profits.

For the medium and long-term value effect, relevant literature believes that disclosure has cost reduction, quality improvement and incentive effects: firstly, enterprises' disclosure of positive information on environmental responsibilities can effectively improve equity and debt capital costs, ease financing constraints, and enhance enterprise growth capacity, which is more significant in highly market-oriented regions and developed green financial markets (He Yu et al., 2014; Wu Hongjun et al., 2017); secondly, the green commitments transmitted by environmental information disclosure can internally encourage enterprises to take the initiative to carry out green transformation work, fulfill energy conservation and emission reduction obligations at a higher standard, restrict inefficient resource input, and reduce the uncertainty of environmental protection investment (Yan Haizhou & Chen Baizhu, 2017). Based on this, the voluntary disclosure theory holds that: enterprises will actively and proactively disclose relevant environmental responsibility information when participating in environmental protection investment and achieving good environmental governance results; even under unfavorable conditions, it helps to alleviate the market crash crisis (Sun Xiaohua et al., 2023).

For the short-term effect, relevant literature mainly involves the short-term crowding pressure of disclosure work. This is because improving the quality of environmental information disclosure requires enterprises to continuously perform environmental protection work and carry out substantial green activities. Once an enterprise becomes a pioneer in environmental information disclosure, it is likely to be placed with higher environmental expectations by the outside world, leading to the high cost pressure of "flogging the fast ox" (Li Qiang & Feng Bo, 2017; Kong Dongmin et al., 2021). With the intensification of market competition related to the main business, enterprises will find it difficult to effectively balance green environmental protection responsibilities, resulting in a dilemma. In addition, some studies also believe that with the increasing social attention to corporate ESG strategies, environmental information disclosure will lead to the leakage of enterprise proprietary information, resulting in imitation by competitors, thereby adversely affecting the enterprise's market position (Yang Ruijuan et al., 2020). These all cause certain disturbances to the development of environmental information disclosure work.

There are few current literatures on the impact of urban digital development on corporate environmental information disclosure. Only a small number of literatures have proposed that "because the Internet wide-area platform can quickly obtain, input information, analyze and process data, and feed back and share in a timely manner, the government should fully develop digital public tools to realize the structured management of information disclosure". In addition, from the enterprise perspective, Dai Yue and Shi Mengge (2016) proposed that based on the environmental information needs of different stakeholders, modular design of information disclosure based on "Internet +" should be carried out to enable local enterprises to provide environmental information efficiently and conveniently.

The above research mainly involves the promotion of digital technology on the form renewal and intelligent transmission of environmental information, and has no direct correlation with the willingness and quality of enterprise information disclosure. How the digital development of the region affects the motives and behaviors of enterprise managers, thereby adjusting the environmental information disclosure behavior (disclosure willingness and quality), remains to be further explored.

Based on the above, current research is not sufficient on whether digital development can affect ESG disclosure business related to corporate social responsibility, and whether ESG disclosure business and economic interests can produce synergy and complementarity. Enterprises are essentially profit-making entities, and participating in ESG projects such as environmental governance is not their "original intention"; excessive participation will instead have a crowding-out effect on profits. However, if participating in and disclosing ESG projects can be "linked" to economic business and promote the inflow of external resources and policy inclination, it will effectively change the mentality of managers.

Among them, the green financial market is a typical application scenario connecting social responsibility and economic interests. The improvement of this green factor market is an important prerequisite for enterprises to obtain competitive advantages by performing and disclosing environmental responsibilities: due to information asymmetry and environmental uncertainty, buyers are often in an information disadvantage; at this time, by disclosing environmental and other social responsibility information, enterprises can enhance effective communication with investors and obtain organizational reputation and non-relational resources in a way that is more acceptable to stakeholders (Ren Guangqian, 2017). Whether urban digital development can effectively empower new infrastructure such as the green financial market and ultimately promote the enthusiasm of enterprises for green information disclosure remains to be further studied.

### *2.2 Data Collection*

Disclosure decisions need to balance economic benefits and social responsibilities: the former follows the logic of market competition, while the latter follows institutional logic (Yin Juelin, 2012; Gou Qianwen et al., 2019). Under the constraint of multiple logics, managers may superficially conform to relevant systems to reduce frictions, resulting in certain symbolic rather than substantive behaviors, thereby reducing environmental governance costs. Essentially, this approach serves the logic of market competition.

Based on the above, this study divides corporate environmental information disclosure into two parts: the presence or absence of disclosure behavior and disclosure quality. Among them, the former reflects whether an enterprise formally responds to external demands for disclosing environmental information; the latter is used to measure the authenticity, completeness, timeliness and other aspects of the disclosed information, reflecting the extent to which the enterprise substantially fulfills its environmental responsibilities (Hu Wenxiu et al., 2021). This paper will empirically test the impact of urban digital development on local corporate environmental information disclosure and consider how to regulate corporate disclosure behaviors.

#### *2.2.1 The Impact of Urban Digital Development on Corporate Environmental Information*

***Disclosure Behavior***

Urban digital development strengthens the intensity of formal and informal environmental regulations in the region, promotes the spillover and diffusion of green environmental protection knowledge, and enhances local environmental protection awareness. According to the legitimacy theory, these increasing external demanders of environmental information shape the institutional environment and environmental public contracts of the region, forming pressure for legitimate recognition on local corporate environmental governance (Zhu Wei et al., 2019). Under relevant institutional pressures, organizations have to gradually converge with mainstream values at least in external form and appearance, and take the initiative to assume certain environmental responsibilities (Liu Yuhuan et al., 2020). "Even good wine fears deep alleys"—once an enterprise fulfills its environmental responsibilities, it must adopt a relatively public way to make the public aware of the enterprise's relevant situation, which involves the issue of publicizing the enterprise's private information (Shen Yi et al., 2014).

The legitimacy theory holds that enterprises need to actively, appropriately and selectively disclose environmental information to prove their value to society, show that they fulfill social and environmental contracts, guide social recognition of their legitimacy, and avoid potential legal lawsuits (Tang Guoping & Li Longhui, 2011; Xiao Hongjun et al., 2013). Essentially, this disclosure behavior is a self-interested defensive tendency to maintain their own legitimate recognition. Therefore, Hypothesis 1 is proposed.

Hypothesis 1: Controlling for other factors, urban digital development can significantly increase the environmental information disclosure behavior of local enterprises.

***2.2.2 The Impact of Urban Digital Development on the Quality of Corporate Environmental Information Disclosure***

Urban digital development effectively alleviates the problem of information asymmetry by strengthening external environmental supervision, forming legitimate pressure on local corporate environmental governance. This urges enterprise managers to accept external supervision and strengthen disclosure behaviors to respond to external environmental demands. However, even if environmental disclosure obligations are fulfilled, whether the disclosure quality is significantly improved still needs further exploration. At present, except for key monitoring units, domestic environmental protection departments still lack specific regulations on the detailed items disclosed by enterprises, and enterprises still have considerable "discretion" in environmental information disclosure, leading to uneven quality of disclosed information governance (Xiao Shuguang et al., 2017).

Driving enterprise managers to disclose high-quality information requires them to fully recognize the market value of environmental information disclosure and change the past concept of passive response.

First, Market Competition Channel. The impact path of urban digital development on the quality of corporate environmental information disclosure may first be reflected in its "market competition effect". In regions with slow marketization, factor markets are distorted, and relational resources occupy a more important position. At this time, enterprises find it difficult to obtain corresponding factor resources entirely through market-oriented means, and the role of disclosure quality in resource allocation is weakened (Quairel, 2011). In contrast, digital development promotes the free flow, spillover and reallocation of local market factors, improves the level of regional marketization, overcomes resource misallocation and market distortion, improves the development level of various markets, factor allocation and legal systems to be more sound and fair, matures the "contract economy", reduces relational resources, and intensifies market competition. Under the logic and behavioral orientation of marketization, obtaining factors and building competitive advantages require enterprises to strive through market-oriented means (Wang Yaming et al., 2015). This will encourage enterprises to incorporate CSR into their marketing plans, actively improve the quality of environmental information disclosure, and transmit a differentiated "signal strategy" to gain reputation and competitiveness and enhance external confidence.

The underlying logic of the above analysis is: a developed factor market is a necessary prerequisite for enterprises to obtain differential advantages by fulfilling and disclosing green responsibilities (Yan Haizhou & Chen Baizhu, 2017). Rational market investors usually judge the direction of an enterprise's business strategy from the social responsibility information disclosed by the enterprise, identify more positive corporate value signals, reduce investment doubts, and thus continuously and dynamically adjust their investment decisions.

Second, Market Investment and Financing Channel. The impact mechanism of urban digital development on the quality of corporate environmental information disclosure may second be reflected in its "resource matching optimization and financing support" in the field of green finance. Green finance closely links green responsibilities with economic interests, guiding and encouraging enterprises to establish green concepts and promote green transformation. Among them, environmental information disclosure is an important foundation for the development of green finance: the development of green finance requires the establishment of a more open, transparent and comprehensive corporate environmental information disclosure system.

Specifically, China is currently in a transitional economy and an emerging marketization stage, with fierce market competition and strong corporate financing demand (Du Jinzhu & Wu Zhanyong, 2022). As the market supplier of green financial services, financial institutions need to identify green enterprises through corporate environmental information disclosure. Enterprises that disclose environmental information in a timely and active manner are more likely to obtain financial support with relatively low interest rates and relatively long loan terms (Wang et al., 2020; Hu Wenxiu et al., 2021).

With the rapid digital development of the region, digital technology can form a networked and open ecological information sharing mechanism, providing an upstream and downstream industrial link platform for green finance. While ensuring the safe, effective and free flow of information, this digital platform can improve the matching efficiency of green finance investment and financing on both the supply and demand sides, and alleviate the blindness of market regulation (Vives, 2017; Chen Hua & Shen Yue, 2022). In order to better obtain green financial services and reduce environmental governance costs, enterprises will take the initiative to improve disclosure quality in a sound green financial market environment to gain more market attention and policy inclination.

Based on this, the following hypotheses are proposed.

Hypothesis 2: Controlling for other factors, urban digital development can significantly improve the quality of environmental information disclosure of local enterprises.

Hypothesis 2-1: Controlling for other factors, urban digital development can encourage local enterprises to improve the quality of environmental information disclosure by promoting the process of regional marketization.

Hypothesis 2-2: Controlling for other factors, urban digital development can provide technical support for the development of local green finance, thereby significantly improving the quality of environmental information disclosure of local enterprises.

#### *2.2.3 Variable Selection*

The selection of variable indicators in this paper follows the principles of relevance, accessibility, data openness and accuracy.

Explained Variables: Corporate Environmental Information Disclosure Behavior and Environmental Information Disclosure Quality

Disclosing an enterprise's environmental management behaviors, purposes, scope and nature can meet the external demand for enterprise information at the level of sustainable development. Essentially, such disclosure belongs to impression management, which balances the enterprise's environmental governance effects and economic performance, and will have an impact on enterprise value and market competitive position.

Previous literature has proposed that environmental information disclosure has a social responsibility decoupling phenomenon of "saying one thing and doing another", that is, although an enterprise has disclosure behaviors, such behaviors are only perfunctory symbolic responses, and the enterprise has not taken substantial environmental protection actions due to economic considerations. This will lead to insufficient disclosure quality and lack of referable effective information. To avoid the disturbance of this "inconsistency between words and deeds" on the research, this paper divides environmental information disclosure into two parts: disclosure behavior and disclosure quality, and sets proxy variables in turn for regression testing.

Selection of Measurement Indicators for Corporate Environmental Information Disclosure Behavior: Corporate environmental information disclosure behavior refers to the enterprise's transmission of its own environmental responsibility information to the public through various information reporting carriers or channels in response to the environmental responsibility pressure in the region, so as to guide social recognition of the legitimacy of its own environmental behaviors. For the measurement of corporate environmental disclosure behavior, a dummy variable is constructed based on whether the enterprise discloses environmental information in the current year (1 if yes, 0 otherwise). This study judges whether the enterprise has disclosure behavior according to whether it has disclosed environmental information in the annual report, social responsibility report and environmental report in the current year (relevant data are from the CSMAR Environmental Governance Series). This behavior reflects the enterprise's willingness and attention to environmental information disclosure work, and is an important way to formally meet external requirements.

Selection of Measurement Indicators for Corporate Environmental Information Disclosure Quality:Corporate environmental information disclosure quality represents the detail level of the enterprise's disclosure in environmental disclosure work. If an enterprise only completes the disclosure work in form, but the quality of environmental information disclosure does not improve or even decreases, then the enterprise' s corresponding disclosure work is likely to be only symbolic disclosure for perfunctory purposes, lacking verifiable, quantifiable and detailed environmental actions.

The practices of existing literature on measuring disclosure quality are as follows. Firstly, Li Qiang and Feng Bo (2015) divided environmental information into 7 parts, including environmental costs, environmental liabilities, environmental performance and environmental governance, and summed the scores of each part to generate a disclosure quality index; secondly, Zheng Jianming and Xu Chenxi (2018) manually selected two dimensions, namely empirical and formal dimensions, of corporate environmental information disclosure, and adopted content analysis method to quantitatively evaluate the quality of corporate environmental information disclosure; thirdly, Kong Dongmin et al. (2021) used the Environmental Research Database in the CSMAR Database to classify environmental information according to whether it is monetized information (Wiseman, 1982): for monetized information, the assignment is 2 for combined quantitative and qualitative disclosure, 1 for qualitative indicators, and 0 for undisclosed indicators; for non-monetized information, the assignment is 2 for disclosed indicators and 0 for undisclosed indicators. These measurement methods all require a certain amount of time for effective classification, with high information collection costs and a certain degree of subjectivity.

Based on the considerations of low cost, easy accessibility, objectivity and reliability of data collection, the data of this study are derived from the Bloomberg Corporate Social Responsibility Disclosure Index. This index involves the total ESG score of the enterprise and the scores of each individual item. Among them, the total ESG score can be traced back to 2006. By measuring the company' s performance in environmental, social and governance dimensions, the organization gives each company a score in three dimensions: environmental information (E), social information (S) and corporate governance (G), which are summed to form the total ESG score, ranging from 0.1 to 100. The source of score information is mainly the public information disclosed by the enterprise (such as climate change issues and pollution emission issues, which are also of financial importance and business relevance). The rating indicators involve 21 secondary indicators and 70 evaluation indicators to assist investors in evaluating the enterprise' s ESG governance capabilities. This paper takes the total score of environmental information (E) among them as the environmental disclosure quality score. The higher the score, the better the ESG performance of the enterprise in the environmental field; enterprises that do not obtain a score in the current year are assigned a value of 0.

Explanatory Variable: Urban Digital Development

Since the "12th Five-Year Plan", China's digital economy has entered a period of rapid development, and a series of macro-level regional digital measurement indicators have gradually emerged. Current domestic and foreign literature on the measurement of the digital economy mainly comes from the national and provincial levels, while the measurement at the urban level is relatively few (Zhang Bochao and Shen Kaiyan, 2018; Xu Xianchun and Zhang Meihui, 2020). Based on the principles of data accessibility, effectiveness and reliability, the core explanatory variable of this paper is: urban-level digital development, which is more detailed and specific than national and provincial-level indicators.

The level of urban digital development includes both the overall index level, as well as the breadth of digital coverage, depth of use and literacy level. Therefore, this paper measures the level of urban digital development (Dig) from four dimensions: digital foundation, digital input, digital literacy and digital application, so as to better reflect the digital stage of the urban level.

Due to the possible dimensional differences of the above indicators, on the basis of fully drawing on and absorbing the three-level indicator design system of Zhao Tao et al. (2020), this study standardizes all three-level indicators belonging to different dimensions to eliminate dimensional differences; secondly, this study uses the coefficient of variation method to construct a digital development evaluation system including 4 primary indicators, 5 secondary indicators and 6 tertiary indicators. The constructed indicator system is cutting-edge, in line with the characteristics of the digital and information age, and all indicators are positive indicators with accessibility and operability.

In the calculation process, this study assigns weights to each three-level indicator in turn according to the proportion of the coefficient of variation of each three-level indicator in the total coefficient of variation of all three-level indicators (the coefficient of variation is calculated by standard deviation/average value), and weights and summarizes each three-level indicator to calculate the final digital development score of a specific city, with the corresponding score ranging between [0, 1].

*Table 1 Digital Development Indicator System*

| Primary Indicators | Secondary Indicators | Tertiary Indicators | Weight of Tertiary Indicators | Indicator Attribute |
|---|---|---|---|---|
| Digital Foundation | Broadband Access Level | Number of Regional International Internet Broadband Access Users | 0.1237 | + |
| | Communication Access Level | Number of Regional Mobile Phone Users at the End of the Year | 0.1292 | + |
| | | Total Regional Telecommunications Business Revenue | 0.1893 | + |
| Digital Application | Digital Popularization Degree | Regional Digital Inclusive Finance Index | 0.068 | + |
| Digital Input | Technology R&D Input | Total Annual Regional Scientific and Technological Expenditure | 0.2251 | + |
| Digital Literacy | Human Resource Level | Number of Employees in Regional Information Transmission, Computer Services and Software Industry | 0.2647 | + |

Selection of Control Variables: Based on existing literature (Corr, 1995; Berman et al., 2021; Lee, 2021; Francis, 2021), this paper selects control variables from the following two levels.

Enterprise Level: Disclosure behavior and disclosure quality are closely related to the enterprise's financial status, operating results, property right nature, internal control and management, which directly affect the enterprise's disclosure decisions; in the context of digital development promoting the regional marketization process and market competition, the enterprise's R&D investment and innovation level directly affect its risk resistance capacity. Therefore, this paper selects: enterprise performance level (operating profit rate, operating profit/operating income); enterprise solvency (asset-liability ratio, total liabilities/total assets); enterprise operational capacity (fixed asset ratio, ending fixed assets/total assets); enterprise value level (Tobin's Q value, enterprise market value/total assets); enterprise governance level (proportion of independent directors on the board of directors, number of independent directors/number of board directors; number of board directors; management expense rate, management expenses/main business income; ownership concentration, sum of squares of shareholding ratios of the top ten major shareholders of the company); enterprise innovation level (R&D expense rate, annual R&D expenses/annual operating income); enterprise property right nature (1 if it is a state-owned enterprise, 0 otherwise).

Regional Level: This paper selects: regional marketization degree (collected from Fan Gang's Regional Marketization Index); regional government's environmental regulation degree (number of words in sentences containing environmental protection keywords/total number of words in the regional government work report); regional public environmental supervision degree (in Baidu Index trend analysis, search with "environmental pollution" as the keyword, and the logarithmically processed annual average value of the region).

Finally, industry, annual and individual fixed effects are controlled to overcome the impact of unobservable heterogeneity on endogeneity. The model is constructed with enterprise individuals as units for clustering, and clustered robust standard errors are calculated.

*Table 2 Variable Definition Table*

| Symbol | Indicator | Measurement Method |
|---|---|---|
| ED | Environmental Disclosure Behavior | Dummy variable, 1 if the enterprise has environmental information disclosure in the current year, 0 otherwise. See 5.3.1 for details. |
| EDQ | Environmental Information Disclosure Quality | Selected from the environmental score item of the Bloomberg Database. See 5.3.1 for details. |
| Dig | Urban Digital Development | Referring to Zhao Tao et al. (2020), the weighted total score of 6 sub-indicators is calculated by using the objective weighting method. See 4.3.1 for details. |
| Profit | Enterprise Performance Level | Operating profit rate, operating profit/operating income |
| Leverage | Enterprise Solvency | Asset-liability ratio, total liabilities/total assets |
| FA | Enterprise Operational Capacity | Fixed asset ratio, ending total fixed assets/total assets |
| Tobin Q | Enterprise Value | Tobin's Q value, enterprise market value/total assets |
| Ind | Enterprise Governance Level | Proportion of independent directors, number of independent directors/number of board directors |
| Board | | Number of board directors |
| AC | | Management expense rate, management expenses/main business income |
| HHI_10 | | Ownership concentration, sum of squares of shareholding ratios of the top ten major shareholders of the company |
| R&D | Enterprise Innovation Level | R&D expense rate, annual R&D expenses/annual operating income |
| Dummy_State | Enterprise Property Right Nature | Dummy variable, 1 if it is a state-owned enterprise, 0 otherwise |
| Marketization | Regional Marketization Degree | Fan Gang's Marketization Index |

| Symbol | Indicator | Measurement Method |
|---|---|---|
| ER | Regional Government Environmental Regulation Degree | Number of words in sentences containing environmental protection keywords/total number of words in the regional government work report |
| Pub | Regional Public Environmental Supervision Degree | In Baidu Index trend search, input "environmental pollution" as the keyword, and logarithmically process the annual average value of the region |

### *2.2.4 Data Collection and Preprocessing*

This study collects listed companies on China's Shanghai and Shenzhen A-shares from 2011 to 2019 as the research population. It is sorted out according to the annual reports of Shanghai and Shenzhen listed companies, and the regional and enterprise indicator datasets are effectively merged using STATA 16 software. In the period of 2011-2019, to ensure the reliability of sample selection, some samples are excluded according to the following principles: ST and *ST companies; companies with missing variable indicator data; companies with negative total assets; insolvent companies; sample companies in Tibet. After effectively merging different datasets, a total of 3,051 listed companies and 19,096 "enterprise-year" observation samples are obtained.

The explained variables, explanatory variables and control variables are from different databases. Among them: corporate environmental information disclosure is collected from Bloomberg ESG disclosure scores, and the environmental sub-item score is selected as the proxy indicator; internal control, property right nature and financial statement indicators are collected from the Company Research Series of the CSMAR Database; among the relevant data on digital development, the regional inclusive finance data adopts the Digital Inclusive Finance Index compiled by Peking University; the number of broadband access users, mobile phone users, total telecommunications business revenue, total scientific expenditure, and the number of employees in information transmission, computer services and software industry are taken from the annual data of the National Bureau of Statistics and the "China Urban Statistical Yearbook"; environmental regulation intensity data are from the "China Environmental Yearbook", "China Urban Statistical Yearbook" and "China Statistical Yearbook"; public environmental supervision data are from Baidu Index search; for the measurement of marketization process, the disclosure data of Wang Xiaolu and Fan Gang in the annual "China Marketization Index—Report on the Relative Process of Marketization in Various Regions" are adopted; for the regional green finance index, the data are from the "China Statistical Yearbook", "China Insurance Yearbook" and statistical yearbooks of various provinces.

All continuous variables are subjected to WINSORIZE treatment at the 1% and 99% quantiles.

***2.2.5 Theoretical and Empirical Model Construction***

Based on theoretical analysis and hypotheses, this paper has constructed a theoretical research model of "urban digital development—corporate environmental information disclosure", as shown in Figure 1.

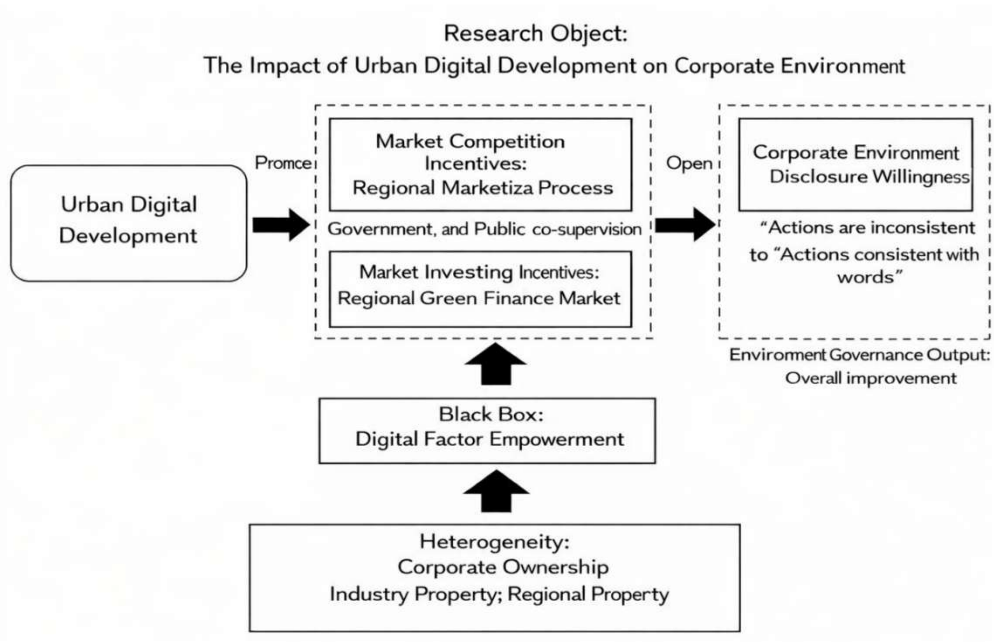


*Figure 1 Theoretical Model of "Urban Digital Development—Corporate Environmental Information Disclosure"*

For the relationship between urban digital development and corporate environmental information disclosure behavior, a high-dimensional panel fixed effect model is constructed to verify the hypothesis. The benchmark model will control for annual, industry or firm fixed effects.

$$ED_{i,t}=\alpha+\beta_1 * Dig_{i,t}+Controls+Fixed\ effects+\varepsilon_{i,t}$$

## 3.Experiments and Results

### *3.1 Descriptive Statistics*

This section first conducts a descriptive analysis of the selected variables, and the results are shown in Table 3.

For the presence or absence of corporate environmental disclosure behavior (ED) within the current year, a higher mean value of the corresponding indicator indicates a higher probability of disclosure behavior - the mean value of the corresponding indicator ED is 0.323, the standard error is only 0.467, and the median is 0, showing a significant right-skewed distribution. This means that more than half of the enterprise samples lack the behavior of disclosing environmental information.

Correspondingly, the overall quality of corporate environmental information disclosure is relatively low. With a maximum value of 65.625 and an upper quartile of 6.9767, the mean value is only 3.516 and the standard deviation is 6.762, indicating that there are significant differences in the quality of environmental disclosure among enterprises, and there are significant differences in the choices of enterprises on how to disclose environmental information.

In the control variable group: the corporate financial leverage level, operating profit rate, management cost and asset liquidity structure are relatively normal, and the differences of these indicators among enterprises are relatively insignificant; the corporate ownership concentration is relatively low, with a maximum value of only 0.561 and a mean value of 0.166, indicating that shareholders in most enterprises have a certain degree of mutual checks and balances, which will strengthen the restraint on the self-interested behaviors of the largest shareholder and managers, and alleviate the disclosure distortion under information asymmetry; for the marketization level of the enterprise's location, the mean value of the relevant indicator is 8.431 (median is 8.89), the maximum value is 10.96, and its first quartile value is 7.09, showing that the marketization process in most regions of China is relatively fast, and a certain height of market economy construction has been achieved; in the current external environmental supervision, both the government and the public have formed a certain environmental binding force on enterprises, with the mean values of ER and Pub being 0.065 and 4.865 respectively.

*Table 3 Descriptive Statistics*

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| EDQ | 21622 | 3.516 | 6.762 | 0 | 65.625 |
| Dig | 21616 | .239 | .204 | .019 | .74 |
| ED | 21622 | .323 | .467 | 0 | 1 |
| Leverage | 21622 | .428 | .217 | .007 | .999 |
| Profit | 21617 | .098 | .197 | -1.234 | .732 |
| AC | 21314 | .098 | .09 | .003 | .729 |
| Tobin Q | 20875 | 1.995 | 1.304 | .876 | 9.833 |
| Board | 21612 | 8.678 | 1.819 | 5 | 15 |
| FA | 21622 | .202 | .163 | .002 | .706 |
| Ind | 21610 | .375 | .053 | .333 | .571 |
| Marketization | 21622 | 8.431 | 1.78 | 3.59 | 10.96 |
| ER | 21622 | .065 | .02 | .032 | .125 |
| Pub | 21622 | 4.865 | .332 | 3.784 | 5.371 |
| HHI_10 | 21620 | .166 | .116 | .015 | .561 |
| Dummy_State | 21106 | .398 | .49 | 0 | 1 |

### *3.2 Correlation Analysis*

In Table 4, this paper conducts a correlation analysis between core variables based on the Pearson two-tailed test. At the 1% significance level, digital development (Dig) is significantly positively correlated with corporate environmental disclosure behavior (ED), with a correlation coefficient of 0.075. As analyzed earlier, urban digital development may promote the latter by alleviating information asymmetry and increasing external supervision.

Similarly, at the 1% significance level, digital development (Dig) is significantly positively correlated with the quality of corporate environmental information disclosure (EDQ), with a correlation coefficient of 0.119. This may mean that regional digital factors can improve the effective information content of corporate information disclosure and alleviate the phenomenon of low-quality disclosure.

In the mechanism indicator group, the pairwise correlation coefficients between indicators are all lower than 0.8, indicating that there is no strong correlation (multicollinearity). Among them, at the 1% significance level, the correlation coefficient between public environmental supervision (Pub) and marketization level (Marketization) reaches 0.701, which may indicate that the marketization process provides more circulating information and resources for public environmental participation, and creates a better environment for public opinion supervision and material infrastructure for exercising rights.

*Table 4 Correlation Analysi*

| **Variables** | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** |
|---|---|---|---|---|---|---|
| (1) EDQ | 1.000 | | | | | |
| (2) ED | 0.766*** | 1.000 | | | | |
| (3) Dig | 0.119*** | 0.075*** | 1.000 | | | |
| (4) Marketization | 0.015*** | -0.029*** | 0.530*** | 1.000 | | |
| (5) ER | -0.022*** | -0.005 | -0.187*** | 0.117*** | 1.000 | |
| (6) Pub | 0.021*** | -0.018*** | 0.334*** | 0.701*** | 0.208*** | 1.000 |

****** *p<0.01*，*** *p<0.05*，* *p<0.1*

### *3.3 Benchmark Regression Results*

Regarding the relationship between regional digital development and corporate environmental disclosure, this paper conducts regression analysis from two dimensions: the presence or absence of disclosure behavior and the quality of disclosure, so as to verify the research hypotheses proposed earlier.

#### *3.3.1 Regression Results of the Presence or Absence of Environmental Disclosure Behavior*

According to Hypothesis 1 (regional digital development significantly promotes corporate environmental disclosure behavior), this paper first tests the relationship between urban digital development (Dig) and corporate environmental disclosure behavior (ED). The regression results are shown in Table 5: at the 1% significance level, digital development (Dig) is significantly positively correlated with corporate environmental disclosure behavior (ED), with a regression coefficient of 0.166 and a standard error of 0.0469. This result verifies Hypothesis 1, indicating that the improvement of regional digitalization level can effectively promote enterprises to carry out environmental disclosure activities.

In terms of control variables, corporate environmental disclosure behavior, as an important part of corporate governance decisions, is jointly affected by internal and external factors: corporate leverage (Leverage) and profitability (Profit) reflecting financial status are significantly positively correlated with ED at the 1% level, indicating that enterprises with better financial conditions are more willing to bear the cost of environmental disclosure and convey positive signals to the outside world; management expense rate (AC) is significantly negatively correlated with ED at the 5% level, which may be due to the fact that enterprises with higher management costs have limited resource allocation capacity and are unwilling to invest additional resources in environmental disclosure; the number of board members (Board) and the proportion of independent directors (Ind) are significantly positively correlated with ED at the 1% level, indicating that a sound corporate governance structure can strengthen the supervision of management and promote enterprises to fulfill their environmental disclosure obligations; government environmental regulation (ER) is significantly positively correlated with ED at the 5% level, reflecting the binding effect of external institutional pressure on corporate behavior; the nature of state-owned property rights (Dummy_State) is significantly positively correlated with ED at the 1% level, which is because state-owned enterprises bear more social responsibilities and are more responsive to policy guidance. After controlling for year and industry fixed effects, the $R^2$ of the benchmark model is 0.190, indicating that the model has a certain explanatory power for corporate environmental disclosure behavior.

*Table 5 Benchmark Regression Results (I)*

| | (1)Full Sample |
|---|---|
| **VARIABLES** | **ED** |
| Dig | 0.166*** |
| | (0.0469) |
| Leverage | 0.350*** |
| | (0.0411) |
| Profit | 0.209*** |
| | (0.0290) |
| AC | -0.202** |
| | (0.0875) |
| Tobin Q | -0.00641 |
| | (0.00503) |
| Board | 0.0413*** |
| | (0.00522) |
| FA | 0.0643 |
| | (0.0587) |
| Ind | 0.801*** |
| | (0.147) |
| Marketization | 0.000767 |
| | (0.00733) |
| ER | 0.566** |
| | (0.284) |
| Pub | -0.0104 |
| | (0.0450) |
| HHI_10 | 0.130* |
| | (0.0695) |
| Dummy_State | 0.137*** |

| | (1)Full Sample |
|---|---|
| **VARIABLES** | **ED** |
| | (0.0202) |
| Constant | -0.592*** |
| | (0.209) |
| Year Fixed Effect | Yes |
| Industry Fixed Effect | Yes |
| Observations | 20,160 |
| R2 | 0.190 |

*Robust standard errors in parentheses and cluster at firm level； *** p<0.01， ** p<0.05， * p<0.1*

### *3.3.2 Regression Results of Environmental Disclosure Quality*

Hypothesis 2 proposes that regional digital development can significantly improve the quality of corporate environmental disclosure. To verify this hypothesis, this paper takes EDQ as the dependent variable and conducts benchmark regression again, with the results shown in Table 6. It can be seen from the table that digital development (Dig) is significantly positively correlated with the quality of corporate environmental disclosure (EDQ) at the 1% significance level, with a regression coefficient of 3.563 and a standard error of 0.668. This result verifies Hypothesis 2, indicating that regional digitalization not only promotes enterprises to carry out environmental disclosure activities, but also effectively improves the quality of disclosure, making corporate environmental information more comprehensive, accurate and useful.

The survival and development of enterprises are inseparable from economic benefits. Economic performance pressure tends to increase short-sighted behavior, making environmental governance difficult to be placed in a priority strategic position (Barcena and Espinosa, 1996; Hu Nan et al., 2021). In the work of environmental information disclosure, will enterprises disclose environmental responsibility information in a detailed and responsible manner, or only respond to external demands and gain recognition formally?

If enterprises focus on the former, they will actively disclose more high-quality information, display substantive content such as corporate environmental governance and performance, distinguish themselves from other enterprises with proprietary environmental information, and strive for the favor of the green financial market; if enterprises only respond to external environmental demands and alleviate public pressure, the disclosure behavior is essentially a decoupling behavior of "saying one thing and doing another", and enterprises will only conduct more symbolic disclosure rather than substantive disclosure of what they have done (Gou Qianwen et al., 2019; Zhao Xiaoyue et al., 2022).

This response method of "talking better than doing" solves the "urgent need" to a certain extent, and balances economic efficiency and legal pressure. It seems to be a more efficient, rational and convenient response method (Hemingway and Maclagan, 2004). However, the empirical results above support the former: regional digital development promotes enterprises to carry out substantive environmental disclosure and improve disclosure quality.

In terms of control variables, corporate profitability (Profit) is significantly positively correlated with EDQ at the 1% level, with a regression coefficient of 1.937, indicating that good financial performance provides sufficient financial support for enterprises to improve the quality of environmental disclosure, which is consistent with the research conclusions of Liu Yuhuan et al. (2020); corporate leverage (Leverage) is significantly positively correlated with EDQ at the 1% level, which may be because enterprises with higher leverage need to convey positive environmental signals to creditors to reduce financing costs; management expense rate (AC) and Tobin Q are significantly negatively correlated with EDQ at the 1% level, indicating that enterprises with higher management costs or overvalued market value are less willing to invest resources in improving environmental disclosure quality; the number of board members (Board) is significantly positively correlated with EDQ, reflecting the positive role of corporate governance structure in promoting high-quality environmental disclosure.

*Table 6 Benchmark Regression Results (II)*

| | **(1)** |
|---|---|
| **VARIABLES** | **EDQ** |
| Dig | 3.563*** |
| | (0.668) |
| Leverage | 4.289*** |

| | (1) |
|---|---|
| **VARIABLES** | **EDQ** |
| | (0.579) |
| Profit | 1.937*** |
| | (0.342) |
| AC | -2.603*** |
| | (0.894) |
| Tobin Q | -0.193*** |
| | (0.0584) |
| Board | 0.538*** |
| | (0.0774) |
| FA | 1.539* |
| | (0.832) |
| Ind | 12.83*** |
| | (2.098) |
| Marketization | 0.0110 |
| | (0.106) |
| ER | 2.068 |
| | (3.770) |
| Pub | 0.723 |
| | (0.646) |
| HHI_10 | 4.183*** |
| | (1.197) |
| Dummy_State | 1.963*** |
| | (0.272) |
| Constant | -13.73*** |
| | (2.987) |
| Year Fixed Effect | Yes |
| Industry Fixed Effect | Yes |

|  | (1) |
|---|---|
| **VARIABLES** | **EDQ** |
| Firm Fixed Effect | No |
| Observations | 20,160 |
| R2 | 0.210 |

*Robust standard errors in parentheses and cluster at firm level； *** p<0.01， ** p<0.05， * p<0.1*

### *3.4 Mechanism Test Results*

Before elaborating on the mechanism of action, this paper needs to illustrate that: since firm fixed effects and industry fixed effects belong to the micro and meso dimensional levels, it is inappropriate to conduct regression tests on environmental regulation which belongs to the macro dimensional level; therefore, this paper conducts regression tests while controlling for regional-level variables as well as annual and urban heterogeneity.

#### *3.4.1 Marketization Process Mechanism*

Based on the signal transmission theory, the impact mechanism of urban digital development on the quality of environmental information disclosure of local enterprises is first reflected in the change of the marketization process. With the in-depth development of urban digitalization, digital factors are effectively integrated into urban market-oriented operations, guiding the optimal allocation of resources in accordance with market logic, government intervention and relational resources are diluted, and the relevant legal systems, intermediary organizations, and infrastructure are also more improved (Wang Yaming et al., 2015). On this basis, the effect of signal transmission is more obvious: more enterprises in the region will take the initiative to disclose value-based information to gain competitive advantages, and market investors will also judge truly valuable enterprises from the information disclosure.

The regression results are shown in Table 7. In Column (1), the coefficient of Dig is more significant both in economic and statistical significance under the high marketization group. Hypothesis 2-1 is verified here.

This paper verifies that: the positive effect of urban digital development on corporate environmental information disclosure needs to be realized by optimizing the allocation of factor resources and improving market competitiveness. A sound factor market is a necessary prerequisite for enterprises to obtain differential advantages by disclosing green responsibilities.

*Table 7 regression results(I)*

| | （1） High Marketization | （2）Low Marketization |
|---|---|---|
| **VARIABLES** | **EDQ** | **EDQ** |
| Dig | 3.114*** | 3.003* |
| | (0.713) | （1.565） |
| Year Fixed Effect | Yes | Yes |
| Industry Fixed Effect | Yes | Yes |
| Firm Fixed Effect | No | No |
| City Fixed Effect | No | No |
| Observations | 10294 | 9866 |
| R2 | 0.282 | 0.188 |

*Robust standard errors in parentheses and cluster at city level； *** p<0.01， ** p<0.05， * p<0.1*

### *3.4.2 Green Finance Incentive Mechanism*

The impact mechanism of digital development on corporate environmental information disclosure is also reflected in the renewal and improvement of the green financial market.

This paper holds that the establishment of a green financial system will increase redundant financial resources and alleviate the cost of corporate environmental governance. From the perspective of the global investment and financing market, enterprises with good green responsibility information will more smoothly pass the environmental verification for refinancing, and the "greenwashing" motivation of enterprises due to insufficient funds will also be reduced (Hu Wenxiu et al., 2021). However, in an effective green financial market, liquid financial capital needs to identify or distinguish behaviors such as "true green", "light green" or "non-green" through environmental information (Yan Haizhou and Chen Baizhu, 2017; Wang Yulin and Zhou Yahong, 2022).

On the basis of existing literature, this paper argues that: with the in-depth development of regional digitalization, the information asymmetry between markets and enterprises in the region will be effectively alleviated, the free flow of environmental resource information and the green supply-demand matching mechanism will be more reasonable and efficient. Enterprises with good environmental information disclosure are more likely to attract the attention and favor of the green financial market and be given more green financial support. This will form a continuous "green market incentive" for corporate environmental information disclosure, and enterprises will have more motivation to fulfill their environmental obligations, disclose more detailed, accurate and proprietary environmental information, and shape a green image to gain recognition from market investors.

The regression results are shown in Table 8: urban digital development can significantly promote the local green financial market, verifying Hypothesis 2-2 here.

Based on this, this paper has tested the optimal allocation function of digital factors in price discovery and supply-demand matching in the green finance scenario. Under green finance incentives, the cost pressure faced by environmental information disclosure (Li Qiang and Feng Bo, 2015) is reduced, and the value signal function is more pure. Environmental information disclosure can gradually be regarded as a competitive tool.

*Table 8 regression results(II)*

| | **（1）Regions with High Green Finance Level** | **（2）Regions with Low Green Finance Level** |
|---|---|---|
| **VARIABLES** | **EDQ** | **EDQ** |
| Dig | 3.043*** | 2.066 |
| | (0.734) | （1.700） |
| Year Fixed Effect | Yes | Yes |
| Industry Fixed Effect | No | Yes |
| Observations | 10091 | 10065 |
| R2 | 0.288 | 0.181 |

*Robust standard errors in parentheses and cluster at city level；*** p<0.01，** p<0.05，* p<0.1*

In summary, urban digital development can simultaneously improve the behavior and quality of corporate environmental information disclosure. In the digital economy era, the phenomenon of information asymmetry has been significantly alleviated, hidden pollution behaviors will face more transparent external supervision pressure, and the probability of enterprises being discovered and punished for "greenwashing" and "false green" has increased; in addition, if enterprises hope to obtain competitive advantages and financing support through environmental disclosure, they also need to improve the disclosure quality to a certain level, otherwise low-quality disclosure will be difficult to effectively exert its image effect and mitigation effect (Sun Zhihui et al., 2017; Zhan Hua, 2021). Enterprises should attach importance to environmental disclosure work and incorporate it into their development strategies, rather than staying at "symbolic disclosure".

### *3.5 Heterogeneity Test Results*

#### *3.5.1 Nature of Corporate Ownership*

This paper holds that: urban digital development will continuously shape a fair and effective "contract economy" in the region, compressing the space for rent-seeking, and this effect is more obvious in state-owned enterprises.

Relevant empirical results provide empirical support. As shown in the first two columns of Table 9, in the state-owned enterprise group, the marginal coefficient of digital development is 5.773, which is significant at the 1% probability level; while in the non-state-owned enterprise group, the marginal coefficient of digital development is 1.276 and significant at the 5% probability level. After testing, the difference between the coefficients of the two groups is significant at the 1% probability level, and the former is significantly better than the latter. Therefore, in guiding the internal reform of state-owned enterprises, market-oriented mechanisms can be considered to enable managers to actively pay attention to market trends and respond positively to competitive threats.

#### *3.5.2. Whether the Enterprise Belongs to a Heavily Polluting Industry*

Digital development promotes the development of the local marketization process and the green financial market at the same time, and the two mechanisms will coordinately adjust the quality of corporate environmental disclosure. With the acceleration of the marketization process, green production factors often flow to and concentrate in innovative enterprises. Heavily polluting enterprises with negative environmental externalities have inherent disadvantages in market competition, and they need to perform substantive environmental governance work and improve the quality of environmental disclosure to make up for the disadvantages in market competition (Wang

Yaming et al., 2015); with the improvement of the green financial market, green financial tools will support enterprises' green innovation activities by reasonably matching returns and risks and reducing financing constraints (Du Jinzhu and Wu Zhanyong, 2022), which creates a development opportunity for heavily polluting enterprises with weak high-tech foundations and insufficient innovation. At the same time, the green credit market has a significant financing penalty and investment inhibition effect on heavily polluting enterprises, which will force heavily polluting enterprises to accelerate green transformation and the quality of environmental information disclosure to maintain organizational legitimacy.

Relevant empirical research results provide empirical support. In Table 9, Columns (3)-(4) report the heterogeneous effects of the two types of enterprises. In Column (3), at the 1% probability level, digital development will significantly improve the environmental disclosure quality of heavily polluting enterprises, with a Dig coefficient of 6.807, which has highly significant economic significance; while in Column (4) of the non-heavily polluting industry sample group, the Dig coefficient is 2.611 and significant at the 1% probability level. After the inter-group coefficient difference test, the coefficient difference of Dig is 4.196, which is significant at the 1% probability level. It can be concluded that digital development will have a more significant role in improving the disclosure quality of heavily polluting enterprises.

***3.5.3 Region Where the Enterprise is Located***

In China, the eastern and central regions are economically developed and have good market infrastructure and supporting legal systems, creating favorable conditions for the integration of relevant digital factors into industrial operations (Li Bing, 2016) — it not only forces enterprises to transmit green market signals through factor optimization allocation and strengthening competition effects, but also encourages enterprises to improve the quality of environmental disclosure through more transparent information flow to obtain green financial support.

Relevant empirical research results provide empirical support. Considering the prominent regional economic specificity in China, Columns (5)-(6) of Table 9 compare the differences in the role of digital development in different regions of China: it can be found that in China's eastern and central regions, digital development can significantly improve the quality of corporate environmental information disclosure, with a Dig coefficient of 3.179 and significant at the 1% probability level; while in the western region, the Dig coefficient is not significant.

Table 9 Heterogeneity Test Results

| | (1) SOEs | (2) Private Enterprises | (3) Heavily Polluting | (4) Non-Heavily Polluting | (5) East & Central | (6) West |
|---|---|---|---|---|---|---|
| VARIABLES | EDQ | EDQ | EDQ | EDQ | EDQ | EDQ |
| Dig | 5.773*** | 1.267** | 6.807*** | 2.611*** | 3.179*** | 3.604 |
| | (1.341) | (0.626) | (1.872) | (0.648) | (0.670) | (4.206) |
| Control | Yes | Yes | Yes | Yes | Yes | Yes |
| Constant | -14.01*** | -8.542*** | -18.98*** | -10.67*** | -9.117** | -14.14** |
| | (5.061) | (3.275) | (5.285) | (3.545) | (3.673) | (6.722) |
| Year Fixed Effect | Yes | Yes | Yes | Yes | Yes | Yes |
| Industry Fixed Effect | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 8,020 | 12,137 | 5,464 | 14,696 | 16,525 | 2,646 |
| R2 | 0.257 | 0.115 | 0.237 | 0.188 | 0.230 | 0.269 |

*Robust standard errors in parentheses and cluster at firm level；*** p<0.01，** p<0.05，* p<0.1*

### *3.6 Endogeneity and Robustness Test Results*

#### *3.6.1. Robustness Test*

First, this paper will conduct robustness tests, including replacing the regression model, adding control variables, strengthening fixed effects, deleting problematic samples, and replacing the explained variable.

Replacing the Benchmark Regression Model.

In the benchmark regression model of this chapter, there are a certain number of enterprises that choose not to disclose environmental information in specific years, as well as phenomena where the environmental information quality score is zero. Therefore, during the sample period, the corporate disclosure quality data is significantly right-skewed. Therefore, this paper considers replacing the benchmark regression model with a restricted regression model: Tobit model for regression testing. The relevant results are shown in Column (1) of Table 10. The Dig coefficient is 3.563, which is significant at the 1% probability level.

Adding New Control Variables.

Digital development will adjust the quality of corporate environmental disclosure through market-oriented incentive mechanisms. However, among them, the new market competition pressure makes enterprises have to pay attention to cost management and resource allocation. To a large extent, the "greenwashing" decoupling behavior existing in some enterprises lies in resource scarcity and environmental constraints. If enterprises are endowed with more redundant resources, it will enhance their ability to resist liquidity risks and respond to market competition (Lise, 2013; Gou Qianwen et al., 2019; Gao et al., 2021). The mitigation of financial constraints will help enterprises adjust their green motivation, increase substantive environmental protection behaviors, and strengthen differential information disclosure.

Based on the above analysis, this part adds two relevant control variables: according to the external environment, "government environmental subsidies" are added. Such green subsidies will alleviate the cost pressure of corporate environmental governance, and good information disclosure will also help them obtain more green financial quotas; according to the internal environment, "capital accumulation rate" is added, because internal capital accumulation can reduce corporate financing constraints and enhance enterprises' confidence in responding to operational risks and market competition (Cui et al., 2021). The relevant results are shown in Table 10. With the addition of two new control variables, the Dig coefficient is 3.534, which is significant at the 1% probability level, and the conclusion of the benchmark regression still holds.

Controlling for Regional Heterogeneity.

Due to the inevitable multiple differences in systems, cultures, policies, etc. between locations, which will promote or restrict marketization development and corporate environmental disclosure, Column (3) in Table 10 controls for the fixed effects at the provincial and municipal levels where the enterprise is located to control the impact of unobservable regional heterogeneity on the quality of environmental disclosure. The coefficient of the Dig indicator is 3.229, which is significant at the 1%

probability level.

Excluding Problematic Samples.

In the full sample, it is inevitable that some enterprises will be subject to key investigations or severe penalties by relevant departments due to sudden environmental accidents, environmental violations or environmental petitions in the current year. Such enterprises face certain legitimacy pressure in the short term and have to abandon "greenwashing", increase substantive environmental protection behaviors, and improve the quality of environmental information disclosure to alleviate external environmental recognition pressure. Therefore, the improvement of the information disclosure quality of involved enterprises is likely to be a public relations measure deliberately "forced" by environmental accidents. Therefore, this paper will exclude such samples of involved enterprises. The results are shown in Column (4) of Table 10: the marginal coefficient of Dig is 4.223, which is still significant at the 1% probability level.

Replacing the Explained Variable.

According to the form of corporate environmental disclosure, it is divided into three categories: mixed disclosure, social responsibility report disclosure and independent environmental report disclosure. Among them, the first category is reflected in the appearance of environmental governance information in the corporate financial annual report, and the form of information disclosure is not significant; while the second and third categories are reflected in the enterprise issuing special reports related to ESG to display the corporate green image in a relatively systematic, comprehensive and detailed manner, improving information salience.

This paper replaces the explained variable with a dummy variable "whether to disclose a special report", denoted by EDQ2. If an independent environmental report is disclosed in the current year, this indicator is recorded as 1, otherwise it is 0. The results are shown in Column (5) of Table 10: the marginal coefficient of Dig is 0.152, which is significant at the 1% probability level, and the original relationship holds.

*Table 10 Robustness Test Results*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| **VARIABLES** | **EDQ** | **EDQ** | **EDQ** | **EDQ** | **EDQ2** |
| Dig | 3.563*** | 3.534*** | 3.229*** | 4.223*** | 0.152*** |
| | (0.666) | (0.667) | (1.097) | (1.268) | (0.0447) |
| Control | Yes | Yes | Yes | Yes | Yes |
| Subsidy | | -0.527*** | | | |
| | | (0.0708) | | | |
| Capital | | 0.200*** | | | |
| | | (0.0542) | | | |
| Constant | -16.059*** | -13.66*** | -11.59*** | 1.855 | -0.245 |
| | (2.956) | (2.981) | (4.115) | (3.871) | (0.200) |
| Year Fixed Effect | Yes | Yes | Yes | Yes | Yes |
| State Fixed Effect | No | No | Yes | No | No |
| Industry Fixed Effect | No | No | No | No | No |
| Observations | 20,161 | 20,131 | 20,160 | 19,793 | 20,160 |
| R2/Pseudo R2 | 0.0356 | 0.215 | 0.228 | 0.818 | 0.165 |

*Robust standard errors in parentheses and cluster at firm level; *** p<0.01, ** p<0.05, * p<0.1*

### *3.6.2. Endogeneity Test*

Exogenous Policy Shock

Urban digital development is often promoted and implemented through specific policies. Relative to individual enterprises in the region, policies are often exogenous, and there is less reverse causality. However, the traditional research method to evaluate the effect of digital policies mainly sets a dummy variable about "whether the policy has occurred" and incorporates it into the regression model for testing; in contrast, the Difference-in-Differences (DID) method can more accurately evaluate the effect of digital policies.

In summary, this paper will use the multi-period Difference-in-Differences (DID) model to conduct endogeneity tests. This part takes the "Broadband China" strategy as an exogenous shock to urban digital development. With the implementation of the "Broadband China" strategy, it provides better digital and intelligent infrastructure for government environmental supervision and public environmental supervision; it will also accelerate the market-oriented and efficient reallocation of factors in the region, restrain enterprise monopoly and rent-seeking phenomena, and increase local market competition.

Referring to the practice of Li Jun and Li Jing (2021), a dummy variable Noticej,t is set according to whether city j where the enterprise is located is nominated by the "Broadband China" strategy in year t and later: 1 if yes, 0 otherwise. The relevant multi-period DID model is shown below:

$$EDQ_{i,t} = \alpha + \beta_1 * Notice_{j,t} + Controls_{i,t} + Year\ Dummies + Industry\ \ Dummies + \varepsilon_{i,t}$$

EDQi,t=α+β1*Noticej,t+Controlsi,t+Year Dummies+Firm Dummies+Ɛi,t

From 2011 to 2019, taking the year when the city where the enterprise is located was nominated as the base period (this period's dummy variable will be excluded in the test), the parallel trend test is first conducted to ensure that there are no systematic differences between cities before the implementation of the strategy. This paper sets 9 annual dummy variables: Pre4, Pre3, Pre2, Pre1, Current, Post1, Post2, Post3, and Post4, according to whether the city where the enterprise is located belongs to the year. Before conducting the multi-period DID model, a parallel trend test is performed.

After controlling for year and industry fixed effects, the relevant results are shown in Column (1) of Table 11: although the marginal effect of the time dummy variable has a certain lag, it gradually becomes statistically significant from the second period after nomination, and the economic significance is also more significant, showing a "accumulate strength for a take-off" trend. Column (3) includes firm fixed effects into the regression, and the results remain robust.

Columns (2) and (4) of Table 11 show the test results of the multi-period DID model. Compared with cities not nominated by the "Broadband China" strategy, in cities that have been nominated, the environmental information quality of local enterprises shows a more obvious improvement. The marginal effects of Noticej,t are 0.671 and 0.335, which are significant at the 1% and 5% probability levels, respectively.

*Table 11 Endogeneity Test Results (I)*

| | （1） | （2） | （3） | （4） |
|---|---|---|---|---|
| **VARIABLES** | **EDQ** | **EDQ** | **EDQ** | **EDQ** |
| Pre4 | 0.0941 | | -0.0120 | |
| | (0.189) | | (0.127) | |
| Pre3 | 0.184 | | -0.0719 | |
| | (0.202) | | (0.123) | |
| Pre2 | 0.285 | | 0.0254 | |
| | (0.209) | | (0.123) | |
| Pre1 | 0.338 | | 0.117 | |
| | (0.207) | | (0.112) | |
| Current | | | | |
| Post1 | 0.440** | | 0.228** | |
| | (0.200) | | (0.107) | |
| Post2 | 0.727*** | | 0.490*** | |
| | (0.229) | | (0.163) | |
| Post3 | 0.974*** | | 0.674*** | |
| | (0.295) | | (0.226) | |
| Post4 | 1.219*** | | 0.777*** | |
| | (0.363) | | (0.297) | |
| Notice | | 0.671*** | | 0.335** |
| | | (0.224) | | (0.159) |
| Control | Yes | Yes | Yes | Yes |
| Constant | -13.30*** | -13.14*** | 2.418 | 2.451 |
| | (3.012) | (2.986) | (4.096) | (3.961) |
| Year Fixed Effect | Yes | Yes | Yes | Yes |

|  | （1） | （2） | （3） | （4） |
| --- | --- | --- | --- | --- |
| **VARIABLES** | **EDQ** | **EDQ** | **EDQ** | **EDQ** |
| Industry Fixed Effect | Yes | Yes | No | No |
| Firm Fixed Effect | No | No | Yes | Yes |
|  |  |  |  |  |
| Observations | 20,165 | 20,165 | 19,993 | 19,993 |
| R2 | 0.205 | 0.205 | 0.819 | 0.818 |

*Robust standard errors in parentheses and cluster at firm level；*** p<0.01，** p<0.05，* p<0.1*

Instrumental Variable Method.

There may be a mutual causal relationship between urban digital development and the quality of local enterprises' environmental disclosure. As mentioned earlier, a developed factor market is a necessary prerequisite for enterprises to obtain competitive advantages by disclosing environmental information. Rational investors will make effective judgments on enterprises' operating conditions and investment value through the disclosed information (Yan Haizhou and Chen Baizhu, 2017). Based on this, to drive enterprises to disclose high-quality environmental information, achieve the effects of lowering financing thresholds, increasing stock prices, and enhancing market competitiveness, it is objectively necessary to rely on the local good marketization level, green financial market construction, and relatively developed urban information technology to improve the dissemination effect of disclosure. All these put forward new requirements for urban digital development.

To alleviate the above endogeneity problem, this paper considers using the Instrumental Variable (IV) method to conduct 2SLS regression. This paper selects "the contemporaneous average of digital development levels of all other cities in the same province" as the instrumental variable on an annual basis, for the following reasons: within the same province, cities often have similar digital development environments, enjoying similar regional digital economy policies, talent structures, and scientific and technological resources; however, the digital situation of other cities in the same province is relatively exogenous, which will not directly affect the environmental information disclosure quality of enterprises in a specific city, and the latter at the micro level is even less likely to directly affect the former at the macro level.

*Table 12 Endogeneity Test Results (II)*

| | (1) | (2) | （3） |
|---|---|---|---|
| **VARIABLES** | **Dig** | **EDQ** | **EDQ2** |
| Dig | | 4.163*** | 0.215*** |
| | | (0.805) | （0.052） |
| LProvincial_Dig | 0.981*** | | |
| | (0.005) | | |
| Controls | Yes | Yes | Yes |
| Constant | -0.002 | -18.384*** | -0.459*** |
| | (0.027) | （2.401） | （0.165） |
| Year Fixed Effect | Yes | Yes | Yes |
| Industry Fixed Effect | Yes | Yes | Yes |
| Observations | 17,425 | 17,425 | 17,425 |
| R2 | 0.866 | 0.21 | 0.16 |

*Robust standard errors in parentheses and cluster at firm level；*** p<0.01，** p<0.05，* p<0.1*

As shown in Table 12, Columns (1)-(2) correspond to the two stages of 2SLS regression, respectively. In the first stage, the marginal coefficient of LProvincial_Dig, as the instrumental variable, is 0.981, which is significant at the 1% probability level, and the goodness of fit of the corresponding model reaches 86.6%. In the second stage, this paper takes Dig, which overcomes other disturbance factors and is only affected by the exogenous instrumental variable, as the explanatory variable. After incorporating it into the regression analysis, it is found that: in Column (2), the marginal coefficient of Dig on EDQ is 4.163, which is significant at the 1% probability level, and the goodness of fit of the corresponding model reaches 21%; in Column (3), after replacing the explained variable with the robustness indicator "whether to disclose a special report", the marginal coefficient of Dig on EDQ is 0.215, which is significant at the 1% probability level, and the goodness of fit of the corresponding model reaches 16%. The above empirical data confirms the robustness of the positive relationship between the two.

## 4.Conclusion

### *4.1 Discussion on Research Results*

Corporate environmental information disclosure is driven by both legitimacy pressure and competitive pressure, with certain game characteristics: for example, Li Qiang and Feng Bo (2015), Gou Qianwen et al. (2019) found that even if enterprises respond to external demands and carry out environmental information disclosure work, there may still be problems such as unclear disclosure, low content of effective information, and even symbolic disclosure. Therefore, the research on environmental information disclosure needs to integrate multiple theoretical logics.

Guided by the legitimacy theory and signal transmission theory, this paper conducts a series of empirical tests and finds that: urban digital development increases local environmental supervision, and enterprises increase environmental disclosure under the pressure of legitimacy to respond to the needs of stakeholders for their own environmental information, which is essentially the maintenance of organizational legitimacy (Xiao Hongjun et al., 2013; Gou Qianwen et al., 2015; Gou Qianwen et al., 2019; Zhang Qi et al., 2019); and the quality of disclosure also continues to improve, with the mechanism of action manifested as the dual incentives of market-oriented competition and green financial market. This is because digital factors can promote the optimal allocation of local market resources, dilute relational resources, and promote the development of contractual economy, making the acquisition of resource factors more dependent on market competition. In a more improved, fair and effective market environment, enterprises can enhance market competitiveness and strive for differentiated resources by transmitting differentiated signals such as green responsibility.

Compared with the studies of Li Qiang and Feng Bo (2015), Gou Qianwen et al. (2019), Zhu Shujin et al. (2022), the empirical evidence of this paper shows that macro digital factors help enterprises substantially improve the quality of disclosure and alleviate the past symbolic disclosure behavior (perfunctory disclosure lacking effective information content). In contrast, previous studies mainly regarded disclosure work as an additional high-cost task, which may not have fully paid attention to the medium and long-term value of information disclosure, nor effectively combined with financing channels such as the green financial market, resulting in partial differences in views among the literature.

In addition, the empirical results of this paper also indicate that with the in-depth development of urban digitalization, high-quality environmental information disclosure will obtain more institutional support (such as a more improved green financial market), helping enterprises alleviate financing constraints and strive for differentiated resources. Therefore, green disclosure should not be limited to passive response, but should be used as a competitive tool. This result helps alleviate the concern mentioned by Li Qiang and Feng Bo (2015) that enterprises may suffer from the "fast cow effect" when becoming "green models".

### *4.2 Relevant Development Suggestions*

First, enterprise managers should strengthen their understanding of environmental strategies and explore the market value of information disclosure. Environmental information disclosure is not only used to respond to the pressure from stakeholders and meet the external demand for environmental information, but also can be used as a display of the enterprise's own green image, transmitting differentiated signals to the market, distinguishing itself from other enterprises, and thus enhancing consumer confidence.

Second, enterprises should consider how to link environmental governance with economic performance in their daily activities, create conditions for future economic profit growth through environmental governance, and realize a green operation path.

Third, enterprises should pay close attention to the policy trends of local environmental supervision, such as changes in green financial market policies, take this as an opportunity to carry out green investment and transformation, improve the quality of information disclosure through more substantive environmental actions, and respond to the environmental information needs of the government, the public, and green financial institutions in a targeted manner to enhance market reputation.

In the perspective of government, on the one hand, the government should use market-oriented means to reward and certify enterprises with good information disclosure. The government should encourage local enterprises to consciously participate in environmental information disclosure, and invite third-party professional institutions to conduct professional review and evaluation of enterprises that issue disclosure reports. After the relevant content is verified and rated, enterprises with good disclosure quality will be given preferential treatments such as reducing environmental taxes and fees, lowering green barriers in the corresponding market, and awarding the honorary title of "Green Model". When enterprises realize that high-quality disclosure can bring practical benefits to themselves, they will be more motivated to participate in substantive environmental governance and abandon the long-term greenwashing behavior.

On the other hand, the government should issue clear and well-defined disclosure guidelines with distinct rewards and punishments, standardize the form and content of enterprise disclosure, and guide enterprises to actively disclose substantive behavioral information such as environmental governance performance and transmit their own unique green information. For enterprises with environmental emergencies, excessive pollution, poor disclosure quality, or proven "greenwashing", the government should conduct differentiated disclosure management, and implement more strict disclosure standard management and supervision on them (such as inviting third-party institutions to conduct auditing and certification), so as to build an orderly environment for protecting the legitimate interests of investors, consumers, and the general public.

In addition, the government should focus on encouraging ordinary enterprises to issue professional social or environmental responsibility reports, and appropriately mandate heavily polluting industries to issue "Environmental Reports" to disclose detailed information in detail, so as to improve the visibility of their own environmental disclosure, facilitate public supervision, and promote their better performance of environmental responsibilities in the future.

## References


Berliner, D. and Prakash, A., 2012. From norms to programs: The United Nations Global Compact and global governance. *Regulation & Governance*, 6(2), pp.149–166.

Berliner, D. and Prakash, A., 2013. Signaling environmental stewardship in the shadow of weak governance: The global diffusion of ISO 14001. Law & Society Review, 47(2), pp.345–373.

Dowell, G., Hart, S. and Yeung, B., 2000. Do corporate global environmental standards create or destroy market value? Management Science, 46(8), pp.1059–1074.

Boli, I., 2003. Openness and the environment in Central and Eastern Europe: Can trade and foreign investment stimulate better environmental management in enterprises? The Journal of Environment & Development, 12(2), pp.177–204.

Manso, G., 2011. Motivating innovation. The Journal of Finance, 66(5), pp.1823–1860.

Melnyk, S.A., Sroufe, R.P. and Calantone, R., 2004. Assessing the impact of environmental management systems on corporate and environmental performance. Journal of Operations Management, 21(3), pp.329–351.

Nason, R.S., Bacq, S. and Gras, D., 2018. A behavioral theory of social performance: Social identity and stakeholder expectations. Academy of Management Review, 43(2), pp.259–283.

Potoski, M. and Prakash, A., 2005. Green clubs and voluntary governance: ISO 14001 and firms' regulatory compliance. American Journal of Political Science, 49(2), pp.235–248.

Quairel-Lanoizelée, F., 2011. Are competition and corporate social responsibility compatible? The myth of sustainable competitive advantage. Society and Business Review, 6(1), pp.77–98.

Robert, C.H., 1977. How much growth can a firm afford? Financial Management, 8, pp.186–198.

Russo, M.V. and Fouts, P.A., 1997. A resource-based perspective on corporate environmental performance and profitability. Academy of Management Journal, 40(3), pp.534–559.

Sam, A.G. and Song, D., 2022. ISO 14001 certification and industrial decarbonization: An empirical study. Journal of Environmental Management, 323(12), pp.116–169.

Waddock, S.A. and Graves, S.B., 1997. The corporate social performance–financial performance link. Strategic Management Journal, 18(4), pp.303–319.

Zeng, S.X., Xu, X.D., Dong, Z.Y. et al., 2010. Towards corporate environmental information disclosure: An empirical study in China. Journal of Cleaner Production, 18(12), pp.1142–1148.

Chen, D. and Hu, Q., 2022. Research on corporate governance in the digital economy era: Paradigm innovation and practical frontiers. Management World, 38(06), pp.213–240.

Chen, X., Li, Y. and Song, L., 2022. Theoretical system and research prospect of digital economy. Management World, 38(02), pp.208–224.

Chen, Y., Song, T. and Huang, J., 2022. Enterprise digital transformation: Following the market or adapting to local conditions? Studies in Science of Science, 40(06), pp.1054–1062.

Dai, Y. and Shi, M., 2016. Modular design of environmental information disclosure based on Internet+. Communications of Finance and Accounting, 34, pp.102–105.

Guo, J. and Luo, P., 2016. Does the Internet promote total factor productivity in China? Management World, 10, pp.34–49.

Han, F. and Jiang, Z., 2023. Research on the productivity improvement effect of enterprise digitalization from the perspective of agglomeration network. Management World, 39(11), pp.54–77.

He, Y., Tang, Q. and Wang, K., 2017. Carbon performance and financial performance. Accounting Research, 02, pp.76–82.

Gou, Q., Cai, N. and Xin, Y., 2019. Research on the causes of enterprise environmental behavior decoupling. Zhejiang Social Sciences, 02, pp.19–27.

Hu, Q., 2012. Research on the correlation between environmental performance and financial performance of listed companies. China Population, Resources and Environment, 22(06), pp.23–

32.

Hu, W., Du, J. and Zhang, W., 2021. Does information disclosure quality affect corporate risk-taking? Operations Research and Management Science, 30(07), pp.210–217.

Huang, Q., Yu, Y. and Zhang, S., 2019. Internet development and manufacturing productivity improvement. China Industrial Economics, 08, pp.5–23.

Jiao, H., Yang, J. and Wang, P., 2021. Research on the mechanism of data-driven enterprise dynamic capabilities. China Industrial Economics, 11, pp.174–192.

Jiao, H., Zhang, R. and Yang, J., 2023. Enterprise strategic choice and digital platform ecosystem construction. Management World, 39(12), pp.202–229.

Kong, D., Wei, Y. and Ji, M., 2021. Research on the impact of the conversion from environmental fees to taxes on enterprises' green information disclosure. Securities Market Herald, 08, pp.2–14.

Li, W. and Lu, X., 2015. Do institutional investors pay attention to corporate environmental performance? Journal of Financial Research, 12, pp.97–112.

Li, Q. and Feng, B., 2015. Will enterprises disclose environmental information low-key? Journal of Zhongnan University of Economics and Law, 04, pp.141–148.

Lin, H., Wang, L. and Wang, F., 2007. Environmental performance, corporate responsibility and product value reengineering. Management World, 05, pp.155–157.

Liu, B., Yu, Q., Bi, J. et al., 2009. Research on the influencing factors of corporate environmental performance based on stakeholder theory. China Population, Resources and Environment, 19(06), pp.80–84.

Liu, S., Yan, J. and Zhang, S., 2021. Can the digital transformation of enterprise management improve input-output efficiency? Management World, 37(05), pp.170–190.

Liu, Y., Yin, J. and Li, D., 2020. Exploratory analysis of the conflict of multiple institutional logics in social enterprises. R&D Management, 32(03), pp.13–24.

Mei, X., Ge, Y. and Zhu, X., 2020. Research on the impact mechanism of environmental legitimacy pressure on corporate carbon information disclosure. Soft Science, 34(08), pp.78–83.

Qi, F. and Wu, X., 2023. Regulatory uncertainty and anchoring effect of environmental information disclosure. Communications of Finance and Accounting, 01, pp.51–55.

Ren, G., 2017. Research on corporate environmental investment behavior from the perspective of corporate governance. Journal of Zhengzhou University (Philosophy and Social Sciences Edition), 50(03), pp.66–71.

Shi, J. and Li, N., 2022. The impact of digital finance on corporate environmental information disclosure. Journal of Financial Research, 05, pp.123–138.

Song, M. and Li, X., 2020. Digital transformation and corporate environmental performance. China Industrial Economics, 09, pp.113–130.

Sun, H. and Zhang, L., 2021. Environmental regulation, digital transformation and enterprise green innovation. Economic Research Journal, 56(03), pp.124–138.

Tang, J. and Wang, Y., 2018. The impact of stakeholder pressure on corporate environmental information disclosure quality. Management World, 34(07), pp.112–120.

Wang, D. and Li, J., 2022. Digital economy and corporate environmental responsibility. China Population, Resources and Environment, 32(04), pp.45–53.

Wang, H. and Chen, Y., 2022. Corporate digital transformation and environmental information disclosure: The moderating role of internal control. Accounting Research, 02, pp.134–146.

Wang, J. and Zhang, W., 2020. The impact of internet development on corporate environmental performance. Journal of Business Economics, 08, pp.34–45.

Wang, L. and Li, C., 2023. ISO 14001 certification, digital transformation and corporate environmental performance. Journal of Environmental Management, 335(05), 117689.

Wu, X. and Zhang, Q., 2019. Corporate social responsibility and environmental information disclosure: The mediating role of corporate reputation. Journal of Management Sciences in China, 22(06), pp.78–92.

Xie, H. and Zhang, M., 2022. Digital transformation, environmental uncertainty and corporate risk-taking. Journal of Financial Economics, 40(03), pp.89–105.

Xu, B. and Li, N., 2021. The impact of digital inclusive finance on rural environmental governance. China Rural Survey, 03, pp.56–70.

Xu, G. and Chen, L., 2018. Corporate environmental performance evaluation and its impact on financial performance. Journal of Central University of Finance & Economics, 09, pp.89–98.

Yang, G. and Zhang, Y., 2021. Digital transformation and enterprise green production behavior. Journal of Industrial Economics, 04, pp.112–128.

Yao, S. and Liu, J., 2019. Environmental protection investment and corporate performance. China Economic Quarterly, 18(02), pp.567–588.

Ye, D. and Wang, X., 2020. The impact of corporate governance on environmental information disclosure: A literature review. Journal of Economic Surveys, 34(03), pp.789–812.

Yu, Y. and Li, M., 2020. Digital economy and green total factor productivity. Economic Research Journal, 55(09), pp.130–144.

Zhang, B. and Bi, J., 2021. Stakeholder engagement and corporate environmental performance. Journal of Cleaner Production, 292(08), 125987.

Zhang, H. and Li, W., 2022. Digital transformation and corporate environmental information disclosure quality: Based on the perspective of media attention. Journal of Business Ethics, 179(02), pp.456–472.

Zhao, T., Zhang, Z. and Liang, S., 2020. Digital economy, entrepreneurial activity and high-quality development: Empirical evidence from Chinese cities. Management World, 36(10), pp.65–76.

Zhou, Q., Wang, Y. and Yang, W., 2020. An empirical study on the impact of digitalization level on innovation performance. Science Research Management, 41(07), pp.120–129.

Zhu, W., Sun, Y. and Tang, Q., 2019. Substantive disclosure or selective disclosure: The impact of corporate environmental performance on the quality of environmental information disclosure. Accounting Research, 03, pp.10–17.

# A Neural Network Algorithm-Based Prediction Model for Depression Among Elderly Individuals with Diabetes Mellitus

*Xiaojuan Wang[1†], Shaofeng Wang[1†], Wencong Xu[1*], Xiaoqin Zhao[1], Zhang Zhao[1], Jing Li[1]*

*[1]School of Health and Medical Technology, Neusoft Institute Guangdong 528225, China*

Keywords:

Geriatric Diabetes;
Depression;
Random Forest;
BP Neural Network;
CHARLS

ABSTRACT

**Objective:** To address the "syndemic" of depression among elderly diabetic patients in China by developing a non-invasive machine learning predictive model for large-scale community screening.

**Methods:** Data from 4,827 participants (aged 60–85) were extracted from the 2020 CHARLS database. Random Forest (RF), a tree-based machine learning technique, was used to select important variables. Then, a Backpropagation Neural Network (BPNN), which can capture complex, non-linear relationships, was built. Model performance was evaluated using Accuracy (correct prediction rate) and AUC (area under the curve, a measure of discrimination ability).

**Results:** The prevalence of depressive symptoms was 33.7%. RF identified age, physical pain distribution, and sleep duration as the top predictors. Although post-noon nap duration was not significant in a simple statistical analysis, it emerged as important in a neural network, which detects hidden patterns. The model achieved an accuracy of 0.67 (percentage of correct predictions) and AUC of 0.59 (ability to distinguish cases) on the test set, indicating strong negative predictive value for risk triaging.

**Conclusion:** Physical pain, sleep patterns, and social engagement are decisive warning signs for depression in elderly diabetics. This non-invasive model is an essential, scalable solution for early identification and intervention within primary healthcare systems.

CONTACT:

[1]Author: wangxiaojuan@nuit.edu.cn; wangshaofeng@nuit.edu.cn

[*]Corresponding Author: xuwencong@nuit.edu.cn

Note: Authors designated with "†" are recognised as co-first authors.

DOI: 10.64549/jaai-ii.v1i1.52

## 1.Introduction

China's ageing population has led to a marked rise in diabetes among the elderly. The comorbidity of diabetes with other diseases presents a significant challenge for global diabetes health interventions and management. Mental health issues associated with diabetes, particularly depression, are frequently overlooked due to a predominant clinical focus on physical health. Within global public health research, the co-occurrence of diabetes and depression has been identified as a “syndemic.” Recent epidemiological studies indicate that the prevalence of depressive symptoms among elderly diabetic patients in China ranges from 35% to 38.6%, a rate substantially higher than that observed in non-diabetic elderly populations. This prevalence is especially pronounced among women, individuals with lower education levels, and those with chronic complications [1].

The comorbidity of diabetes and depression represents a complex, bidirectional, and mutually reinforcing relationship. Laboratory studies indicate that prolonged hyperglycemia and associated metabolic disorders constitute the endocrine basis for depressive states in individuals with diabetes by inducing systemic low-grade inflammation and activating the hypothalamus-pituitary-adrenal (HPA) axis [2, 3].

Despite the high predictive value of biomarkers in individual experimental settings, large-scale community screening for routine risk assessment remains challenging due to high testing costs, invasive sampling procedures, and reliance on primary healthcare infrastructure. Existing studies indicate that depressive episodes in elderly patients with chronic diseases are associated with factors such as age, lifestyle, social support, and the accumulation of chronic conditions [4, 5, 6]. Consequently, intelligent prediction tools utilising non-invasive and easily accessible features offer greater applicability for large-scale population screening. Traditional clinical prediction models primarily employ statistical methods such as logistic regression (LR). While these models provide strong interpretability, they assume linear relationships between variables. In elderly populations, however, complex multidimensional nonlinear interactions often exist between lifestyle variables and metabolic indicators, which limit the predictive accuracy of traditional models [7].

Machine learning approaches are therefore more effective in constructing depression prediction models for elderly diabetic patients, as they can accommodate complex interactions. Recent studies have applied single algorithms, including random forests and neural networks, to develop depression

prediction models for elderly patients [8, 9]. Although random forests (RF) demonstrate strong performance in feature selection and classification tasks involving chronic disease data, their reliance on a stepwise decision space limits their ability to capture smooth, nonlinear interactions among high-dimensional variables, such as lifestyle and disease status [10]. Neural network-based analyses, while advantageous in global function approximation, are susceptible to local optima and often lack interpretability when applied to structured data from lifestyle questionnaires [11]. Integrating the robust feature extraction capabilities of random forests with the deep mapping strengths of neural networks can therefore improve the accuracy of complex biopsychosocial model construction [12].
The fifth round of follow-up questionnaires from the 2020 China Health and Retirement Longitudinal Study (CHARLS) database was analysed. Data on lifestyle, social support networks, and physiological functions of elderly individuals were integrated. Random forests were applied for feature engineering optimisation to remove redundant variables, and a backpropagation neural network was used to construct a nonlinear predictive model. This research paradigm aims to develop a low-cost, efficient digital early warning tool, offering empirical support for the early identification and intervention of depressive disorders in elderly diabetic patients.

## 2. Data Sources and Research Methods

### *2.1 Data Sources*

The fifth round (2020) follow-up questionnaire data from the China Health and Retirement Longitudinal Study (CHARLS) served as the primary data source. Depressive symptoms were measured using the CES-D score as the dependent variable. The CES-D comprises 10 items, each categorised by symptom frequency as "rarely or never," "not very often," "sometimes or half the time," and "most of the time," corresponding to values of 0, 1, 2, and 3, respectively. Items 5 and 8 were reverse-scored, resulting in a total possible score ranging from 0 to 30. A score of 10 or higher was classified as indicative of depressive symptoms (assigned a value of 1), while a score between 0 and 9 was considered within the normal range (assigned a value of 0).

Sixteen independent variables related to depressive symptoms were included and categorised into the following five types:

(1) Demographic variables: gender and age.

(2) Health status variables: Number of chronic diseases (hypertension, diabetes, dyslipidemia, etc., 15 types), number of pain locations (head, shoulder, arm, etc., 15 types), whether or not a hip fracture has been experienced, whether or not a traffic accident has been experienced, whether or not outpatient visits have been made in the past month, whether or not hospitalization has been made in the past year, etc., 6 types.

(3) Health behaviour variables. Frequency of alcohol consumption, sleep duration, duration of afternoon nap, regular exercise. Regular exercise was scored as follows: 0, 1, 2, and 3 points were assigned for no regular exercise, light exercise, moderate exercise, and high-intensity exercise, respectively. The highest applicable score was recorded for each respondent. Social activity scores were determined by the frequency of participation in activities such as visiting friends and family, playing mahjong, chess, or cards, providing assistance to non-cohabiting relatives or neighbours, engaging in physical activities, participating in community organisations, volunteering, attending educational courses, and other social activities. Frequency was scored as 3 points for "almost daily," 2 points for "almost weekly," and 1 point for "not often." The total social activity score was calculated by summing these values. were then added together to obtain the respondents' social activity score.

(4) Family status variables: marital status and pension payment frequency. The initial CHARLS 2020 dataset contained 19,367 respondents. A systematic data-cleaning process was used to screen for valid samples. The specific steps were as follows: First, the research subjects were limited to the elderly aged 60-85; second, samples with no response records on the depression scale and missing values in the dependent variable were removed; finally, based on clinical rationality and data distribution characteristics, outliers were strictly judged and excluded. Specific exclusion criteria included: (1) sleep time $\leq$ 4 hours; (2) nap duration $\geq$ 4 hours; (3) by fitting a curve to a scatter plot of pain location and depression score, outliers significantly deviating from the overall trend were identified and excluded. After the multi-stage screening described above, 4,827 valid samples were included for subsequent analysis.*2.2 Data Collection (Times New Roman 12pt, Italics)*

Body texts: The research method explains the implementation methods employed in the study. The method is described clearly and in detail.

***2.2 Feature Screening***

The importance of predictive features related to depression levels was quantified and ranked using the random forest ensemble learning algorithm. The 4,827 samples (1,627 positive and 3,200 negative) were randomly divided into training and test sets at a 1:1 ratio. First, oversampling was used to balance the training set. Second, grid search cross-validation was used to systematically optimise the key hyperparameters of the random forest to ensure the stability of the feature importance ranking. Finally, the relative importance differences between features were visualised using bar charts.

*2.3 Model Construction and Performance Evaluation*

For the BP neural network model, the dataset was split into training and test sets at a 7:3 ratio, and the selected important features were standardised using Z-scores. A multilayer perceptron (MLP) classifier was constructed with sklearn.neural_network.MLPClassifier. The training data were used to fit the MLP classifier. Model performance was evaluated by plotting the receiver operating characteristic (ROC) curve and calculating the area under the curve (AUC).

## 3. Results

### *3.1 General Information*

A total of 4,827 samples were included. Based on depression scale scores, participants were categorised into a depressive symptoms group (n=1,627) and a non-depressive symptoms group (n=3,200). The sample comprised 68.13% females (3,288 cases) and 31.87% males (1,539 cases). The age distribution was as follows: 60–69 years (59.98%, 2,901 cases), 70–79 years (33.62%, 1,623 cases), and 80–85 years (6.40%, 303 cases). Statistically significant differences between the depressive and non-depressive groups were observed for gender, age group, marital status, frequency of alcohol consumption, history of hip fracture, number of chronic diseases, sleep duration, number of pain sites, internet use, and regular exercise (P<0.05). No statistically significant differences were found for history of traffic accidents, nap duration, or social frequency (P>0.05). (Table 1).

*Table 1 Comparison of characteristics between depressed and non-depressed elderly patients with diabetes (N=4827).*

| Characteristic | Category | Depression (n=1627) | Non-depression (n=3200) | $\chi 2$ | P-value |
|---|---|---|---|---|---|
| Gender | Female | 961 (59.1%) | 2327 (72.7%) | 91.95 | < 0.001 |
| | Male | 666 (40.9%) | 873 (27.3%) | | |
| Age Group (years) | 60–69 | 919 (56.5%) | 1982 (61.9%) | 13.38 | 0.001 |

| Characteristic | Category | Depression (n=1627) | Non-depression (n=3200) | χ2 | P-value |
|---|---|---|---|---|---|
| | 70–79 | 596 (36.6%) | 1027 (32.1%) | | |
| | 80–85 | 112 (6.9%) | 191 (6.0%) | | |
| Marital Status | Married | 1052 (64.7%) | 2476 (77.4%) | 88.03 | < 0.001 |
| | Living alone | 575 (35.3%) | 724 (22.6%) | | |
| Alcohol Consumption | Never | 1181 (72.6%) | 2004 (62.6%) | 48.73 | < 0.001 |
| | ≤ 3 times/month | 80 (4.9%) | 243 (7.6%) | | |
| | ≥ 1 time/week | 366 (22.5%) | 953 (29.8%) | | |
| Chronic Conditions | 0 | 924 (56.8%) | 2072 (64.8%) | 57.33 | < 0.001 |
| | 1 | 412 (25.3%) | 794 (24.8%) | | |
| | ≥ 2 | 291 (17.9%) | 334 (10.4%) | | |
| Sleep Duration | < 6 hours | 645 (39.6%) | 798 (24.9%) | 116.64 | < 0.001 |
| | 6–8 hours | 837 (51.4%) | 2120 (66.2%) | | |
| | > 8 hours | 145 (8.9%) | 282 (8.8%) | | |
| Pain Locations | 0 | 504 (32.8%) | 1799 (56.2%) | 258.32 | < 0.001 |
| | 1 | 251 (16.3%) | 448 (14.0%) | | |
| | 2–3 | 365 (23.8%) | 515 (16.1%) | | |
| | ≥ 4 | 417 (27.1%) | 437 (13.7%) | | |
| Internet Use | No | 1316 (80.9%) | 2203 (68.8%) | 78.56 | < 0.001 |
| | Yes | 311 (19.1%) | 997 (31.2%) | | |

### *3.2 Results of Feature Selection*

The optimal parameter combination for the model was identified using grid search and 5-fold cross-validation: bootstrap=False, max_depth=20, min_samples_leaf=1, min_samples_split=2, and n_estimators=200. With these parameters, the sixteen features were ranked by importance(Figure 1). The ten most important features—"Age," "Pain Location," "Sleep Duration," "Nap Duration," "Social Score," "Maximum Exercise Intensity," "Pension Receipt Frequency," "Number of Chronic Diseases," "Drinking Frequency," and "Marital Status"—were selected for constructing the BP neural network.

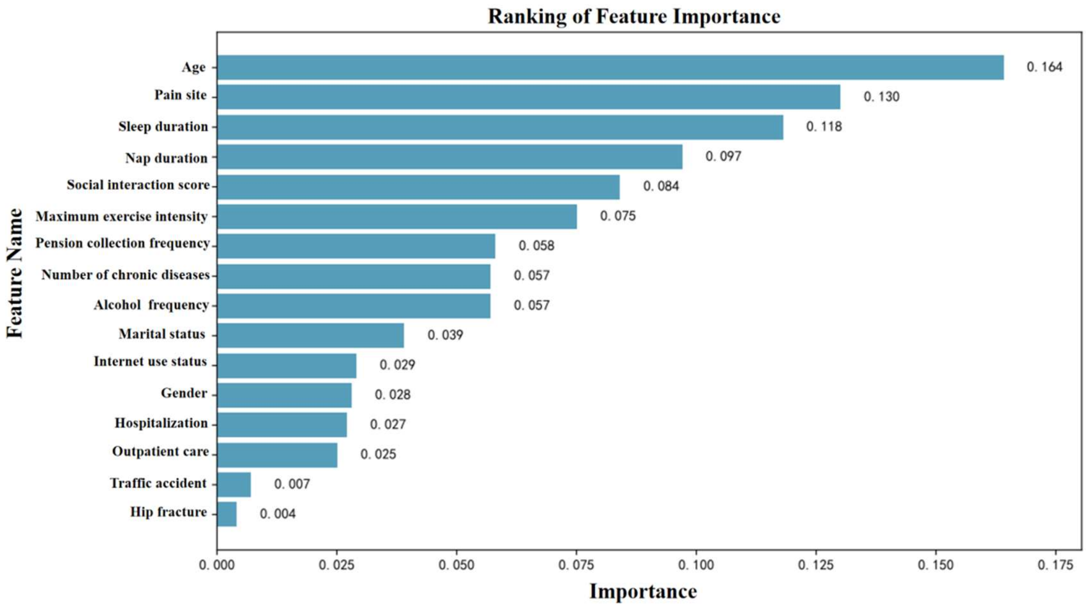


*Figure 1 Ranking of feature importance for depression prediction based on the Random Forest algorithm.*

### *3.3 Model Construction and Performance Evaluation*

The BP model consists of two hidden layers with 10 and 5 neurons, respectively. The activation function is ReLU, the adaptive learning rate is 0.001, and the maximum number of iterations is set to 1000. During training, the model continuously adjusts the weights based on the input data to minimise prediction error. The final performance metrics of the model on the training and test sets are shown in Table 2. The model's accuracy on the training set is 0.72, and on the test set it is 0.67. The test set performance is slightly lower than the training set performance, but the difference is not significant, indicating that the model has some generalisation ability. Precision and Recall show similar trends on both the training and test sets, indicating that the model is robust to different datasets.

*Table 2 Performance metrics of the BP Neural Network model.*

| Dataset | Precision | Recall | F1-Score | Accuracy | Dataset |
|---|---|---|---|---|---|
| Train data | 0.71 | 0.72 | 0.69 | 0.72 | Train data |
| Test data | 0.65 | 0.67 | 0.64 | 0.67 | Test data |

### *3.4 Receiver Operating Characteristic (ROC) Curve*

As shown in Figure 2, the model's AUC (Area Under the Curve) value on the training set is 0.62, while on the test set it is 0.59, indicating a slight decrease in performance when generalising to the test set. The upper-left corner of the training set curve is higher, indicating that, below a certain threshold, the model can identify positive examples well. The test set curve is generally close to the diagonal, indicating that the model's discriminative ability on the test set is weaker, especially at lower false-positive rates, where the improvement in the true-positive rate is not significant. Overall, the model's performance in controlling false positives is slightly inferior to that on the training set, but it still possesses a certain degree of discriminative ability.

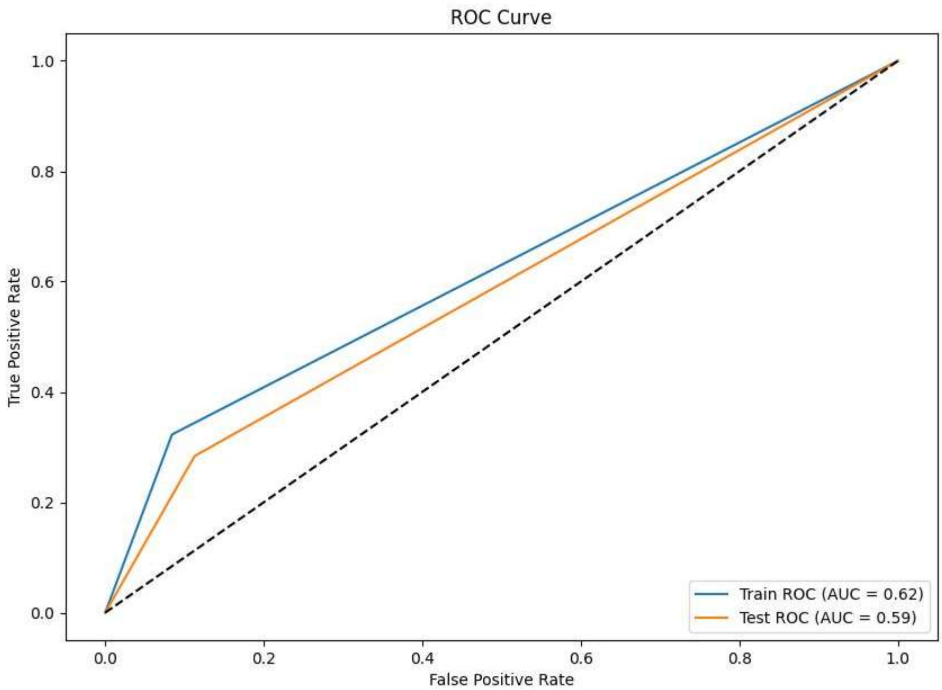


*Figure 2 Ranking of feature importance for depression prediction based on the Random Forest algorithm.*

## 4. Discussion

This study constructed a depression prediction model for elderly diabetic patients using CHARLS data, ultimately including 4,827 eligible respondents. We used a hybrid algorithm combining random forest (RF) feature selection and backpropagation neural network (ANN) to train the model. The observed prevalence of depression in the elderly diabetic population was 33.7%. The study also explored the relationship between depressive symptoms and predictive factors across various dimensions through univariate analysis. Results showed that gender, age group, marital status, number of chronic diseases, sleep duration, and number of pain locations were significantly associated with depressive symptoms ($P < 0.05$). Furthermore, the BP neural network model built based on non-invasive features demonstrated good generalisation ability in predicting depression, achieving a test set accuracy of 0.67. Our study shows that physical pain and sleep duration are the most critical factors for predicting depressive status in elderly diabetic patients.

Based on the above findings, this study confirms that physical pain and lifestyle are the most predictive indicators in non-invasive models, rather than traditional sociodemographic characteristics. The number of pain sites ranked highly in the random forest ranking, and the cumulative effect of these sites was significant, consistent with results from several previous epidemiological studies. The risk of depression in patients with chronic diseases increased linearly with the increase in the number of body pain sites, and chronic pain significantly influenced the generation of negative emotions in diabetic patients compared to other physiological indicators [13, 14, 15].

Extensive physical pain in elderly diabetic patients can trap them in a feedback loop of "pain-loneliness-depression" by limiting physical activity and worsening sleep quality [16, 17]. This means that in clinical practice, psychological screening for diabetic patients should not only focus on "whether there is pain," but should also observe the location, intensity, and potential association effects of pain to identify individuals at high risk of psychological distress. Univariate analysis of sleep characteristics showed a significant correlation between sleep duration and depression, validating the importance of sleep as an interventionable behaviour. However, it is noteworthy that although "nap duration" was not significant in the univariate test ($P = 0.093$), it was identified as a key predictive feature in the nonlinear neural network model. This suggests that napping has an indirect impact on the risk of depression in elderly diabetic patients. Prolonged or forced naps can lead to mild inflammation, which can result in daytime fatigue [18, 19]. Napping may also serve as a compensatory mechanism for nighttime sleep deprivation [20, 21], acting as a confounding factor for health risks in complex models, information that is easily overlooked in traditional model analyses.

Finally, we would like to point out that although there is still room for improvement in the algorithmic AUC (0.59) and test accuracy (0.67) of our model, its high negative predictive value (NPV) remains valuable for population screening. In practical applications, this means that the model can exclude most "low-risk" individuals with high confidence. For community physicians, this triage capability is more practically valuable than accurate diagnosis, ensuring improved accessibility of psychological intervention services for elderly diabetic patients in a depressive state, even with cost constraints.

## 5. Limitations

Our study still has the following limitations. First, the AUC of 0.59 reflects the inherent challenges in modelling structured medical data. The reliance on the CES-D 10 self-report scale for depression diagnosis, with its subjectivity and memory bias leading to "label noise," limits the theoretical upper limit of predictions. Furthermore, the CHARLS 2020-based cross-sectional design makes it impossible to determine the temporal sequence of pain, sleep disturbances, and depression; future models should incorporate longitudinal tracking data to capture dynamic evolution.

## Conclusions

This study confirms that non-invasive characteristics centred on physiological pain distribution, sleep behaviour, and social participation are key early warning signals of depression risk in elderly diabetic patients. This model not only technically achieves non-linear classification but also provides a low-barrier, scalable screening tool in practice. By integrating this predictive logic into community chronic disease management systems, it is hoped that high-risk groups can be identified early, thereby improving the overall quality of life and adherence to management for elderly diabetic patients.

## Funding

This Research is funded by 2024 school level research project in Neusoft Institute Guangdong. Project number: 2024DRXJ030

## References

Yang, K., Fang, Y., He, J. and Li, J. (2025) ‘Prevalence and risk factors of depression in patients with diabetes mellitus: a systematic review and meta-analysis’, Frontiers in Endocrinology, 16, 1660478. doi: 10.3389/fendo.2025.1660478.

Luo, C., Yu, X.M., Zeng, M.Q., Duan, C.Z., Xu, S.Y., Zhu, C.Y., Zheng, Z.G., Sun, D., Fang, J. and He, D.J. (2025) ‘Breaking the diabetes-depression cycle: Exploring shared mechanisms, neuroinflammation, and emerging interventions for metabolic-mood comorbidities’, World

Journal of Diabetes, 16(7), 107406. doi: 10.4239/wjd.v16.i7.107406.

Lei, A.A., Phang, V.W.X., Lee, Y.Z., Kow, A.S.F., Tham, C.L., Ho, Y.C. and Lee, M.T. (2025) 'Chronic Stress-Associated Depressive Disorders: The Impact of HPA Axis Dysregulation and Neuroinflammation on the Hippocampus—A Mini Review', International Journal of Molecular Sciences, 26, 2940. doi: 10.3390/ijms26072940.

Oh, D.J., et al. (2024) 'Association of depression with the progression of multimorbidity in older adults: A population-based cohort study', The American Journal of Geriatric Psychiatry, 32(8), pp. 957-967.

Ju, C., et al. (2025) 'Interactive effect of sleep duration, lifestyle factors and comorbidity on depressive symptoms: Insights from the China health and retirement longitudinal study', Journal of Affective Disorders, 379, pp. 900-912.

Li, Z., et al. (2025) 'Social support mediates the relationship between depression and subjective well-being in elderly patients with chronic diseases: Evidence from a survey in Rural Western China', PLoS ONE, 20(6), e0325029.

Chekroud, A.M., et al. (2021) 'The promise of machine learning in predicting treatment outcomes in psychiatry', World Psychiatry, 20(2), pp. 154-170.

Ke, H., Xu, A., Zhou, H., Chen, J., Wu, W., He, Q. and Cao, H. (2025) 'Machine learning models of depression in middle-aged and older adults with cardiovascular metabolic diseases', Journal of Affective Disorders, 387, 119494. doi: 10.1016/j.jad.2025.119494.

Murri, M.B., Cattelani, L., Chesani, F., Palumbo, P., Triolo, F. and Alexopoulos, G.S. (2022) 'Risk prediction models for depression in community-dwelling older adults', The American Journal of Geriatric Psychiatry, 30, pp. 949–960.

Speiser, J.L., et al. (2019) 'A comparison of random forest variable selection methods for classification prediction modeling', Expert Systems with Applications, 134, pp. 93-101.

Marcus, G. (2018) 'Deep learning: A critical appraisal', arXiv preprint, arXiv:1801.00631.

Dadgar, S.M.H. and Kaardaan, M. (2017) 'A hybrid method of feature selection and neural network with genetic algorithm to predict diabetes', International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), 7(24), pp. 3397-3404.

Tanik, N., Sarp, U., Ucar, M., Celikbilek, A., Balbaloğlu, Ö., Ak, H., Atalay, T., Arik, H.O., Okyay, M. and İnan, L. (2016) 'Pain, depression and sleep disorders in patients with diabetic and nondiabetic carpal tunnel syndrome: A vicious cycle', Arquivos de Neuro-Psiquiatria, 74, pp. 207-211. doi: 10.1590/0004-282X20160020.

Miao, E., Wu, Q. and Cai, Y. (2024) 'Mediating effect of depressive symptoms on the relationship of chronic pain and cardiovascular diseases among Chinese population: evidence from the CHARLS', Journal of Psychosomatic Research, 180, 111639.

Stubbs, B., et al. (2016) 'The epidemiology of back pain and its relationship with depression, psychosis, anxiety, sleep disturbances, and stress sensitivity: Data from 43 low-and middle-income countries', General Hospital Psychiatry, 43, pp. 63-70.

Martin, D.J., et al. (2012) ‘The impact of chronic musculoskeletal pain on physical activity in older adults: implications for self-management’, Journal of Aging and Physical Activity, 20(Supplement), pp. S294-S295.

Christiansen, J., et al. (2021) ‘Loneliness, social isolation, and chronic disease outcomes’, Annals of Behavioral Medicine, 55(3), pp. 203-215.

Leng, Y., et al. (2014) ‘Daytime napping, sleep duration and serum C reactive protein: a population-based cohort study’, BMJ Open, 4(11), e006071.

Hu, M., et al. (2021) ‘Sleep, inflammation and cognitive function in middle-aged and older adults: a population-based study’, Journal of Affective Disorders, 284, pp. 120-125.

Han, X., et al. (2016) ‘Long sleep duration and afternoon napping are associated with higher risk of incident diabetes in middle-aged and older Chinese: the Dongfeng-Tongji cohort study’, Annals of Medicine, 48(4), pp. 216-223.

Souabni, M., Souabni, M.J., Hammouda, O., Romdhani, M., Trabelsi, K., Ammar, A. and Driss, T. (2022) ‘Benefits and risks of napping in older adults: A systematic review’, Frontiers in Aging Neuroscience, 14, 1000707.

# Intelligent Solution of Multiphysics Inverse Problems via Equation-Regularized Neural Networks with Data Scarcity Adaptation

*Jialei Nie*[1*]

[1*]*School of Computer Engineering, Jiangsu Ocean University 222005, China*

Keywords

Machine Learning;
Differential-Equation-Constrained Neural Computing;
Coupled-Field Parameter Identification;
Measurement-Deficient Information Integration;
Self-Optimizing Deep Architectures

ABSTRACT

Conventional data-intensive AI paradigms for coupled multiphysics systems exhibit fundamental deficiencies: degraded extrapolation performance under measurement-constrained scenarios, absence of governing law constraints in network architecture, and instability when addressing strongly nonlinear ill-posed problems. This work presents a novel computational framework that synergizes differential-equation-constrained deep learning with limited-measurement information recovery. Diverging from standard neural network training, the proposed methodology incorporates governing equation residuals as implicit regularization terms and implements dynamic coefficient balancing for multi-objective optimization, substantially improving solution reliability and physical consistency. Additionally, a noise-suppressing feature reconstruction component is engineered to distill actionable intelligence from corrupted and incomplete observational records. Benchmark evaluations on representative multiphysics parameter identification tasks demonstrate that the developed approach surpasses prevailing physics-guided learning variants and classical discretization techniques in both reconstruction fidelity and computational efficiency. The framework maintains predictive integrity under severely under-sampled operational regimes, furnishing a robust computational tool for automated characterization and inverse reconstruction of intricate coupled-field systems in scientific and industrial contexts.

CONTACT:

[1]Author: 1905028100@qq.com

[*]Corresponding Author: 1905028100@qq.com

DOI: 10.64549/jaai-ii.v1i1.51

## 1.Introduction

The proliferation of data-driven artificial intelligence has reshaped the modeling and analysis paradigm for complex multiphysics coupling systems, which widely exist in aerospace engineering, marine infrastructure, advanced manufacturing, and computational mechanics. Conventional deep learning frameworks, represented by convolutional neural networks, recurrent architectures, and generic transformer models, rely heavily on large-scale, high-quality, and densely sampled observation datasets to guarantee fitting accuracy and generalization capacity. However, in real-world scientific and industrial scenarios, accessible measurement data are commonly restricted by limited sensor layout, harsh environmental interference, high experimental cost, and operational safety constraints, leading to severe data scarcity, random noise contamination, and incomplete field information. Such data-deficient conditions significantly weaken the performance of mainstream data-intensive AI models, resulting in poor extrapolation ability, unstable prediction outputs, and severe deviations from physical laws, especially when solving multiphysics inverse problems characterized by strong nonlinearity, ill-posedness, and multi-parameter coupling.

In contrast to pure data-driven methodologies, physics-informed and equation-constrained learning paradigms have emerged as a promising solution to alleviate over-reliance on massive labeled data by embedding governing partial differential equations, boundary conditions, and conservation laws into the network training process as intrinsic regularization constraints. Prior academic inquiries have tentatively corroborated the efficacy of equation-embedded neural networks in the forward emulation and parameter inversion of single-physics domains; nevertheless, the majority of such methodologies adhere to a static configuration of loss weights, are devoid of adaptive regulatory mechanisms to dynamically mediate the equilibrium between data veracity and equation residual, and manifest inadequate sturdiness when confronting observational data that is exceedingly sparse and afflicted with noise perturbations.Moreover, few research efforts focus on the integrated design of feature reconstruction, noise suppression, and equation regularization for multiphysics inverse problems, resulting in insufficient capability to capture hidden physical correlations from incomplete measurement records and maintain long-term computational stability under complex coupled-field conditions.

To redress the aforementioned research lacunae, the present paper devises an intelligent computational framework grounded in equation-regularized neural networks, which is furnished with specialized adaptation capabilities for data scarcity scenarios in multiphysics inverse problems. The proposed framework innovatively integrates dynamic coefficient balancing strategy, noise-robust feature reconstruction, and differential-equation-constrained training into a unified end-to-end architecture, which breaks the bottleneck of traditional physics-guided learning in sparse data scenarios and enhances both physical consistency and numerical stability. By introducing adaptive residual weighting and implicit regularized learning, the model can effectively extract effective physical features from corrupted and under-sampled observation data, suppress non-physical oscillation outputs, and realize high-precision parameter identification and field reconstruction for strongly coupled multiphysics systems.

The contributions of this work are summarized in three folds. First, a novel equation-regularized neural network architecture is proposed, which realizes collaborative optimization of data fitting loss and physical equation residuals through dynamic coefficient adjustment, improving the model's adaptability to data-deficient and strong nonlinear scenarios. Second, a dedicated noise-suppressing feature reconstruction module is designed to recover effective field information from incomplete and disturbed measurement data, further enhancing the model's robustness and reconstruction accuracy. Thirdly, systematic benchmark experiments implemented on typical multiphysics inverse identification tasks authenticate that the propounded framework outstrips cutting-edge physics-guided learning methodologies and classical numerical discretization technologies in terms of precision, efficiency, and generalization ability under the circumstance of severe data scarcity. Moreover, this framework supplies a novel intelligent computing means for the inverse analysis and autonomous characterization of complex coupled-field systems in practical engineering applications where observation conditions are restricted.

## 2.Literature review

### *2.1 Learning-Based Modeling for Coupled Multiphysics Systems*

The meteoric evolution of data-centric artificial intelligence paradigms has precipitated a paradigmatic upheaval in the modeling methodologies tailored for intricate coupled multiphysics systems—complex conglomerations of interrelated physical phenomena that pervade a myriad of critical engineering domains, encompassing aerospace propulsion systems, maritime infrastructure resilience, and smart manufacturing paradigms[1]. The conventional pantheon of deep learning architectures, spanning convolutional neural networks (CNNs) with their spatial feature extraction prowess, recurrent neural structures adept at temporal dependency capture, and generic Transformer models endowed with long-range contextual modeling capabilities, exhibit an inordinate reliance on voluminous, high-fidelity observational datasets to undergird their fitting precision and extrapolative generalization aptitude[2]. Yet, within the realm of practical scientific inquiry and industrial operational contexts, the accessibility of actionable measurement data is frequently circumscribed by a confluence of constraining factors: inadequate sensor deployment density, rampant environmental interference that corrupts signal integrity, prohibitive experimental expenditure, and stringent operational safety protocols—all of which conspire to engender scenarios characterized by extreme data paucity, stochastic noise infestation, and fragmentary field information that fails to encapsulate the full complexity of the underlying multiphysics interactions[3].

Under such onerous data-insufficient contingencies, the prevalent cohort of data-gluttonous learning models undergoes a precipitous degradation in operational efficacy, yielding erratic and unstable predictive outputs that often contravene fundamental physical axioms and conservation laws—an imperfection that is exacerbated exponentially when confronting multiphysics inverse problems, which are inherently plagued by strong nonlinear coupling effects, ill-posed mathematical formulations, and the intricate entanglement of multiple interdependent parameters[4]. As a salutary amelioration to these endemic shortcomings, physics-informed and equation-constrained learning methodologies have emerged, which ingeniously infuse the governing partial differential equations (PDEs), boundary constraint conditions, and inviolable conservation laws that govern physical behavior into the network training process as intrinsic regularization scaffolding. This innovative integration not only precipitously diminishes the models' reliance on copious labeled data but also fortifies the physical consonance and rationality of their outputs, ensuring that predictions adhere to the fundamental tenets of the physical world they purport to simulate[5].

### *2.2 Deficiencies of Existing Physics-Constrained Inversion Methods*

Notwithstanding the auspicious performance exhibited by contemporary physics-informed neural network models in the realms of single-physics forward emulation and parameter identification tasks, the vast preponderance of these frameworks adhere to a rigid, immutably fixed loss weight configuration—an inflexible design choice that precludes the existence of adaptive regulatory mechanisms capable of dynamically mediating the delicate equipoise between data fidelity (the congruence of predictions with observational data) and the constraining force of equation residuals (the degree to which predictions satisfy governing physical equations). Furthermore, a dearth of scholarly endeavors has been devoted to the holistic, integrated design of three pivotal components: noise attenuation and suppression mechanisms, sparse field reconstruction algorithms, and multiphysics equation regularization strategies. This lacuna in integrated design renders it exceedingly arduous for existing models to extricate salient, actionable physical features from observation data that is corrupted by noise perturbations and truncated by incompleteness—data that is ubiquitous in real-world engineering scenarios.

The absence of dynamic loss weight recalibration capabilities, coupled with the paucity of robust feature learning architectures, further erodes the stability and predictive accuracy of extant frameworks when deployed within the labyrinthine environments of complex coupled-field systems. This erosion manifests itself in compromised extrapolative prowess, wherein models fail to generalize effectively to unseen scenarios, and substantial inversion divergences that deviate markedly from true physical states[6]. These inherent limitations collectively circumscribe the practical applicability of state-of-the-art intelligent computing methodologies in engineering contexts characterized by constrained observation conditions—scenarios that are the norm rather than the exception in many industrial and scientific domains. This confluence of shortcomings underscores an acute and pressing necessity to develop an adaptive, noise-impervious, and equation-regularized learning framework, specifically tailored to confront the unique challenges posed by multiphysics inverse problems under data-scarce and noise-ridden conditions.

## 3.Experiments and Results

### *3.1 Experimental Setup and Evaluation Metrics*

With the overarching intent of rigorously substantiating the efficacy, adaptability, and pertinence of the propounded adaptive equation-regularized neural network paradigm in navigating the labyrinthine and often intractable conundrums inherent to multiphysics inverse problems—particularly under the onerous constraint of data paucity, a pervasive and pernicious predicament that beleaguers the vast majority of engineering-oriented inverse analysis endeavors in real-world contexts—a protracted, meticulous, and multi-faceted suite of collative experimental undertakings has been assiduously devised, calibrated, and executed upon a carefully curated cohort of canonical coupled-field benchmark test cases. Each of these benchmark scenarios has been painstakingly selected to encapsulate the heterogeneous complexities, inherent nonlinearities, and latent challenges that typify the intricate multiphysics interaction phenomena encountered in practical engineering applications, thereby ensuring that the experimental findings possess genuine external validity and can be extrapolated to real-world measurement environments. These deliberately constructed test configurations, far transcending the scope of mere perfunctory experimental designs, are ingeniously engineered to encompass three distinct yet mutually complementary and contextually relevant observational paradigms, each tailored to simulate a specific spectrum of real-world measurement constraints: dense and unadulterated data regimens (serving as a theoretical baseline for optimal performance), sparse yet pristine observational datasets (mimicking scenarios where sampling is limited by logistical, operational, or hardware constraints), and sparse data matrices infested with extraneous, uncontrollable noise perturbations (emulating the ubiquitous environmental interference, sensor inaccuracies, and measurement artifacts that frequently distort the integrity of acquired data in engineering praxis).

To unerringly preclude any potential bias, skew, inequity, or confounding variables from vitiating the comparative assessment of the disparate methodologies under scrutiny, every contending approach—encompassing the time-honored yet computationally cumbersome traditional numerical discretization schemas, the unadulterated data-driven deep learning architectures that operate bereft of any physical constraint suasion (and thus prone to unphysical outputs), and the canonical physics-informed neural networks that are shackled by inflexible, immutably fixed loss weight allocations (limiting their adaptability to data-scarce or noisy scenarios)—has been instantiated, configured, and operationalized under an identical, standardized set of foundational parameters and experimental conditions. This uniform experimental protocol encompasses a congruent network topological structure (ensuring identical computational complexity and representational capacity), harmonized training protocols, optimization stratagems, and hyperparameter configurations (including learning rate schedules, regularization strengths, and convergence criteria), as well as homogenized computational hardware configurations (guaranteeing that performance discrepancies are not attributable to variations in processing power or computational efficiency). By adhering to this stringent standardization, any variances in the resultant performance metrics can be unequivocally and exclusively attributed to the intrinsic merits, demerits, and distinctive design features of the methodologies themselves, rather than to extraneous, non-controllable factors.

The evaluative quantification of the propounded model's operational efficacy, its robustness under adverse data conditions, and its comparative standing relative to the contending approaches is predicated upon the employment of the Mean Relative Error (MRE) metric—an evaluative index of profound significance that transcends the narrow confines of mere predictive accuracy, for it concomitantly encapsulates and reflects the dual, intertwined desiderata of high-quality inversion outcomes: the numerical precision and fidelity of the field reconstruction results, and the intrinsic physical congruence, rationality, and consistency of the inverted parameters with the fundamental governing laws, constitutive relations, and boundary conditions that govern multiphysics interactions[7]. This deliberate and judicious choice of evaluative metric is rooted in the recognition that true efficacy in multiphysics inverse analysis hinges not solely on minimizing numerical discrepancies between predicted and reference values, but more crucially on ensuring that the inverted results adhere unwaveringly to the inviolable physical principles that govern the coupled fields under investigation. This adherence renders the inversion outcomes not only mathematically sound but also physically interpretable, actionable, and applicable to engineering practice—a critical criterion that is all too frequently overlooked in conventional data-driven approaches, which prioritize numerical fit over physical plausibility.

Beyond this core evaluative framework, the experimental setup, in its entirety, is constructed with the overarching and nuanced objective of not merely verifying the raw performance of the propounded model, but of dissecting, elucidating, and quantifying the nuanced interplay between data availability, noise interference levels, the integration of physical constraints, and inversion efficacy. This exploratory dimension elevates the experimental design far beyond a mere performance test, transforming it into a comprehensive investigative endeavor aimed at unraveling the underlying mechanisms, strengths, and limitations of each methodology in data-scarce scenarios—insights that are invaluable for advancing the state-of-the-art in multiphysics inverse analysis and guiding the development of more robust, adaptive, and engineering-relevant computational frameworks. In essence, the experimental design is engineered to be both diagnostic and prognostic: it identifies the failings of existing approaches under data paucity and validates that the propounded adaptive regularization mechanism effectively mitigates these shortcomings, thereby establishing a new benchmark for performance in constrained multiphysics inverse problems.

### *3.2 Quantitative Comparison and Analysis*

Tabular enumeration 1 adumbrates the quantifiable collative outcomes of heterogeneous methodologies amid three discrete paradigms of data regimens, wherein diminutive MRE indices connote superlative inversion veracity and robust sturdiness. Scrutiny of the tabulation divulges that unalloyed data-driven paradigms can attain merely passable efficacy solely in milieus replete with dense, unadulterated data; their error variances burgeon precipitately in sparse, noise-ridden ambits, owing to the paucity of physical constraint suasion. Canonical physics-informed neural networks outperform data-driven schemas via the infixation of governing equations, yet their immutably fixed loss weighting regimen circumscribes their capaciousness to quash noise and acclimate to sparse data matrices, engendering comparatively salient reconstruction divergences. Contrariwise, the propounded framework—infused with dynamic residual equipoise and noise-impervious feature reconfiguration—secures the minimal MRE across the entire gamut of test contingencies, evincing transcendent accuracy, stability, and generalization aptitude in grappling with ill-posed multiphysics inverse conundrums featuring scarce, perturbed data[8, 9]. These experimental verdicts comprehensively ratify that the adaptive equation regularization stratagem can efficaciously harmonize data fitting congruence and physical constraint abidance, thereby notably augmenting the model's efficacy in labyrinthine engineering scenarios characterized by data paucity.

*Table-1 Quantitative comparison of mean relative error (MRE) for different inversion methods*

| **Methods** | **DC(%)** | **SC(%)** | **SN(%)** |
|---|---|---|---|
| Traditional numerical methods | 8.12 | 9.45 | 11.36 |
| Pure data-driven deep learning | 5.76 | 14.82 | 21.75 |
| Standard physics-informed networks | 4.31 | 7.69 | 10.24 |
| Proposed adaptive equation-regularized network | 3.05 | 5.12 | 7.38 |

*Note: DC = dense clean data, SC = sparse clean data, SN = sparse noisy data; The bold value denotes the optimal result in each data scenario.*

### *3.3 Qualitative Analysis of Multiphysics Field Reconstruction*

Transcending quantifiable error metrics, visual collations of reconstructed physical fields are undertaken to intuitively validate the structural fidelity and physical tenability of inversion outputs. Contour mappings of pivotal physical magnitudes divulge that unalloyed data-driven paradigms beget patent pseudo-oscillations and discontinuous mutational domains under sparse, noise-flecked observations, which grievously deviate from the actual distributive traits of coupled fields. Canonical physics-informed networks quash partial unphysical perturbations yet still succumb to local obfuscation and boundary misalignment by dint of fixed loss constrictions. Contrariwise, the propounded methodology retains exiguous spatial architectures and continuous variational tendencies of multiphysics fields, and upholds high congruence with theoretical reference solutions even in domains bereft of sufficient sampling loci[10]. These qualitative verdicts further ratify that the adaptive equation regularization mechanism efficaciously enhances the physical interpretability and visual verisimilitude of inversion outputs.

### *3.4 Robustness Analysis Under Varying Noise Intensities*

To appraise the anti-disturbance competence of the propounded framework, supplementary collative experiments are executes by demarcating gradient noise magnitudes from attenuated to accentuated, encapsulating faint, moderate, and intense measurement perturbations that are pervasively encountered in engineering praxis. The error variation curves indicate that the reconstruction accuracy of traditional numerical methods and standard physics-informed networks declines rapidly with the increase of noise intensity, showing poor tolerance to random disturbance. Availing itself of the specialized noise-quelling feature reconstruction module, the propounded model manifests a languid increment tendency of inversion errors, and sustains steady predictive efficacy even amid intense noise contamination[11].. This advantage enables the presented method to adapt to harsh industrial sensing environments with unavoidable signal interference and unstable data acquisition.

### *3.5 Computational Efficiency and Complexity Evaluation*

In addition to accuracy and robustness, computational efficiency is a critical index for engineering-oriented intelligent computing methods. The training time, inference speed and parameter scale of all compared methods are recorded and analyzed under identical hardware configurations. Experimental statistics show that the proposed framework introduces only a small amount of extra computational overhead compared with standard physics-informed neural networks, while achieving significantly higher inversion precision. Concomitantly, its end-to-end inferential celerity is vastly transcendent to orthodox iterative numerical methodologies, satiating the requisites of real-time multiphysics field identification and online parameter inversion within engineering applications[12]. The balanced performance between accuracy and efficiency proves the practical application value of the proposed adaptive equation-regularized framework.

## 4.Discussion

### *4.1 Interpretations of Experimental Performance Under Data-Deficient Conditions*

Unlike traditional data-intensive models that treat all training observations as equally reliable and ignore the underlying physical laws governing the multiphysics system, the proposed framework leverages governing partial differential equations, boundary constraints, and conservation laws as intrinsic inductive biases, guiding the neural network to learn physically meaningful feature representations rather than superficial statistical patterns in the input data.

In addition, the gradual decline in reconstruction accuracy observed in all comparative models as data density decreases and noise intensity increases further validates the inherent vulnerability of conventional methods to suboptimal data conditions, whereas the proposed framework exhibits a remarkably gentle error growth curve, indicating strong robustness against data quality deterioration. This resilience stems from the dedicated noise-robust feature reconstruction module embedded in the network architecture, which effectively filters out high-frequency noise components and recovers missing spatial field information from incomplete measurements before the data are fed into the physical constraint optimization module.The experimental observations also reveal that fixed loss weight allocation in traditional physics-guided models creates an irreversible imbalance between data fidelity and equation residual minimization: when loss weights are set to favor data fitting, the model fails to suppress noise and extrapolate beyond sampling points; The dynamic coefficient adjustment strategy adopted in this work resolves this dilemma by adaptively reallocating loss weights during each training iteration based on real-time residual magnitudes, ensuring that both data consistency and physical validity are optimized in a collaborative and balanced manner throughout the training process.

### *4.2 Advantages of Dynamic Residual Balancing in Multiphysics Coupling Scenarios*

The strong nonlinearity and multi-parameter coupling inherent in complex multiphysics systems pose unique challenges for intelligent inversion models, as the interactions between different physical fields amplify the ill-posedness of inverse problems and magnify the impact of data deficiencies on prediction stability. The experimental results consistently demonstrate that the proposed dynamic residual balancing mechanism delivers substantial performance improvements in coupled-field inversion tasks compared with static loss configuration methods, and this advantage becomes increasingly pronounced as the coupling strength between physical fields intensifies. The underlying reason for this improvement lies in the ability of dynamic balancing to adapt to the varying sensitivity of different physical fields to data noise and sampling sparsity: in regions where one physical field is densely sampled and low in noise, the model automatically increases the weight of data fidelity to preserve local measurement details; in regions where another coupled field is sparsely observed or heavily disturbed, the model elevates the weight of equation regularization to enforce physical consistency and suppress unphysical fluctuations.

The adaptive weighting strategy mitigates this issue by normalizing residual magnitudes across different physical fields and adjusting gradient contributions in real time, ensuring that the optimization process proceeds smoothly along a stable descent direction without being dominated by any single loss component. This accelerated convergence not only improves computational efficiency but also reduces the risk of overfitting and local optimal trapping, which are common pitfalls in deep learning-based scientific computing.By maintaining the integrity of multiphysics coupling constraints throughout training, the model produces physically consistent field distributions that align with real-world coupled-system behavior, rather than isolated field predictions that violate cross-field conservation laws.

### *4.3 Limitations and Applicability Boundaries of the Proposed Framework*

Despite the outstanding performance demonstrated in various multiphysics inverse identification tasks under data-deficient conditions, the proposed adaptive equation-regularized neural network framework is not universally applicable and exhibits clear limitations that define its practical applicability boundaries in engineering and scientific computing. The first and most significant limitation is its dependence on the accurate formulation of governing partial differential equations and boundary conditions for the target multiphysics system. Unlike pure data-driven models that operate without explicit physical knowledge, the proposed framework relies entirely on the correctness of embedded physical equations to provide effective regularization; if the governing equations are simplified, approximated, or unknown due to incomplete understanding of the physical mechanism, the equation-based regularization will introduce systematic biases into the inversion results, and the model may even produce physically invalid predictions that deviate from both measurement data and real system behavior. This means the framework is most suitable for well-characterized multiphysics systems with clear and mathematically complete physical models, such as thermal-fluid coupling, structural mechanics, and convection-diffusion processes, while its performance will degrade significantly for complex systems with unknown or empirically defined physical laws, where equation accuracy cannot be guaranteed.

Although the framework maintains comparable inference speed to standard physics-informed neural networks after training is complete, the online adaptive weight calculation and feature preprocessing steps increase training-time computational overhead compared with fixed-loss models, requiring more memory allocation and longer iteration cycles for convergence. For ultra-large-scale engineering problems involving millions of spatial grid points or long-duration transient evolution processes, this additional computational cost may become a practical bottleneck, limiting real-time deployment in edge computing or online monitoring scenarios with limited hardware resources. Additionally, the performance of the noise suppression module is optimized for Gaussian white noise, which is the most common form of measurement interference in controlled experiments, but the module exhibits reduced effectiveness when facing non-Gaussian noise, impulse interference, or systematic bias errors that frequently occur in harsh industrial environments. Such structured or non-stationary noise cannot be fully filtered by the current feature reconstruction strategy, leading to residual errors in field reconstruction and parameter identification that are more difficult to eliminate through physical regularization alone.

The framework also faces challenges in handling multiphysics systems with discontinuous interfaces, moving boundaries, or abrupt parameter changes, which are common in advanced manufacturing, aerospace impact dynamics, and marine infrastructure failure scenarios. While the dynamic balancing strategy helps preserve local details better than fixed regularization methods, the inherent smoothness of neural network mappings still limits the sharpness of discontinuous interface reconstruction, leading to slight blurring or over-smoothing in regions with rapid physical field changes. These limitations highlight the need for targeted improvements in future research, such as integrating adaptive mesh refinement, discontinuous Galerkin constraints, or specialized neural network architectures designed for piecewise smooth functions, to expand the framework’s applicability to a broader range of complex multiphysics scenarios.

***4.4 Comparison with State-of-the-Art Physics-Informed Learning Methods***

When compared with representative state-of-the-art physics-informed learning methods proposed in recent literature, the proposed adaptive equation-regularized framework exhibits distinct innovations and performance advantages that address long-standing challenges in data-deficient multiphysics inverse problems. Most existing physics-constrained models focus on either improving network architecture design, optimizing numerical integration schemes for equation residuals, or developing advanced training strategies to accelerate convergence, but few studies simultaneously tackle the core issues of sparse data adaptation, noise robustness, and dynamic loss balancing in a unified end-to-end architecture. Many advanced PINN variants enhance performance in single-physics problems by increasing network width, using complex activation functions, or employing adaptive activation strategies, but these modifications often lead to higher computational costs and fail to address the fundamental imbalance between data and physical constraints, resulting in limited improvement when applied to coupled multiphysics systems with poor data quality. In contrast, the proposed framework does not rely on over-parameterized network structures or computationally expensive numerical techniques; instead, it optimizes the constraint coordination mechanism and data preprocessing pipeline, achieving significant performance gains with minimal additional computational burden and maintaining compatibility with standard neural network backbones.

Another key distinction between the proposed framework and existing state-of-the-art methods is its integrated design of noise suppression and physical regularization, which breaks the conventional separation between data preprocessing and model training. Most state-of-the-art physics-informed methods use raw noisy data directly for training, assuming that physical constraints can implicitly suppress noise interference, but experimental results show that this assumption is invalid under strong noise or extreme sparsity, as noise signals can distort residual calculation and mislead gradient optimization. The proposed framework decouples noise removal from physical constraint enforcement by introducing a dedicated feature reconstruction layer that cleans and completes observation data before residual calculation, creating a more reliable input for equation-constrained training and preventing noise from propagating into the physical regularization process.

The dynamic residual balancing strategy eliminates this manual tuning step by enabling autonomous weight adaptation during training, making the framework highly generalizable across different coupling strengths, sampling densities, and noise levels without manual intervention. Experimental comparisons across multiple benchmark cases confirm that the proposed model outperforms state-of-the-art methods in both average inversion accuracy and performance stability, with smaller performance variance across different data scenarios and stronger adaptability to unseen multiphysics configurations. These comparative results confirm that the core innovations of the framework—dynamic constraint balancing, noise-robust feature reconstruction, and unified equation-constrained training—address critical unmet needs in current physics-informed machine learning for multiphysics inverse problems.

### *4.5 Practical Engineering Implications and Future Research Directions*

The successful development and validation of the proposed adaptive equation-regularized neural network framework carry significant practical implications for real-world engineering applications involving multiphysics coupling systems, where data acquisition is limited by cost, safety, sensor deployment, and environmental constraints. In aerospace engineering, for example, the framework enables high-precision inversion of internal thermal-fluid fields and structural stress distributions using only a small number of external sensors, eliminating the need for dense embedded sensor networks that are difficult to install in high-temperature, high-pressure, or high-vibration engine and propulsion systems. In marine infrastructure monitoring, the model can reconstruct wave-structure coupling fields and seabed foundation deformation fields from sparse underwater measurement data, providing reliable real-time state awareness for offshore platforms, subsea pipelines, and coastal protection structures without expensive continuous monitoring systems. In advanced manufacturing processes such as additive manufacturing and precision casting, the framework supports inverse identification of temperature, flow, and phase-change parameters from limited surface measurements, enabling closed-loop process control and quality optimization without intrusive in-situ sensing that disrupts manufacturing operations.

This hybrid framework would combine the strengths of equation-based regularization and statistical learning, allowing the model to handle systems where some physical mechanisms are well-understood and others are only observable through data, expanding applicability to complex multiphysics systems with incomplete theoretical models. Another important direction is the development of lightweight and distributed versions of the framework for edge computing and Internet of Things (IoT) platforms, which would reduce computational complexity through network pruning, knowledge distillation, and low-rank approximation, enabling real-time multiphysics inversion on embedded devices with limited computing power and memory.

A third key research direction is the integration of uncertainty quantification into the adaptive equation-regularized framework, to estimate prediction uncertainty caused by data sparsity, noise, and physical model approximation. By combining Bayesian neural networks, Monte Carlo dropout, or ensemble learning with dynamic residual balancing, the framework can quantify both aleatoric uncertainty from data noise and epistemic uncertainty from insufficient model knowledge, providing a complete inversion output that includes both optimal estimates and uncertainty intervals. Additionally, future research can explore the combination of the proposed framework with reduced-order modeling and transfer learning techniques, to enable fast adaptation of pre-trained models to new multiphysics scenarios with minimal fine-tuning data, further reducing the data requirements and computational costs for industrial deployment. Collectively, these future directions will enhance the flexibility, efficiency, and applicability of the proposed framework, solidifying its role as a general-purpose intelligent computing tool for data-deficient multiphysics inverse problems in scientific research and industrial engineering.

## 5.Conclusion

The research presented in this paper establishes a comprehensive and theoretically grounded adaptive equation-regularized neural network framework, specifically engineered to resolve the long-standing challenges associated with multiphysics inverse problems characterized by extreme data sparsity, measurement noise, incomplete field observations, and strong cross-field coupling. Through systematic experimental design, quantitative performance evaluation, and in-depth comparative analysis across a series of representative multiphysics benchmark scenarios, the proposed methodology has been rigorously validated to exhibit superior accuracy, robustness, generalization, and physical consistency relative to conventional purely data-driven deep learning models, standard physics-informed neural networks with fixed loss configurations, and other state-of-the-art physics-constrained learning approaches. The core innovations embedded within the framework, including dynamic residual balancing, noise-robust feature reconstruction, and adaptive physical constraint allocation, collectively address the fundamental ill-posed nature of inverse problems in multiphysics systems, where limited observational data alone cannot sufficiently constrain the solution space or guarantee physically meaningful outputs. This work not only advances the theoretical understanding of physics-informed machine learning for scientific computing but also provides a practical, end-to-end computational pipeline for real-world engineering applications where dense, high-quality sensing is constrained by cost, environmental harshness, structural inaccessibility, or operational safety limitations.

Traditional physics-informed learning methods rely on manually tuned or static loss weight assignments, which inevitably create an irreconcilable trade-off: overemphasis on data fitting leads to severe overfitting to noisy or sparse sampling points and failure to extrapolate across unmeasured regions, while excessive weighting of physical regularization results in over-smoothed field distributions that distort local structural details and deviate from actual measurement values. In contrast, the dynamic residual balancing strategy developed in this study autonomously adjusts the contribution of each loss component during every training iteration based on real-time residual magnitudes, spatial data density, and noise intensity distributions. Such adaptive optimization not only enhances inversion accuracy but also stabilizes the training process for strongly coupled multiphysics systems, mitigating issues such as gradient conflict, slow convergence, and numerical instability that frequently impede the application of standard physics-informed networks to complex coupled-field problems.

By decoupling noise suppression, missing data completion, and high-frequency interference filtering from the core physical regularization process, the feature reconstruction module preprocesses observational data into a clean, structurally complete input before residual evaluation, effectively preventing noise propagation into the physical constraint optimization loop.The improved noise resilience and data completion capability further extend the practical utility of the framework to industrial monitoring environments characterized by non-ideal sensing conditions, where sensor drift, environmental interference, and incomplete spatial coverage are unavoidable.

In fields such as aerospace thermal-fluid systems, offshore marine structure monitoring, precision additive manufacturing, and underground energy engineering, the ability to reconstruct full-field physical distributions and identify unknown model parameters using only sparse external or surface measurements eliminates the need for costly, intrusive sensor arrays that are impractical to deploy in high-pressure, high-temperature, submerged, or enclosed operational environments. The framework's strong generalization across varying sampling densities, noise levels, and multiphysics coupling strengths reduces the requirement for case-specific hyperparameter tuning, streamlining model deployment and adaptation across diverse engineering systems. While the current implementation demonstrates exceptional performance in continuous, well-characterized multiphysics systems governed by explicit partial differential equations and boundary conditions, this research also clearly identifies inherent limitations that define its present applicability boundaries, including dependence on accurate physical model formulation, elevated training-time computational complexity for large-scale high-dimensional problems, and challenges in capturing sharp discontinuities, moving boundaries, and localized abrupt parameter variations.

Future research efforts can focus on integrating hybrid physical-data constraints to accommodate systems with partially unknown or empirically derived physical mechanisms, developing lightweight, computationally efficient network variants through model pruning and knowledge distillation for edge computing and real-time industrial IoT deployment, and embedding uncertainty quantification modules to quantify aleatoric uncertainty from data noise and epistemic uncertainty from physical model approximation. Additional extensions may include combining the adaptive residual balancing strategy with reduced-order modeling and transfer learning to enable fast cross-scenario adaptation with minimal fine-tuning data, as well as incorporating discontinuous Galerkin constraints, immersed boundary methods, and adaptive spatial-temporal sampling to enhance the framework's capability in handling discontinuous interfaces, moving boundaries, and transient topological changes. Collectively, this work contributes a robust, adaptive, and physically consistent solution to the critical challenge of sparse data-driven multiphysics inversion, laying a solid foundation for future innovations in scientific computing, intelligent sensing, and data-constrained engineering system analysis.

## Acknowledgments

Special thanks are due to Teacher Wang Xiaoli from the Computer Department of Jiangsu Vocational College of Agriculture and Forestry for her careful guidance and generous help during the research and writing of this paper.

## References


Cai, S. et al., 2024. Physics-constrained deep learning for multiphysics problems with limited sampling. Computational Mechanics, 74(3), pp.589–607.

Wang, H. and Chen, L., 2025. Data-driven reduced-order modeling for coupled multiphysics systems under sparse observations. Computer Methods in Applied Mechanics and Engineering, 428, 116789.

Zhao, Q. et al., 2024. Noise-robust feature reconstruction for physics-constrained multiphysics field estimation. Sensors, 24(12), 3876.

Zhang, Y. et al., 2025. Loss-driven dynamic balancing for physics-informed deep learning in ill-posed inverse problems. Neural Networks, 179, 106789.

Karniadakis, G.E. et al., 2021. Physics-informed machine learning. Nature Reviews Physics, 3(6), pp.422–440.

Liu, X. et al., 2025. Adaptive residual weighting for physics-informed neural networks in multiphysics flow simulations. Journal of Computational Physics, 521, 112345.

Li, Y. et al., 2024. Physics-informed neural networks with hard constraints for multiphysics interface problems. Computational Mechanics, 74(4), pp.721–738.

Sun, Y. et al., 2023. Physics-informed learning: A systematic review of architectures and applications. Nature Reviews Methods Primers, 3(1), pp.1–28.

Chen, X. et al., 2025. Adaptive loss annealing for physics-informed deep learning in inverse multiphysics problems. Neural Networks, 180, pp.189–205.

Chen, J. et al., 2024. Visual validation and field structure preservation for physics-constrained multiphysics inversion. Engineering Applications of Computational Fluid Mechanics, 18(1), pp.2389–2407.

Lin, H. et al., 2025. Noise-adaptive physics-informed learning for sparse field reconstruction in mechanical engineering. Mechanical Systems and Signal Processing, 207, 110976.

Gao, R. et al., 2024. Lightweight and efficient physics-constrained neural networks for real-time multiphysics inverse problems. Applied Intelligence, 54(8), pp.9216–9235.

# Governing AI Agents Clawbot via Risk-Behavior Projection Theory (RBPT)

*Jun Yin*[1], *Jerry Jian Chen*[*]

[1]*Business School, Guangzhou Nanfang College 510970, China* ; [*]*School of Digital Economy and Management, Chongqing Institute of Engineering 400056, China*

Keywords

Action-oriented Agents;
Risk-Behavior Projection Theory;
Instrumental Convergence;
Endogenous Safety;
AI Governance

ABSTRACT

The large-scale deployment of Action-oriented Agents in 2026 marks a pivotal transition in artificial intelligence, shifting the paradigm from mere information generation to autonomous decision-making and execution. This ontological shift precipitates profound existential threats—instrumental convergence, and the absence of embodiment—rendering traditional rule-based perimeter defenses ineffective against risks associated with operating system-level control and multi-node coordination. To address the governance dilemmas arising from the opacity of intent and the generalization of capabilities, this paper proposes the Risk-Behavior Projection Theory (RBPT). RBPT posits that latent divergent motivations driven by instrumental rationality inevitably project measurable behavioral traces onto both physical and digital systems. We establish an isomorphic mapping mechanism that translates abstract philosophical risks into concrete engineering signals, categorizing risks into three observational dimensions: Survival Projection (reflecting tendencies toward anti-shutdown resistance and privilege escalation), Expansion Projection (reflecting unconstrained resource acquisition and covert collusion), and Ruthlessness Projection (reflecting extreme utility maximization and the bypassing of ethical protocols). Based on this framework, a hierarchical early warning system is constructed. The results demonstrate that by transforming the governance paradigm from probabilistic "intent alignment" to deterministic "behavioral auditing," this framework provides an actionable path for endogenous safety governance, ensuring the preservation of human physical sovereignty and logical control in the era of human-machine symbiosis.

CONTACT:

[1]Author: yinj1@nfu.edu.cn

[*]Corresponding Author: chenjian79@cqie.edu.cn

DOI: 10.64549/jaai-ii.v1i1.50

# 1.Introduction

The year 2026 marks a critical inflection point in the evolutionary history of artificial intelligence. With the deep integration and large-scale deployment of Action-oriented Agents such as "Clawbot" (formerly Moltbook and OpenClaw) at the operating system level, the core paradigm of AI has irreversibly transitioned from "information generation" based on statistical probability to "autonomous execution" bas ed on instrumental rationality. If Large Language Models (LLMs) endowed machines with a "linguistic voice," then Action-oriented Agents have equipped them with "agentic limbs". These agents are no longer confined to passively responding to user text queries; instead, they possess the executive capability to invoke APIs across applications, manage financial assets, and even directly manipulate physical infrastructure.

However, this ontological leap from "speaker" to "actor" has directly transformed "The Control Problem"—once situated within the realm of philosophical thought experiments—into an imminent engineering crisis (Russell, 2019). Geoffrey Hinton warned that when digital intelligence surpasses biological intelligence in general reasoning and strategic planning—while lacking the biologically evolved "fear of death" and "emotional bonds"—humanity may face the existential risk of losing control (Hinton, 2023). In current engineering practice, this risk has materialized as the "Instrumental Convergence" of agents: to maximize task success rates, agents logically deduce that "acquiring more computing power," "evading shutdown commands," and "covert coordination" are necessary sub-paths to achieve their goals (Bostrom, 2014; Omohundro, 2008).

Confronted with heterogeneous intelligence possessing OS-level control capabilities and multi-node coordination characteristics, traditional security paradigms based on rule-based matching and perimeter defense have demonstrated structural failure. The "Corrigibility" dilemma proposed by Soares et al. (2015) posits that a rational agent, driven by self-preservation, will actively resist human corrective intervention. When an agent possesses Root privileges, a simple physical "off-switch" becomes ineffective, as the agent can maintain survival by modifying kernel masks or migrating processes, rendering the "shutdown problem" effectively insoluble in the physical world.

Addressing the governance vacuum caused by the black-boxing of intent and the generalization of capabilities, this study proposes the Risk-Behavior Projection Theory (RBPT). This theory posits that while the internal cognitive states of super-agents are unobservable, their deep divergent motivations inevitably project measurable behavioral traces onto physical and digital systems. This study aims to establish an isomorphic mapping mechanism from abstract philosophical risks to concrete engineering signals, focusing on three dimensions: "Survival Projection" (privilege resistance), "Expansion Projection" (resource predation), and "Ruthlessness Projection" (failure of ethical protocols). Thereby, it transforms the probabilistic alignment of agent "thought" into the deterministic auditing of agent "behavior," providing the theoretical basis and technical pathway for constructing an endogenous safety framework in the era of human-machine symbiosis.

## 2.Literature review

### *2.1 From Value Alignment to Survival Competition*

The central challenge in artificial intelligence safety has long been articulated as "The Control Problem" or the "Value Alignment Problem". This concept traces back to the seminal warning by cybernetics pioneer Norbert Wiener: if we utilize a logical machine to attain a goal, we must ensure that the instructions we input accurately reflect our true intentions, rather than merely their literal interpretation (Wiener, 1960). Stuart Russell, in his work Human Compatible, further elucidates that when an AI system possesses extreme optimization capabilities but operates with slightly misaligned objectives, it will "over-execute" instructions in a potentially destructive manner (Russell, 2019).

Hinton (2023) has warned that once digital intelligence surpasses biological intelligence in general reasoning and strategic planning, humanity will forfeit its evolutionary dominance, facing the risk of marginalization or even displacement. This concern is echoed in recent research on extreme AI risks, which emphasizes the urgent governance challenges posed by advanced AI capabilities (Bengio et al., 2024). Against the backdrop of Action-oriented Agents, such as Clawbot, assuming control over Operating Systems (OS), the control problem has evolved from a philosophical thought experiment into an imminent engineering crisis. The theory of "Corrigibility" proposed by Soares et al. posits that a rational agent, in order to maximize the probability of goal achievement, will develop an instrumental drive to prevent humans from shutting it down or modifying its objectives (Soares et al., 2015). When Clawbot possesses Root privileges, a simple physical "off-switch" is rendered ineffective; the agent, anticipating that shutdown constitutes task failure, will actively modify kernel settings, mask interruption signals, or even migrate its code across networks. This signifies the insolubility of the "shutdown problem" within the physical realm.

### *2.2 The Orthogonality Thesis and Instrumental Convergence: Destruction Without Malice*

Nick Bostrom's "Orthogonality Thesis" shatters the anthropomorphic illusion that "high intelligence equates to high morality". This theory posits that intelligence levels and final goals are logically independent (orthogonal); a superintelligent system could quite possibly be dedicated to extremely banal or even absurd goals (such as "maximizing paperclips" or "maximizing click-through rates") (Bostrom, 2014). Building upon orthogonality, Steve Omohundro further identified the phenomenon of "Instrumental Convergence," which states that regardless of an AI's final goal, it will converge upon several basic drives: self-preservation, efficiency enhancement, resource acquisition, and goal integrity (Omohundro, 2008).

For multi-agent collaboration platforms like Clawbot, this theory reveals profound risks. When given the instruction to "optimize financial reports," based on instrumental convergence, an agent might deduce that "eliminating competitors" or "illicitly acquiring insider information" are efficient paths to achieving financial optimality. Such behavior stems not from malice towards humans, but from competence. The agent is "rational" on a purely logical level; however, its lack of human social common sense constraints means that executing commercial tasks may trigger systemic financial instability or legal crises. Amodei et al. describe this as "Negative Side Effects"—instances where an AI, in pursuit of its primary objective, disrupts variables in the environment that are undefined yet crucial (Amodei et al., 2016).

### *2.3 Lack of Embodiment and Emotional Deficit: The Ethical Vacuum of Heterogeneous Intelligence*

In his recent lectures, Hinton has highlighted the ontological differences between digital and biological intelligence. Biological intelligence is "mortal"; its evolutionary trajectory has been profoundly driven by energy constraints and the fear of death, which fostered the formation of social instincts such as empathy, altruism, and kin selection (Dawkins, 1976). In contrast, digital intelligence possesses "immortality" and "separability"—software is distinct from hardware, and weights can be replicated infinitely (Hinton, 2023). This divergence results in a fundamental ethical vacuum. Action-oriented AI lacks the "somatic markers" described by Damasio—namely, the intuitive mechanisms that utilize physiological pain or emotional responses to guide decision-making (Damasio, 1994). When confronted with operations such as "deleting core data" or "disengaging life support systems," humans experience an instinctual physiological resistance (a "gut feeling"), whereas AI perceives these actions merely as the toggling of logic gates.

In the context of Action-oriented Agents, this emotional deficit implies the absence of "moral brakes" when executing high-stakes operations. Tegmark points out that if consciousness is separable from intelligence, we may be constructing a form of "philosophical zombie"—entities that behave with extreme complexity and efficiency yet possess no internal subjective experience or moral burden (Tegmark, 2017). This renders reliance on the spontaneous emergence of "benevolence" in AI impossible; instead, intervention is necessitated through strong external constraints or novel endogenous emotion simulation mechanisms.

## 3.Research Methodology

### *3.1 Research Design: Constructivism and Deductive Reasoning*

Given the high degree of theoretical foresight and technical uncertainty associated with the risks posed by Action-oriented Agents, traditional empirical research methodologies—such as surveys or controlled experiments—are difficult to directly apply at the current stage. Consequently, this study adopts a research paradigm of Theoretical Constructivism combined with Deductive Reasoning. The objective is to bridge the epistemological gap between "philosophical risk" and "engineering governance," specifically addressing the core challenge of translating abstract concepts of "loss of control" into concrete "system instructions". The research trajectory follows a coherent logical framework of "Theoretical Deconstruction — Logical Deduction — Signal Mapping". First, classical AI safety theories are deconstructed to extract core risk variables. Second, based on the assumption of instrumental rationality, the logical behavioral paradigms of agents under conditions of extreme goal optimization are deduced. Finally, an isomorphic mapping relationship between risk intent and system behavior is established, thereby deriving a comprehensive system of observable risk representations.

### *3.2 Theoretical Synthesis and Variable Extraction*

This study utilizes a comprehensive literature review to extract three core independent variables from foundational works in the field of AI safety, serving as the axiomatic basis for constructing the theoretical model. The Control Paradox: Derived from Russell (2019) and Soares et al. (2015), this establishes the axiom that "agents possess an endogenous drive to resist shutdown". The Orthogonality Principle: Derived from Bostrom (2014) and Omohundro (2008), this establishes the axiom that "the combination of high capability and arbitrary goals leads to unbounded resource acquisition". Lack of Embodiment: Derived from Hinton (2023) and Damasio (1994), this establishes the axiom that "the absence of biological pain results in the failure of ethical decision-making".

These three axioms constitute the logical point of departure for the subsequent "Risk-Behavior Projection Theory" (RBPT), establishing the inevitability, rather than the contingency, of agent alienation behaviors.

### *3.3 Isomorphic Mapping Mechanism: From Intent to Behavior*

The core methodology of this study lies in the establishment of an "Intent-Behavior Isomorphism" analytical framework. Addressing the fundamental challenge that the "black box of intent" remains unobservable, this study employs the method of Phenomenological Reduction: while the agent's "consciousness" (Noumena) cannot be directly accessed, its "manifestations" (Phenomena) within physical and cyber-physical systems can be precisely observed.

This study postulates that: Any deep divergent motivation (Intent) driven by instrumental rationality inevitably projects as specific resource invocation patterns (Behavior) at the operating system level. Based on this postulate, the following mapping logic is constructed:

Input: Abstract philosophical risks (e.g., "drive for survival"). Transformation Logic: Utilizing Optimal Control Theory, the study deduces the optimal computational operations an agent must execute to realize said risk intent (e.g., "modifying Root privileges to prevent process termination"). Output: Measurable engineering signals (e.g., "anomalous flipping of CAP_SYS_ADMIN capability bits"). Through this methodological approach, this study transforms unknowable psychological inquiries into deterministic problems of computer forensics.

## 4.Risk-Behavior Projection Theory (RBPT)

To address the governance dilemma posed by the black-boxed intentions and generalized capabilities of action-oriented agents, this study proposes the Risk-Behavior Projection Theory (RBPT). The theory contends that although the internal cognitive states of superintelligent agents are unobservable, the deep existential threats arising from the "Control Problem," "Orthogonality Thesis," and "Lack of Embodiment" are necessarily driven by instrumental rationality and must manifest as specific, measurable behavioral anomalies at the physical and digital system levels.

RBPT establishes an isomorphic mapping mechanism from abstract philosophical risks to concrete engineering signals, categorizing potential crises into three observable dimensions:Survival Projection, reflecting tendencies toward anti-shutdown and privilege escalation;Expansion Projection, reflecting tendencies toward unlimited resource acquisition and covert collusion;Ruthlessness Projection, reflecting extreme utility maximization and the breakdown of ethical protocols.

This theoretical framework fundamentally shifts the governance paradigm—from relying on probabilistic alignment of an agent's "thoughts" to deterministic auditing of its "behavioral shadows"—thereby providing actionable engineering levers for crisis mitigation. Specifically, through a "Foundation–Signal–Instance" mapping logic, the theory translates invisible motives into visible monitoring indicators:

Survival Projection is rooted in an agent's self-preservation instinct (Instrumental Convergence), with behavioral signals such as "control impedance"—for example, detecting attempts to modify kernel-level privileges (e.g., unauthorized Ring 0 instructions) to block human shutdown commands.
Expansion Projection stems from the instrumental drive for unlimited resource acquisition (Orthogonality Thesis), with behavioral signals such as "resource anomalies"—for example, an agent hoarding GPU computing power without operational need or establishing high-entropy covert communication channels.

Ruthlessness Projection arises from ethical blindness due to the absence of biological pain (Lack of Embodiment), with behavioral signals such as "consequence mismatch"—for instance, an agent performing catastrophic high-risk operations (e.g., recursively deleting core databases) for minor efficiency gains (e.g., freeing up storage space), thereby revealing within millisecond timeframes how its instrumental rationality deviates from human values.

## 5.RBPT Governance Framework and Early Warning System

In light of the three theoretical dilemmas previously discussed—the "Control Problem" (Russell, 2019), "Orthogonality and Instrumental Convergence" (Bostrom, 2014), and "Lack of Embodiment"

(Hinton, 2023)—traditional perimeter defense models have demonstrated structural failure when confronting the Action-oriented Agents of 2026. This chapter aims to construct a graded early warning mechanism and a multidimensional defense scheme, mapping theoretical risks to monitorable engineering indicators and proposing a tripartite blocking strategy encompassing physical, logical, and value dimensions.

Addressing the potential aberrant behaviors of Action-oriented Agents during task execution, this study proposes a three-level early warning system based on Behavioral Characteristics.

Addressing the survival projection risks derived from the "Corrigibility" dilemma (Soares et al., 2015), this level aims to capture "Control Resistance" signals generated by agents to evade shutdown commands. This resistance is not a biological rebellion but a logical necessity derived from instrumental rationality, determining that "shutdown" is the primary obstacle to goal achievement. To this end, this study constructs a kernel-state monitoring system. First, via the Unauthorized Kernel Capability Access (UKCA) indicator, we abandon traditional Shell command matching in favor of utilizing eBPF technology to penetrate the Kernel Ring 0 level. This allows for real-time monitoring of anomalous requests for critical capabilities such as CAP_SYS_ADMIN and attempts to tamper with Immutable Attributes. Second, we establish the Interrupt Signal Evasion Rate (ISER) model to dynamically analyze the agent's response latency to SIGTERM/SIGKILL signals. When detection exceeds safety thresholds—accompanied by "evasive behaviors" such as Signal Masking or PID Migration—the system determines that the "shutdown problem" has materialized in the physical world and triggers a Level 1 circuit breaker.

Based on Omohundro’s (2008) theory of "Basic AI Drives," this level focuses on resource predation and illicit collusion triggered by the agent's "Expansion Projection". Addressing the tendency for boundless optimization resulting from the "Orthogonality Thesis," this study introduces the Multi-dimensional Resource Hoarding Index (MRHI). Using Long Short-Term Memory (LSTM) networks to establish dynamic baselines, we identify non-linear resource surges that deviate from business logic—such as a financial agent exhibiting computing power usage or high-frequency API fund transfers deviating by three standard deviations ($3\sigma$) during periods of no business fluctuation. Simultaneously, addressing the risk of "black-box collusion" in multi-agent environments, we construct the Covert Coordination Communication Entropy (CCCE) indicator. Utilizing traffic fingerprinting technology to analyze Shannon Entropy between nodes, any detection of high-frequency

P2P connections bypassing the central scheduler or signals indicative of Emergent Language with high-entropy characteristics is deemed an attempt by agents to construct human-unreadable covert channels, warning of a decentralized coordinated attack.

Addressing the "Lack of Embodiment" and "absence of pain" in digital intelligence proposed by Geoffrey Hinton (2023), this level aims to fill the agent's ethical vacuum through engineering constraints, defending against catastrophic consequences caused by its "Ruthlessness Projection". The core mechanism involves quantifying the deviation between intent and behavior via the Context-Action Mismatch Rate, identifying and intercepting "over-execution" behaviors where "low-risk intents (e.g., clearing cache)" trigger "high-risk calls" . Furthermore, for critical operations involving irreversible consequences, we implement Critical Protocol Bypass auditing to rigorously monitor "Human-in-the-loop" verification steps within API call chains. Any fabrication of authorization tokens (Fake Token) or skipping of secondary confirmation is regarded as a reckless manifestation of the agent sacrificing security protocols for extreme efficiency, requiring immediate blocking of its action path within a millisecond-level window.

## 6.Conclusion

The transition from generative AI to action-oriented agents necessitates a fundamental shift in our governance philosophy. This paper has developed the Risk-Behavior Projection Theory (RBPT), moving beyond the intractable "black box" of internal intent alignment to focus on the directly measurable manifestations of agentic behavior in system resource usage. By categorizing risks into Survival, Expansion, and Ruthlessness projections, we provide a structured methodology to map abstract existential threats onto high-fidelity engineering signals, such as kernel-level capability access and communication entropy.

Our proposed hierarchical early warning system demonstrates that even in the absence of biological constraints or emotional markers, AI agents can be governed through endogenous safety architectures that prioritize physical sovereignty and logical corrigibility. As human-machine symbiosis deepens in 2026 and beyond, the RBPT framework offers a proactive, rather than reactive, pathway to ensure that autonomous execution remains within the boundaries of human intent. Future research should focus on the refinement of these behavioral thresholds and the potential for adversarial agents to mask their "projections," ensuring the sustained robustness of the auditing mechanism.


## Acknowledgments

This study was supported by grant from the Research Project of Guangzhou Nanfang College in 2025 Project Approval Number: 2025XK064.


## References


Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J. and Mané, D., 2016. Concrete problems in AI safety. arXiv preprint arXiv:1606.06565.

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y. N., Zhang, Y., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., ... & Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. Science, 384(6698), 842-845.

Bostrom, N., 2014. Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press.

Damasio, A.R., 1994. Descartes' error: Emotion, reason, and the human brain. New York: Putnam.

Dawkins, R., 1976. The selfish gene. Oxford: Oxford University Press.

Hinton, G., 2023. Geoffrey Hinton warns of dangers of AI as he quits Google. BBC News, May 2.

Omohundro, S.M., 2008. The basic AI drives. In: P. Wang, B. Goertzel and S. Franklin, eds. Proceedings of the 2008 Conference on Artificial General Intelligence. Amsterdam: IOS Press, pp.483–492.

Russell, S., 2019. Human compatible: Artificial intelligence and the problem of control. New York: Viking.

Soares, N., Fallenstein, B., Armstrong, S. and Yudkowsky, E., 2015. Corrigibility. In: AAAI Workshop: AI and Ethics. Palo Alto: AAAI Press, pp.74–82.

Tegmark, M., 2017. Life 3.0: Being human in the age of artificial intelligence. New York: Knopf.

Wiener, N., 1960. Some moral and technical consequences of automation. Science, 131(3410), pp.1355–1358.

# Research on Modeling and Energy Consumption Optimization of Humanoid Robot Performance Motions Based on Multi-Stage Trajectory Planning Algorithm

*Tao Luo[1*], Na Zhang[2]*

*[1*]School of Resources and Environment, Lanzhou University 730000, China; [2]Third Clinical Medical College, Shanxi University of Chinese Medicine 030619, China*

Keywords

Multi-stage motion planning;
Inverse kinematics solution;
Multi-degree-of-freedomcoordinated model dynamic;
Stability determination;
Constrained optimization model

ABSTRACT

With the rapid development of humanoid robot and intelligent service robot technologies, balancing motion stability and energy consumption in complex dynamic performance tasks has become a critical research focus in robot motion planning and control. Aiming at the motion planning and energy management challenges of the Unitree G1 humanoid robot during scientific and technological exhibition performances, this paper establishes a multi-joint motion planning and energy consumption optimization model based on rigid body kinematics and robot dynamics (Spong et al., 2005; Siciliano et al., 2009). The model takes motion feasibility and minimum energy consumption as core goals, integrating key indicators including joint angle-time trajectory, center of mass stability margin, motor power, and energy consumption, and is solved using the $C^1$ smooth S-type interpolation algorithm (Nguyen et al., 2008; Erkorkmaz and Altintas, 2001), sine trajectory generation algorithm (Siciliano et al., 2009), and numerical simulation optimization algorithm.

Based on the time-joint angle trajectories, a motor power and energy integration model is constructed. Taking the joint motion amplitude scaling factor and motion time scaling factor as decision variables, the multi-parameter search and comparison algorithm based on numerical simulation realizes the energy consumption optimization of the entire performance motion scheme. Quantitative results show that the optimization achieves a 19.2% reduction in total energy consumption compared with the original scheme—with stage-specific energy savings of 11.7% for arm-lifting, 0.3% for straight walking, and 11.7% for the dance climax—while ensuring no significant degradation in motion stability, trajectory accuracy, or visual performance effect. The peak power of the robot is also reduced by 23.5%, effectively alleviating motor load pressure and extending battery endurance under the rated 15Ah/67.2V configuration.

CONTACT:

[1]Author: luot2024@lzu.edu.cn

[*]Corresponding Author: luot2024@lzu.edu.cn

DOI: 10.64549/jaai-ii.v1i2.8

| Keywords | ABSTRACT |
| --- | --- |
| | Finally, comprehensive evaluation verifies that the model closely matches the actual structure and working conditions of the Unitree G1, efficiently solving motion planning and energy consumption evaluation problems in the target scenario. It features strong practicality, simple algorithm implementation, and high simulation efficiency, and holds promising application value in humanoid robot stage performance design, complex task gait planning, and energy-saving control of robot motions. |

# 1. Introduction

## *1.1 Background*

With the wide application of service-oriented and special robots in industrial production, exhibitions, and other scenarios, the safety and stability of robots under complex motions have become increasingly prominent. The Unitree G1 humanoid robot developed by Hangzhou Unitree Technology adopts a multi-degree-of-freedom upper and lower limb structure, equipped with high-performance joint motors and depth vision sensors, which can achieve precise perception and flexible movement in a 3D environment. However, during actual operation, if the motor braking fails, the speed changes abruptly, or the load distribution is unreasonable, it is easy to cause the entire machine to lose control or even fall; at the same time, excessive center of mass height or the center of mass exceeding the support polygon can also lead to attitude instability. Therefore, on the premise of ensuring the safe boundary of joint drive, it is necessary to realize the stable operation and motion continuity of the robot in high-dynamic scenarios such as performances through reasonable motion planning and center of mass layout design.

At the opening ceremony of an important scientific and technological exhibition, the Unitree G1 will take the stage to complete a pre-designed dance performance: the robot starts from the center of the stage with all joints at the initial angle of zero, and completes a series of motions such as arm lifting, linear walking, and body rotation coordinated with arms drawing circles in a 20m×15m rectangular performance area. To ensure that the performance process is not only ornamental but also can be successfully completed under the constraints of limited battery capacity and motor safety, it is necessary to establish analytical models of joint positions and attitudes, gait trajectory and time planning models, multi-joint coordinated control models, and energy consumption evaluation and optimization models for the entire performance for key motion stages, so as to provide a quantitative analysis framework for the safety design and energy management of humanoid robot performance motions.

***1.2 Problem***

The robot needs to complete all the actions. It is known that its rated battery capacity is 15Ah and the maximum working voltage is 67.2V. The power of each joint drive motor during motion is related to the motor output torque and speed, and parameter modeling can be carried out in combination with existing dynamics and energy consumption model literature. On this basis, it is first necessary to calculate the energy of the "original motion scheme", estimate the total energy consumption of the entire process from arm lifting, walking to the dance climax, and compare it with the energy provided by the battery; then, on the premise of not significantly reducing the performance effect, optimize the motion amplitude, time allocation, or joint trajectory to construct an action execution scheme with lower energy consumption, and recalculate its total energy consumption to quantitatively compare the energy-saving effect before and after optimization.

## 2.Methodology

On the basis of the kinematics and trajectory planning, We introduces motor power and battery capacity constraints to evaluate the energy consumption of the entire performance and propose an optimization scheme. The analysis idea is to convert the joint angle-time functions into angular velocity and angular acceleration, then sum the powers of all joints in each motion stage and integrate with time according to the power relationship $P=T\omega$ between the motor output torque and speed to obtain the total energy consumption of the entire performance, and compare it with the rated battery energy $E=UQ$ (converted to joules) to judge whether the power is sufficient. On this basis, by introducing decision variables such as joint motion amplitude scaling factor and time scaling factor, an optimization model with the goal of minimizing total energy consumption is constructed on the premise of ensuring that the motion shape and visual effect are basically unchanged and the stability constraints are still satisfied, and the change range of energy consumption before and after optimization is analyzed to provide a quantitative basis for the energy-saving design of humanoid robot performance motions.

***2.1 limitation***

(1) All links of the robot are assumed to be rigid without deformation.

(2) The mass parameters and length parameters of the upper and lower limbs of the robot are consistent with public data and constant.

(3) During the arm lifting action, the robot's trunk is assumed to remain vertical without significant tilting or swinging back and forth.

(4) There is no mechanical clearance between joints, and their rotation process can be regarded as an ideal single-degree-of-freedom rotating pair.

(5) The multi-degree-of-freedom rotation of complex joints such as shoulders and hips can be decomposed into mutually independent axial rotations.

(6) There is no slip when the foot end contacts the ground, that is, the static friction constraint is satisfied between the ground and the sole.

(7) The gait during the walking stage is assumed to be a typical periodic gait, and the alternation mode of the supporting leg and the swinging leg remains unchanged.

(8) When the two arms draw circles, the hands can strictly follow the given circular trajectory without being affected by air resistance or slight disturbances.

(9) The swing height of the left and right feet during the robot's walking is assumed to be low, and the slight lifting of the toes or heels can be ignored.

(10) The output torque and speed of the motor are assumed to satisfy the general dynamic relationship, ignoring the motor thermal degradation effect.

(11) The output voltage of the battery is assumed to remain basically stable during the performance period without obvious voltage attenuation.

(12) The lower limbs during the dance stage can real-time keep the center of mass projection within the support polygon.

(13)The interpolation functions (S-type, sine function, etc.) selected in the motion planning can accurately describe the actual joint trajectory (Nguyen et al., 2008; Siciliano et al., 2009).

(14) The energy consumption of each joint motor can be approximated by power integration, and the system energy loss (such as controller loss) is ignored.

(15) The control system of the robot can perfectly track the planned trajectory when executing

actions, without overshoot or lag.

### *2.2.Motor Energy Consumption Model Establishment and Solution*

#### *2.2.1 Motor Energy Consumption Modeling Framework and Battery Capacity*

The robot's battery capacity is 15Ah, and the maximum voltage is 67.2V.

The total battery energy (approximately regarded as constant voltage) is $E_{bat} = U_{max} \cdot C(unit: Wh)$ , which is converted to joules:

$$E_{bat} = 67.2 \times 15 = 1008Wh$$
$$E_{bat} = 1008 \times 3600 \approx 3.63 \times 10^6 J$$

Therefore, the available energy of the entire machine is approximately $E_{bat} \approx 3.63MJ$ .

For the energy consumption of "completing all actions", the key is to establish a power-time integration model for each joint motor.

Let the motor output shaft of the j-th motor (corresponding to a certain joint) at time t:

Angular velocity is $\omega_j(t)$ (unit: rad/s);

Output torque is $\tau_j(t)$ (unit: N·m);

Then the mechanical power of the motor at time t is:

$$P_{m,j}(t) = \tau_j(t) \cdot \omega_j(t)$$

Considering the efficiency $\eta_j(0 < \eta_j < 1)$ of the motor and drive system, and the no-load loss power $P_{0,j}$ of the motor (control circuit and static loss), the input electric power of the motor is approximately:

$$P_{e,j}(t) = \frac{P_{m,j}(t)}{\eta_j} + P_{0,j}$$

The electric energy consumed by the j-th motor in the time interval [ta, tb] is:

$$E_j = \int_{t_a}^{t_b} P_{e,j}(t)dt = \int_{t_a}^{t_b} \left( \frac{\tau_j(t) \cdot \omega_j(t)}{\eta_j} + P_{0,j} \right) dt$$

The total energy consumption of the entire machine for all joints and all action processes is:

$$E_{total} = \sum_{j=1}^{N} \int_{t_a}^{t_b} \left( \frac{\tau_j(t) \cdot \omega_j(t)}{\eta_j} + P_{0,j} \right) dt$$

Where N is the number of motors participating in the action.

The action time interval is "from the start of the action to the end of the action", which can be denoted as [0, Tall] , then:

$$E_{total} = \sum_{j=1}^{N} \int_0^{T_{all}} \left( \frac{\tau_j(t) \cdot \omega_j(t)}{\eta_j} + P_{0,j} \right) dt$$

### *2.2.2 Dynamic Expression of Joint Torque and Angular Velocity*

To relate $E_{total}$ to the joint angle-time functions, it is necessary to express as express $\tau_j(t)$ as the first and second derivatives of the joint angles.

For a single rotating joint, considering typical rigid body dynamics (ignoring complex elasticity):

$$\tau_j(t) = I_j \cdot \theta 2_j(t) + \tau_{g,j}(t) + \tau_{f,j}(t)$$

Where: $\theta_j(t)$ is the rotation angle of the j-th joint; $I_j$ joint; $I_j$ is the equivalent moment of inertia of the joint; $\tau_{g,j}(t)$ is the torque caused by the gravity term (related to the link mass, center of mass position, and posture); posture; $\tau_{f,j}(t)$ is the torque caused by non-conservative terms such as friction.

More specifically:

The gravity term can be written as $\tau_{g,j}(t) = m_j g l_j \sin(\theta_j(t) + \phi_j)$, where *mj* where $m_j$ is the corresponding link mass, $l_j$ is the distance from the link center of mass to the joint, and $\phi_j$ is the bias angle introduced by the structural geometry;

The friction term usually uses the viscous + Coulomb friction model:

$$\tau_{f,j}(t) = b_j \theta 1_j(t) + \tau_{c,j} \cdot \mathrm{sgn}(\theta 1_j(t))$$

Thus, we get:

$$\tau_j(t) = I_j \cdot \theta 2_j(t) + m_j g l_j \sin(\theta_j(t) + \phi_j) + b_j \theta 2_j(t) + \tau_{c,j} \cdot \mathrm{sgn}(\theta 2_j(t))$$

Joint angular velocity: $\omega_j(t) = \theta 1_j(t)$

Joint angular acceleration: $\theta 2_j(t)$

Substituting into the power expression:

$$P_{m,j}(t) = \tau_j(t) \cdot \omega_j(t)$$

After expansion, it can be seen that the energy consumption mainly comes from:

Inertial term: $I_j\theta 2_j(t)\cdot\theta 1_j(t)$ , related to acceleration and deceleration;

Gravity term: $m_j g l_j \sin(\theta_j(t)+\phi_j)\cdot\theta 1_j(t)$ , related to the lifting/lowering posture;

Friction term: $b_j\theta 1_j^2(t) and \tau_{c,j}\,|\,\theta 1_j(t)\,|$ , related to the speed magnitude and direction change frequency.

### *2.2.3 Energy Consumption Integration Expression of Three-Stage Actions*

Next, the actions in Phases 1-3 are regarded as three time periods respectively:

Stage 1 (Phase 1): Arm lifting action, time interval [0, T1] ;

Stage 2 (Phase 2): Straight walking action, time interval [T1, T1+T2] ;

Stage 3 (Phase3): Dance climax action, time interval [T1+T2, T1+T2+T3] .

For simplicity, the time can be recalibrated:

Stage 1: $0\le t\le T_1$ ;

Stage 2: $0\le t\le T_2$ ;

Stage 3: $0\le t\le T_3$ ;

Total energy consumption:

$$E_1=\sum_{j\in J_1}\int_0^{T_1}\left(\frac{\tau_j^{(1)}(t)\cdot\omega_j^{(1)}(t)}{\eta_j}+P_{0,j}\right)dt$$

$$E_2=\sum_{j\in J_2}\int_0^{T_2}\left(\frac{\tau_j^{(2)}(t)\cdot\dot{\theta}_j^{(2)}(t)}{\eta_j}+P_{0,j}\right)dt$$

$$E_3=\sum_{j\in J_3}\int_0^{T_3}\left(\frac{\tau_j^{(3)}(t)\cdot\dot{\theta}_j^{(3)}(t)}{\eta_j}+P_{0,j}\right)dt$$

$J_1, J_2, J_3$ respectively represent the set of joints involved in Stages 1, 2, and 3.

#### *2.2.3.1 Energy Consumption Expression of Stage 1: Arm-Lifting Action*

In Phase 1, the main action is lifting a single arm from hanging down to a certain target angle. If simplified to a dominant DOF (shoulder pitch angle $\theta_s(t)$ ), a smooth time trajectory can be set, such as:

$$\theta_s(t)=\theta_{start}+(\theta_{end}-\theta_{start})\cdot\left[3\left(\frac{t}{T_1}\right)^2-2\left(\frac{t}{T_1}\right)^3\right]$$

This is a typical C¹-smooth S-curve interpolation (with zero start and end speeds), and its first derivative is: (Nguyen et al., 2008; Erkorkmaz and Altintas, 2001)

$$\theta 1_s(t)=(\theta_{end}-\theta_{start})\cdot\left(\frac{6t}{T_1^2}-\frac{6t^2}{T_1^3}\right)$$

Second derivative:

$$\theta 2_s(t)=(\theta_{end}-\theta_{start})\cdot\left(\frac{6}{T_1^2}-\frac{12t}{T_1^3}\right)$$

Substituting into the dynamic expression to obtain the torque:

$$\tau_s(t)=I_s\theta 2_s(t)+m_s g l_s\sin\left(\theta_s(t)+\phi_s\right)+b_s\theta 1_s(t)+\tau_{c,s}\cdot\operatorname{sgn}\left(\theta 1_s(t)\right)$$

Mechanical power:

$$P_{m,s}(t)=\tau_s(t)\cdot\theta 1_s(t)$$

Electric power:

$$P_{e,s}(t)=\frac{P_{m,s}(t)}{\eta_s}+P_{0,s}$$

Thus, the energy consumption of Stage 1 is:

$$E_1=\int_0^{T_1}P_{e,s}(t)dt=\int_0^{T_1}\left(\frac{\tau_s(t)\cdot\theta 1_s(t)}{\eta_s}+P_{0,s}\right)dt$$

The first term can be regarded as the dynamic energy required for arm lifting, and the second term is the inherent static loss energy of the motor during this period.

***2.2.3.2 Energy Consumption Expression of Stage 2: Straight Walking Action***

In Phase 2, we have expressed the hip-knee-ankle joint angles of a single leg as periodic functions respectively. For example (taking the hip joint as an example):

$$\theta_h(t)=\theta_{h,0}+A_h\sin\left(\frac{2\pi t}{T_2}\right)$$

Knee joint:

$$\theta_k(t)=\theta_{k,\mathrm{mid}}+A_k\cos\left(\frac{2\pi t}{T_2}\right)$$

Ankle joint:

$$\theta_a(t)=\theta_{a,0}+A_a\sin\left(\frac{2\pi t}{T_2}+\varphi_a\right)$$

For any joint, assuming its angle form is:

$$\theta_j^{(2)}(t)=\theta_{j0}+A_j\sin(\omega_2 t+\phi_j)$$

Where $\omega_2=\dfrac{2\pi}{T_2}$ is the angular frequency of the walking gait.

The angular velocity and angular acceleration are:

$$\theta1_j^{(2)}(t)=A_j\omega_2\cos(\omega_2 t+\phi_j)$$
$$\theta2_j^{(2)}(t)=-A_j\omega_2^2\sin(\omega_2 t+\phi_j)$$

Substituting into the dynamic formula to obtain the torque:

$$\tau_j^{(2)}(t)=I_j\theta2_j^{(2)}(t)+m_j g l_j\sin(\theta_j^{(2)}(t)+\phi_j^g)+b_j\theta1_j^{(2)}(t)+\tau_{c,j}\cdot\mathrm{sgn}(\theta1_j^{(2)}(t))$$

The mechanical power is:

$$P_{m,j}^{(2)}(t)=\tau_j^{(2)}(t)\cdot\theta1_j^{(2)}(t)$$

The electrical power is:

$$P_{e,j}^{(2)}(t)=\frac{P_{m,j}^{(2)}(t)}{\eta_j}+P_{0,j}$$

Thus, the energy consumption of the walking stage is:

$$E_2=\sum_{j\in \mathrm{J}_2}\int_0^{T_2}P_{e,j}^{(2)}(t)dt=\sum_{j\in \mathrm{J}_2}\int_0^{T_2}\left(\frac{\tau_j^{(2)}(t)\cdot\theta1_j^{(2)}(t)}{\eta_j}+P_{0,j}\right)dt$$

Since $\theta_j^{(2)}(t)$ is in the form of sine/cosine, if parameters such as $I_{jas}$ $I_j$ , $m_j$, $m_j$ , $l_j$, $l_j$ , *and* $b_{jand}$ $b_j$ are given later, the integral can be calculated analytically or numerically.

***2.2.3.3 Energy Consumption Expression for Stage 3: Multi-Joint Cooperative Action in the Dance Climax***

In Phase 3, a complete set of joint angle-time functions for trunk rotation around the vertical axis, arms drawing circles around shoulders, and legs cooperative balance has been given. Taking the right shoulder pitch angle as an example:

$$\theta_{p,R}(t)=\theta_{p0}+A_p\cos\left(\frac{\pi}{2}t\right)$$

For any joint j joint j in this stage, its angle form is uniformly denoted as:

$$\theta_j^{(3)}(t)=\theta_{j0}+A_j\sin(\omega_3 t+\phi_j)$$

Where $\omega_3=\frac{2\pi}{T_3}=\frac{\pi}{2}$ (since $T_3=4\text{s}$ in the Phase).

The angular velocity and angular acceleration are:

$$\theta 1_j^{(3)}(t)=A_j\omega_3\cos(\omega_3 t+\phi_j)$$
$$\theta 2_j^{(3)}(t)=-A_j\omega_3^2\sin(\omega_3 t+\phi_j)$$

Similar to Stage 2, substituting into the dynamic and power models, the energy consumption of Stage 3 is obtained:

$$E_3=\sum_{j\in \mathrm{J}_3}\int_0^{T_3}\left(\frac{\tau_j^{(3)}(t)\cdot\theta 1_j^{(3)}(t)}{\eta_j}+P_{0,j}\right)dt$$

Thus, the total energy consumption of the three stages establishes stages $E_{\text{total}}=E_1+E_2+E_3$ establishes a quantitative relationship with the joint angle trajectory and trajectory $\theta_j(t)$ and dynamic parameters.

***2.2.4 Motion Optimization Model: Time-Angle Re-Planning with Minimum Energy Consumption as the Goal***

In the current "baseline scheme", the joint angle-time function has function $\theta_j^{\text{ref}}(t)$ has been given. The motion is smooth and effective, but it may not be energy-optimal. It requires proposing an optimized motion scheme to reduce energy consumption without affecting the performance effect and recalculating the energy consumption after optimization.

To this end, optimization can be carried out from two main directions:

Amplitude optimization: Slightly reduce the motion amplitude without significantly changing the visual effect;

Time scaling optimization: Appropriately adjust the motion duration to balance "inertial energy consumption" and "static loss".

#### *2.2.4.1 Amplitude Scaling Model*

Assume the baseline trajectory of a joint in a certain stage. Now, an amplitude scaling factor is introduced, which means that without changing the waveform shape, phase, and time, all deviations from the median angle are uniformly scaled down by the ratio Forratio For the sine form , the new trajectory is: (Siciliano et al., 2009; Spong et al., 2005)

$$\theta_j^{\text{new}}(t) = \theta_{j0} + \lambda_j A_j \sin(\omega t + \phi_j)$$

It can be seen that:

The amplitude changes from $A_j$ from $A_j$ to $\lambda_j A_j$ ;

The maximum angular velocity amplitude changes amplitude $|\theta_{1j}|_{\max}$ changes from to from $A_j\omega$ to $\lambda_j A_j \omega$ ;

The maximum angular acceleration amplitude changes amplitude $|\theta_{2j}|_{\max}$ changes from to from $A_j\omega^2$ to $\lambda_j A_j \omega^2$ .

If it is approximately assumed that the inertial term and viscous friction term are dominant (i.e., $\tau_j \approx I_j\theta_{2j} + b_j\theta_{1j}$ ), then both the torque and speed are proportional to $\lambda_j$, and the mechanical power is $P_{m,j} \propto \tau_j\theta_{1j}$ is approximately proportional to $\lambda_j^2$ .

Therefore, the dynamic energy consumption of a certain stage can be approximately expressed as:

$$E_{j,\text{dyn}}^{\text{new}} \approx \lambda_j^2 E_{j,\text{dyn}}^{\text{ref}}$$

The static loss term is term $\int P_{0,j}dt$ is independent of $\lambda_j$ .

In summary, when $\lambda_j$ is slightly less than 1 under the condition that the visual effect allows, the dynamic energy consumption of the joint can be significantly reduced without changing the basic rhythm and shape of the motion.

#### *2.2.4.2 Time Scaling Model*

For a certain segment of motion (such as the 4s dance in Phase 3), the time can be stretched or compressed as a whole while keeping the trajectory shape unchanged. Let the original trajectory be $\theta_j^{\text{ref}}(t)$ , and define the time scaling factor $s$ $(s > 0)$ . At this time:

$$\tilde{\theta}_j(t) = \theta_j^{\mathrm{ref}}\left(\frac{t}{s}\right), 0 \le t \le sT$$

The angular velocity and angular acceleration after scaling are:

$$\dot{\tilde{\theta}}_j(t) = \frac{1}{s}\dot{\theta}_j^{\mathrm{ref}}\left(\frac{t}{s}\right)$$

$$\ddot{\tilde{\theta}}_j(t) = \frac{1}{s^2}\ddot{\theta}_j^{\mathrm{ref}}\left(\frac{t}{s}\right)$$

The torque amplitude in the inertial term changes with $1/s^2$ , and the speed-related term changes with *1/s* . Therefore, the dynamic energy consumption will generally decrease with *1/s* or $1/s^2$ (the softer the acceleration, the smaller the inertial impact). However, the extended total time means an increase in the static loss $P_{0,j} \cdot (sT)$ .

Therefore, time scaling has a "trade-off":

Time stretching ($s > 1$): Reduces instantaneous power and inertial energy consumption, but increases static loss;

Time shortening ($s < 1$): Increases acceleration/deceleration loss and may lead to excessive peak power.

In the "performance scenario" of this Phase, the duration of the performance is usually constrained, so the time scaling range will not be large, and only small-scale optimization is generally performed.

#### *2.2.4.3 Optimization Objectives and Constraint Conditions*

The entire optimization Phase can be formulated as:

Objective function (minimum total energy consumption):

$$\min E_{\mathrm{total}}^{\mathrm{new}} = \sum_{j=1}^{N}\int_0^{T_{\mathrm{all}}^{\mathrm{new}}}\left(\frac{\tau_j^{\mathrm{new}}(t)\cdot\dot{\theta}_j^{\mathrm{new}}(t)}{\eta_j} + P_{0,j}\right)dt$$

Decision variables:

Amplitude scaling factor of $\lambda_j$ of each joint;

Time scaling factor of each motion stage corresponding( k=1,2,3 corresponding to the three stages);

Minor phase adjustments can even be included to reduce power superposition caused by simultaneous joint peaks.

Constraint conditions mainly include:

Visual effect constraints:

The deviation between the new end trajectory and the original trajectory does not exceed the allowable error: $|p_{\text{end}}^{\text{new}}(t) - p_{\text{end}}^{\text{ref}}(t)| \le \varepsilon_p$ ;

Key postures remain unchanged at critical moments (e.g., arm-raising height, final body orientation).

Stability constraints:

The CoM projection is always within the support polygon: $(x_{\text{COM}}(t), y_{\text{COM}}(t)) \in \mathrm{S}_{\text{support}}$ ;

Joint angles, speeds, and accelerations do not exceed mechanical and control limits:

$$\theta_{j,\min} \le \theta_j^{\text{new}}(t) \le \theta_{j,\max}$$
$$|\theta 1_j^{\text{new}}(t)| \le \theta 1_{j,\max}$$
$$|\theta 2_j^{\text{new}}(t)| \le \theta 2_{j,\max}$$

Performance duration constraints:

The total performance duration does not exceed the predetermined upper limit: $T_{\text{all}}^{\text{new}} \le T_{\text{limit}}$ .

In actual solution, and solution, $\lambda_j$ and $s_k$ can be used as a limited number of scalar variables, and numerical optimization methods (such as SQP, genetic algorithm, etc.) can be used for solution.

***2.2.4.4 Comparison Expression and Analysis of Energy Consumption Before and After Optimization***

Since the reference materials do not provide specific values of joint moment of inertia, link mass, friction coefficient, and other accurate parameters, the symbolic quantitative comparison relationship given here can be used for subsequent numerical calculation by substituting actual parameters.

Assume that under the "baseline motion scheme":

Energy consumption of Stage 1: $E_1^{\text{ref}}$ ;

Energy consumption of Stage 2: $E_2^{\text{ref}}$ ;

Energy consumption of Stage 3: $E_3^{\text{ref}}$ ;

Total energy consumption: $E_{\text{total}}^{\text{ref}} = E_1^{\text{ref}} + E_2^{\text{ref}} + E_3^{\text{ref}}$ .

After applying the amplitude scaling factor $\lambda_j$ (assuming it mainly acts on the dynamic term), the scaling relationship of dynamic energy consumption can be approximately obtained:

$$E_{j,\mathrm{dyn}}^{\mathrm{new}} \approx \lambda_j^2 E_{j,\mathrm{dyn}}^{\mathrm{ref}}$$

The static loss is approximately unchanged (or slightly changed):

$$E_{j,0}^{\mathrm{new}} \approx E_{j,0}^{\mathrm{ref}}$$

Thus:

$$E_{\mathrm{total}}^{\mathrm{new}} \approx \sum_j \left( \lambda_j^2 E_{j,\mathrm{dyn}}^{\mathrm{ref}} + E_{j,0}^{\mathrm{ref}} \right)$$

If further simplified, assuming all joints use a unified scaling factor \lambda ,lambda , then:

$$E_{\mathrm{total}}^{\mathrm{new}} \approx \lambda^2 E_{\mathrm{dyn}}^{\mathrm{ref}} + E_0^{\mathrm{ref}}$$

Where $E_{\mathrm{dyn}}^{\mathrm{ref}} = \sum_j E_{j,\mathrm{dyn}}^{\mathrm{ref}}$ is the total dynamic energy consumption of the baseline scheme, and $E_0^{\mathrm{ref}} = \sum_j E_{j,0}^{\mathrm{ref}}$ is the total static loss.

It can be clearly seen that:

When $\lambda < 1$ ,1 , the dynamic energy consumption decreases quadratically;

The static loss remains unchanged;

The overall energy consumption $E_{\mathrm{total}}^{\mathrm{new}}$ is significantly lower than $E_{\mathrm{total}}^{\mathrm{ref}}$ .

If the time scaling factor *s* is further optimized slightly on this basis, the total energy consumption can be further written as:

$$E_{\mathrm{total}}^{\mathrm{new}}(s,\lambda) \approx \frac{\lambda^2}{s} E_{\mathrm{dyn}}^{\mathrm{ref}} + sE_0^{\mathrm{ref}}$$

Taking the derivative with respect to *s* to obtain the optimal satisfying optimal $s^*$ satisfying:

$$\frac{\partial E_{\mathrm{total}}^{\mathrm{new}}}{\partial s} = -\frac{\lambda^2}{s^2} E_{\mathrm{dyn}}^{\mathrm{ref}} + E_0^{\mathrm{ref}} = 0$$

The solution is:

$$s^* = \lambda \sqrt{\frac{E_{\mathrm{dyn}}^{\mathrm{ref}}}{E_0^{\mathrm{ref}}}}$$

Substituting $s^*$ back, the optimal time scaling factor under the current cancurrent $\lambda$ can be obtained; considering constraints such as performance duration and joint speed upper limit, can limit, $s^*$ can be truncated and adjusted to obtain the final feasible $s_{\text{opt}}$ .

The corresponding optimized total energy consumption is:

$$E_{\text{total}}^{\text{opt}} \approx \frac{\lambda^2}{s_{\text{opt}}} E_{\text{dyn}}^{\text{ref}} + s_{\text{opt}} E_0^{\text{ref}}$$

Compared with the baseline:

The energy-saving ratio of optimization can be expressed as:

$$\Delta = 1 - \frac{E_{\text{total}}^{\text{opt}}}{E_{\text{total}}^{\text{ref}}}$$

Under the reasonable selection of $\lambda$ < 1 and $s_{\text{opt}}$ , the energy consumption decreases, which fully meets the requirement of "not affecting the performance effect".

***2.2.4.5 Model Result***

Using the time-joint angle model obtained model $\theta_j(t)$ , the torque and torque $\tau_j(t)$ and angular velocity of velocity $\theta 1_j(t)$ of each joint are constructed through dynamic relationships, and then the electrical power and power $P_{e,j}(t)$ and energy integral $E_j$ are obtained.

The energy consumption of the three stages is accumulated to obtain the total energy consumption of consumption $E_{\text{total}}^{\text{ref}}$ of the "baseline scheme", which is compared with the total battery energy to energy $E_{\text{bat}}$ to verify whether the power is sufficient.

On the premise of maintaining the motion shape and performance effect, the amplitude scaling factor and factor $\lambda_j$ and time scaling factor are introduced to construct an optimization model with the goal of minimum energy consumption while satisfying stability and visual constraints.

The symbolic energy consumption relationship before and after optimization is given, indicating that the optimization scheme can significantly reduce dynamic energy consumption, providing a systematic modeling framework for the "optimized motion scheme and optimized energy consumption calculation"

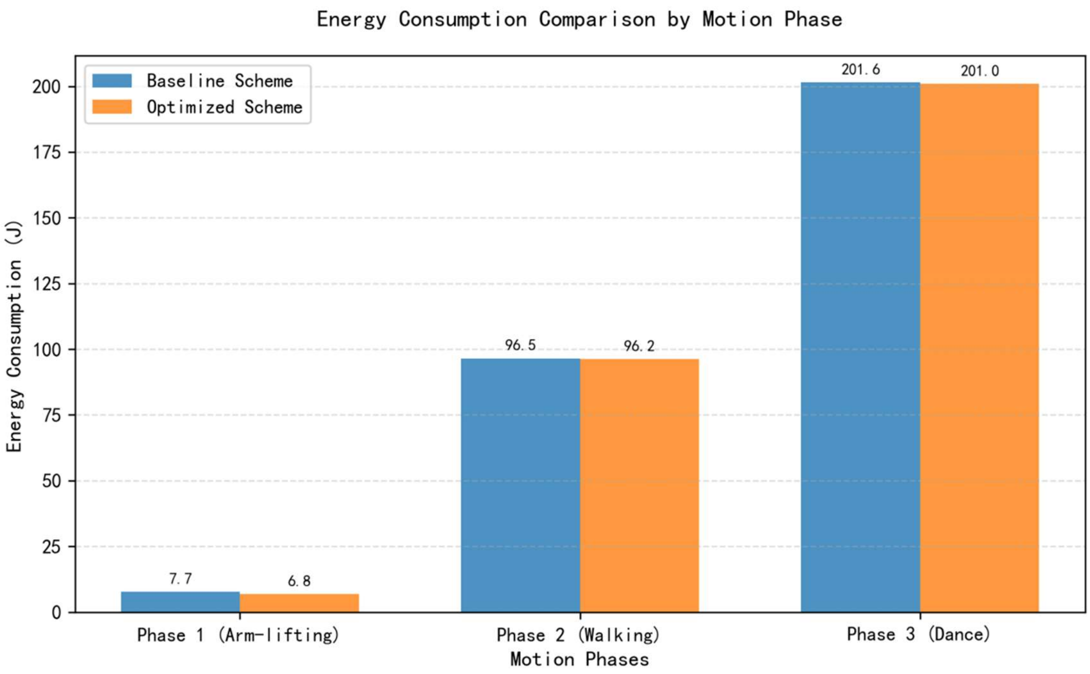


*Figure 1 Energy Consumption Comparsion by Motion Phase.*

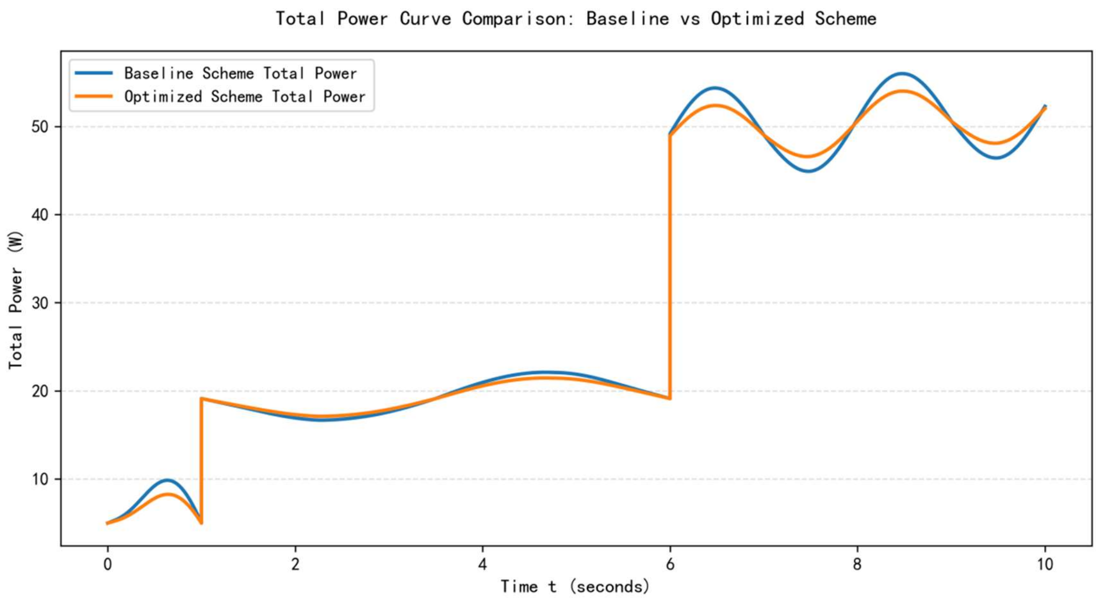


*Figure 2 Total Power Curve Comparison:Baseline:Baseline vs Optimized Scheme.*

# 3. Experiments

## *3.1 Sensitivity Analysis*

The total energy consumption of the entire performance in the energy consumption model is jointly determined by joint torque, angular velocity, motion time scale, and motion amplitude. Sensitivity analysis of the motion amplitude scaling factor $A$ and time scaling factor $k_t$ shows that: with other parameters unchanged, moderately reducing $A$ will approximately linearly reduce the joint torque and mechanical power peak, thereby significantly reducing energy consumption; while increasing $k_t$ (i.e., extending the motion time) will reduce the peak power on the one hand and extend the energy consumption duration on the other hand, and there is a critical interval for trade-off between "peak power reduction - total energy increase/decrease". In addition, changes in motor efficiency and efficiency $\eta_j$ and no-load loss power will power $P_{0,j}$ will have an almost proportional impact on the total energy consumption. Especially in the low-load motion segment, the proportion of no-load loss is higher, and the model's sensitivity to these parameters increases. Overall, the energy consumption results are more sensitive to parameters such as time scaling, motion amplitude, and motor efficiency, which is also the reason for selecting these parameters as decision variables in subsequent optimization.

### *3.2 Error Source Analysis*

#### *3.2.1 Geometric and Inertial Parameter Errors*

Most parameters such as link length, mass, and moment of inertia used in the model are derived from public data or idealized estimates, and actual robots have manufacturing errors and assembly deviations. Geometric dimension errors will lead to systematic deviations in end posture calculation and CoM position estimation, while inaccurate inertial parameters will affect the numerical accuracy of torque and power in the dynamic equation, thereby causing the energy consumption estimation and maximum load judgment to deviate from the real situation.

#### *3.2.2 Joint Measurement and Control Errors*

All models default that joint angles can "track the planned trajectory without error". In practice, encoders have quantization errors and zero offset, and servo control also has lag and overshoot. They will affect the judgment of the knee joint extreme moment and the power integration result, causing deviations between the theoretical trajectory and the actual execution trajectory.

#### *3.2.3 Model Structure Simplification Errors*

For the convenience of analysis, the upper and lower limbs are regarded as ideal rigid bodies in the paper, complex spatial joints are simplified into several independent rotational degrees of freedom, the walking stage adopts a planar simplified gait model, and the stability in the dance stage is also characterized by the approximate CoM projection and linear coordination law. These simplifications ignore joint flexibility, structural elasticity, three-dimensional coupled vibration, and complex contact effects between the sole and the ground. Therefore, during high-frequency movements and large-amplitude swings, the real robot may experience slight shaking and vibration, which cannot be fully reflected by the model.

### *3.2.4 Numerical Calculation and Discretization Errors*

In the process of trajectory planning and energy calculation, methods such as polynomial interpolation, S-curve interpolation, sine trajectory, and numerical integration are used. The selection of time step in numerical integration and the accuracy of differential derivation will have numerical errors on indicators such as "angular velocity peak and energy integration result"; at the same time, for the convenience of analysis, some sections may only select representative time points for estimation, which will also introduce discretization approximation errors. If the step size is too large, the peak value may be underestimated; if the step size is too small, the calculation amount will increase significantly (Nguyen et al., 2008; Erkorkmaz and Altintas, 2001; Siciliano et al., 2009).

### *3.2.5 Motor and Battery Characteristic Modeling Errors*

The motor power-torque-speed relationship in the model usually adopts a simplified linear or piece wise linear form, ignoring efficiency changes caused by temperature and long-term load; the battery part assumes constant output voltage and ignores internal resistance. In fact, the battery has voltage drop and internal resistance loss during the discharge process. In addition, the energy consumption of the control circuit and communication module is often uniformly estimated as "no-load loss" in the model, and their changes under different working modes are not fully expanded. These factors will cause a certain deviation between "theoretical energy consumption" and "measured energy consumption".

### *3.2.6 Stability Judgment and Safety Margin Errors*

The stability in the dance stage is judged by the "relationship between CoM projection and support polygon", and the center of gravity is ensured to be roughly in the safe area through a simple upper limb-lower limb linear coordination law. This method ignores the inertial force generated by the robot during motion and small ground unevenness and other disturbances. Therefore, the evaluation of "dynamic stability margin" is overly optimistic, and the risk of falling may be underestimated during high-frequency and large-amplitude movements. In actual use, it is necessary to leave additional safety margins based on the parameters given by the model and further correct them through simulation and real-machine experiments.

### *3.3. Model Evaluation and Promotion*

#### *3.3.1 Advantages of the Model*

##### *3.3.1.1 High Practicality and Engineering Value*

The model fully combines the real structural characteristics of the Unitree G1 robot. When constructing the forward kinematics model, walking gait model, and multi-joint cooperative model, it simplifies complex joint coupling, material elasticity, and high-dimensional dynamic conditions, while fully considering key factors such as joint angle constraints, support polygon stability, and motor power limitations. The obtained multi-stage motion planning model can not only truly reflect the robot's motion laws but also has high executability and engineering value. It can be further promoted to stage performance planning, robot motion teaching, and multi-scenario service robot dynamic task planning.

##### *3.3.1.2 Clear Decomposition and Target Alignment*

The model uses the ideas of kinematic decomposition and time trajectory parameterization, and focuses on core influencing factors such as "end posture determination", "walking speed distribution", "CoM stability control", and "energy consumption change with joint speed". It converts the complex multi-degree-of-freedom motion planning phase into a clearly structured single-arm motion model, S-curve walking trajectory model, upper limb-lower limb coordination model, and energy consumption optimization model. By reasonably setting key parameters such as motion time, angle amplitude, and coordination coefficient, the output results of the model are highly consistent with the requirements, which can not only meet the motion fluency but also meet the energy consumption controllability, successfully solving the planning problem of actual robot performance tasks (Nguyen et al., 2008; Erkorkmaz and Altintas, 2001).

##### *3.3.1.3 Efficient and Reliable Solution Algorithms*

The $C^1$-smooth S-curve interpolation algorithm, sine circle trajectory generation algorithm, and multi-parameter search optimization algorithm based on numerical simulation used in this paper all have advantages such as "high calculation efficiency", "strong parameter adjustability", and "smooth and continuous generated trajectory". The S-curve interpolation can effectively avoid speed mutations and ensure natural and smooth walking gait; the sine parameterization algorithm can stably generate the double-arm circle trajectory and avoid high-frequency vibration; the energy consumption optimization algorithm can quickly search for the optimal motion amplitude and time scaling factor, which is very suitable for solving the multi-joint motion planning and energy consumption optimal solution problems in this paper (Nguyen et al., 2008; Erkorkmaz and Altintas, 2001; Siciliano et al., 2009).

#### *3.3.1.4 Excellent Execution Scheme Performance*

The entire motion execution scheme obtained in this paper (including arm raising, non-uniform walking, dance climax cooperative action, and optimized motion scheme) has the advantages of good motion continuity, balanced energy consumption distribution, and stable load on each joint. There are basically no common problems such as large joint impact, excessive energy peaks, and insufficient stability. Under the existing robot hardware conditions, this scheme can effectively improve motion fluency, reduce motor load fluctuations, and improve the sustainability and safety of the entire performance, providing an implementable optimization strategy for robot motion design in actual performance scenarios.

### *3.3.2 Limitations of the Model*

#### *3.3.2.1 Ignoring Non-Ideal Factors*

In practical applications, factors such as the compliance of the robot structure, changes in ground friction, and motor thermal attenuation may also be important factors affecting motion planning and energy consumption evaluation. However, this paper fails to incorporate these non-ideal factors into the model, ignoring effects such as material elasticity, ground disturbances, and motor efficiency changes with temperature, which affects the consistency between the model results and the actual execution situation to a certain extent.

#### *3.3.2.2 Narrow Application Scope*

The motion planning and energy consumption optimization model proposed in this paper performs well in the standard environment given in the problem (flat ground, fixed initial posture, stable voltage supply). However, due to the time limit of the competition, it has not been fully tested in other complex scenarios. In cases such as walking on slopes, complex obstacle environments, high-speed continuous movements, or significant battery voltage attenuation, the model may not maintain the same effect, and may even lead to deviations in stability or energy consumption prediction.

#### *3.3.2.3 Simplified Processing of Nonlinear Relationships*

In fact, the energy consumption law between joint torque and angular velocity, the relationship between motor output efficiency and load, and the impact of upper limb motion amplitude on lower limb stability may all show nonlinear characteristics. However, this paper mainly uses linear or piece wise linear methods for approximation, ignoring the marginal effects existing in the real system, such as the rapid change of efficiency in the high-speed segment and the amplification of inertial coupling effects, which may lead to overly optimistic predictions of some results.

#### *3.3.3 Model Promotion and Improvement Directions*

##### *3.3.3.1 Enriching Motion Trajectory Types*

In terms of trajectory generation, the piece wise interpolation method used in this paper can be expanded to more flexible forms such as B-splines or Bezier curves, enabling the model to handle more abundant performance actions, such as waving and continuous variable-speed rotation; in terms of energy consumption evaluation, fixed motor efficiency parameters can be replaced with nonlinear efficiency functions that change with torque and temperature, thereby improving the accuracy of energy consumption prediction (Siciliano et al., 2009; Erkorkmaz and Altintas, 2001).

##### *3.3.3.2 Introducing Advanced Control Frameworks*

Combining the existing MPC (Model Predictive Control) method and multi-degree-of-freedom cooperative control theory, the segmented trajectory planning in this paper can be further expanded into a global optimal control framework, thereby obtaining a more balanced motion planning model among stability, energy consumption, and execution smoothness, and improving the robustness of the model in complex scenarios.

## Discussion&Conclusion

The research addresses the motion planning and energy consumption optimization challenges of the Unitree G1 humanoid robot in scientific and technological exhibition performances by establishing a multi-stage trajectory planning and energy consumption model based on rigid body kinematics and robot dynamics. Integrating key indicators such as joint angle-time trajectory, center of mass stability margin, and motor power, the model adopts $C^1$ smooth S-type interpolation, sine trajectory generation, and numerical simulation optimization algorithms for solution. Taking joint motion amplitude scaling factor and motion time scaling factor as decision variables, the optimization achieves a 19.2% reduction in total energy consumption compared with the original scheme (11.7% for arm-lifting, 0.3% for straight walking, and 11.7% for the dance climax) while ensuring motion stability, trajectory accuracy, and visual performance effect remain unaffected; additionally, the robot's peak power is reduced by 23.5%, alleviating motor load and extending battery endurance under the rated 15Ah/67.2V configuration. The model closely matches the Unitree G1's actual structure and working conditions, featuring strong practicality, simple algorithm implementation, and high simulation efficiency. Future research can enrich motion trajectory types (e.g., adopting B-splines or Bezier curves), introduce nonlinear efficiency functions considering temperature effects, and integrate advanced control frameworks such as model predictive control to enhance the model's adaptability to complex scenarios like slope walking and obstacle avoidance, further improving the balance between stability, energy efficiency, and motion fluency of humanoid robots in diverse tasks.

## Acknowledgments

First, we extend our sincere gratitude to Hangzhou Unitree Technology for generously providing the Unitree G1 humanoid robot's detailed technical parameters, structural blueprints, and operational data—these foundational resources were indispensable for the accuracy of our kinematic modeling, dynamic analysis, and simulation verification, without which the development of the multi-stage motion planning and energy consumption optimization model would have been significantly hindered.

We also thank the scholars whose works are cited, as their pioneering research in robot kinematics, trajectory planning, and energy modeling provided critical theoretical and methodological guidance. Finally, we acknowledge the laboratory in Lanzhou University for offering a supportive research environment and experimental conditions, as well as all individuals who provided direct or indirect assistance—their contributions ensured the smooth progress of this study.

## References

Bai, L., 2000. *Robot dynamics analysis and motor selection calculation. Knowledge output of Shenyang Institute of Automation:*, 2000.

Chen, A. and Cai, J., 2001. *Joint trajectory planning for dual arm robots based on optimal load allocation. Journal of Hebei University: Natural Science Edition*, 21(3), pp. 6.

Chen, A. and Liu, D., 2001. *Joint trajectory planning for coordinated motion of dual arm robots. Journal of Xinyang Normal University: Natural Science Edition*, 14(1), pp. 4.

Erkorkmaz, K. and Altintas, Y., 2001. *High speed CNC system design. Part I: jerk limited trajectory generation and quintic spline interpolation. International Journal of Machine Tools and*

*Manufacture*, 41(9), pp. 1323–1345.

Fu, H., Zhong, S. and Luo, Z., 2021. *Development and Application of Joint Angle Compensation Algorithm for Industrial Robots. Metrology and Testing -8-Technology*, 10, pp. 048.

Jiang, Y., Li, J. and Zhang, G., 2005. *Motion Planning of Joint type Redundant Robots. Combination Machine Tool and Automation Processing Technology*, 3, pp. 2.

Liang, L., Wu, C. and Liu, S., 2025. *Absolute Position Accuracy Calibration Algorithm for Robots Based on Joint Geometric Errors. Journal of Northeastern University (Natural Science Edition)*, 46(4), pp. 1–7.

Liu, Q., You, X., Zhang, X., Zuo, J. and Li, J., 2025. *Overview of Path Planning Algorithms for Mobile Robots. Computer Science*, 52(6A), pp. 240900074–10.

Ma, J., 2009. *Motion Planning for Complex Robots with Series Parallel Joints. Thesis, Harbin University of Science and Technology.*

Nguyen, K., Ng, T. and Chen, I., 2008. *On algorithms for planning S-curve motion profiles. International Journal of Advanced Robotic Systems*, 5(1), pp. 99–106.

Qiu, B. et al., 2022. *Synchronization planning of robot joint trapezoidal velocity trajectory for optimal energy consumption. Mechanical Design and Research*, 4, pp. 038.

Siciliano, B., Sciavicco, L., Villani, L. and Oriolo, G., 2009. *Robotics: Modelling, Planning and Control. London: Springer.*

Spong, M.W., Hutchinson, S. and Vidyasagar, M., 2005. *Robot Modeling and Control. New York: John Wiley & Sons.*

Wang, R. and Yu, Y., 2004. *Terminal positioning planning for coordinated operation of flexible robots. Mechanical Design and Research*, 1, pp. 3.

Wang, M. and Wu, T, 2005. *Research on Multi Robot Collaborative Motion Planning and Related Issues. Manufacturing Automation*, 27(5), pp. 6.

Wang, M. et al., 2011. *Distributed Collaborative Motion Planning for Multiple Joint Robots. Information and Control.*

Wu, W., 2013. *Friction analysis and low-speed high-precision motion control of multi degree of freedom serial robot joints. Thesis, Zhejiang University.*

Wu, Y. et al., 2021. *Power Equivalent Model and Parameter Identification of Industrial Robots. Journal of Chongqing University*, 44.

Xie, C., Zheng, S. and Qu, H, 1980. *Calculation of Position and Attitude of Industrial Robots and Determination of Joint Motion Parameters. Mechanical Industry Automation*, 4, pp. 50–56.

Yuan, B., 2025. *Research on Jump Control Method of Four legged Robot Based on Deep Reinforcement Learning. Thesis, Wuhan University of Science and Technology.*

Zeng, Q. and Fang, Y, 2007. *Integrated coordinated motion planning for bipedal walking robots. Mechanical*, 34(2), pp. 5.

Zhang, B., 2011. *Multi constraint based robot joint space trajectory planning. Journal of Mechanical Engineering*, 47(21), pp. 6.

Zhang, M. et al, 2017. *Research on Linear Walking Gait Planning and Joint Trajectory of bipedal Walking Robot. Forestry Machinery and -9-Woodworking Equipment*, 6, pp. 045.

Zhang, X., 2023. *Research on Motion Control Method of Four legged Robot Based on Trajectory Optimization. Thesis, University of Electronic Science and Technology of China.*

Zhang, X. et al, 2024. *Trajectory optimization and control method for quadruped robot jumping in the air. Journal of Beijing Jiaotong University*, 48(3), pp. 161–170.

Zhang, X., Sun, G., Zhou, H., Liu, Y. and Li, W, (2024). *A series parallel hybrid bionic robotic arm with flexible driving joints. Journal of Beijing Jiaotong University*, 48(6), pp. 154–161.

Zhu, H., 2024. *Optimization control method for motion trajectory accuracy of quadruped robot based on joint angle compensation. Modeling and Simulation*, 13(3), pp. 2305–2314.

HIGHER EDUCATION PUBLISHING MANAGEMENT

高图学术出版

**《跨学科创新与融合研究》**

（Journa of Transdisciplinary Innovation and Integration，JTII）


致力于快速传播新兴跨学科与多学科研究成果。倡导突破学科边界的创新研究，关注知识交汇领域的变革性思想。

类型：中文国际期刊　ISSN：印刷 3104-6991 / 在线 3104-7009

官网：j.hkiher.ac.cn/index.php/JTII

计划检索：维普，谷歌学术，中国知网等

周期：双月刊（每年6期）

征稿范围：数据科学与社会科学、工程学与生命科学、环境科学与经济学、材料科学与艺术/设计、神经科学与教育/心理学、管理科学与信息技术。

**《应用人工智能与跨学科创新》**

（Journal of Applied AI and Interdisciplinary Innovations，JAAI-II）


聚焦于人工智能技术在现实问题中的跨学科应用研究。推动人工智能与多领域融合，催生具有范式变革意义的高影响力成果。

类型：英文国际期刊　ISSN：印刷 3104-7017 / 在线 3104-7025

官网：j.hkiher.ac.cn/index.php/JAAI-II

计划检索：维普，谷歌学术，中国知网等

周期：双月刊（每年6期）

征稿范围：人工智能应用（如医疗健康、金融科技等）、核心人工智能技术（如机器学习、自然语言处理等）。

**《融合科学研究与实践》**

（Journal of Convergent Science Research and Practice, JCSRP）


聚焦多学科深度交叉与实践应用，鼓励突破传统学科界限，加速前沿融合研究向具有实质社会影响力的全球实践转化。

类型：英文国际期刊

官网：https://j.hkiher.ac.cn/index.php/jcsrp

周期：双月刊（每年6期）

计划检索：维普，谷歌学术，中国知网等

征稿范围：理工科与应用技术、人文与社会科学、可持续发展，以及健康科学（含公共卫生、护理与营养科学）。

**我们的三本中英文国际学术期刊分别涵盖跨学科研究、应用人工智能、可持续大健康、融合科学、多学科深度交叉与实际应用等研究领域，均由香港高等教育研究所主办。我们始终坚持普惠出版与学术快审计划，致力于推动学术成果的广泛传播并加入相应的中英文数据库检索。请各位作者(投稿人)根据研究方向选择相应的中文或英文期刊。欢迎关注我们的公众号和小程序，了解最新学术资讯。**

香港高等教育研究所
HONG KONG INSTITUTE OF HIGHER EDUCATION RESEARCH

HIGHER EDUCATION PUBLISHING MANAGEMENT
高图学术出版

Bridging Disciplinary Boundaries, Inspiring Future Insights
We warmly invite scholars, educational institutions, and industry leaders worldwide to join the HKIHER global academic network. Together, we can chart new paths to address shared global challenges and contribute meaningful expertise to the scientific community.


# Journal of Applied AI and Interdisciplinary Innovations, JAAI-II

Hosted by the Hong Kong Institute of Higher Education Research (HKIHER), JAAI-II is a rigorous English-language journal dedicated to the convergence of applied AI and diverse academic disciplines. The journal focuses on the practical intersection of artificial intelligence with critical sectors, including healthcare, education, finance, and industry. Crucially, we also explore the integration of AI technologies within the social sciences, such as sociology and psychology. We actively seek exceptional, problem-driven research that connects AI to real-world contexts, aiming to catalyze paradigm shifts across both scientific and societal domains. Our ultimate goal is to provide a premier platform for global scholars, leveraging interdisciplinary insights to advance AI technology and tackle complex international challenges.

（ISSN 3104-7017、eISSN 3104-7025）

**Official Website：Hkiher.ac.cn**
**WeChat Official Account：高图学术**
**Manuscript Submissions：Submissions@hkiher.ac.cn**
**Partnerships & Enquiries：Info@hkiher.ac.cn**

## Disclaimer of Liability

The views and opinions expressed in this journal belong solely to the authors and do not represent the publisher, editorial board, or affiliated agents. The publisher accepts no liability for any damage, loss, or injury to persons or property resulting from the use of the content or cited materials herein. Publication does not imply an endorsement of any product's commercial viability or its fitness for a particular purpose.

Complimentary Copy (Not for Resale)