



Optimizing Local Government Performance Evaluation Systems Under Digital Transformation: A Public-Value Framework and an Expanded Case Study of Hangzhou, China

Taizhen Ge

Liyang Municipal Bureau of Industry and Information Technology, Liyang, Jiangsu, China
E-mail: getaizhen1999@163.com

Abstract: Digital transformation has moved local government from electronic service delivery to data-driven, platform-based, and increasingly algorithmic governance. Yet many performance evaluation systems still rely on periodic reporting, manually submitted indicators, fragmented departmental targets, and internal administrative assessment. These arrangements are poorly aligned with the speed, complexity, inclusiveness, and accountability requirements of digital governance. This article develops an optimized framework for local government performance evaluation under digital transformation and strengthens it with an expanded case analysis of Hangzhou, China. Drawing on digital-era governance, public value theory, data-driven public sector theory, and algorithmic accountability, the study proposes a parsimonious five-dimensional architecture: governance effectiveness, public service value, data and platform capability, innovation and learning capacity, and trust, transparency, and accountability. Methodologically, the article adopts a design-science-oriented conceptual approach, policy-document analysis, and an illustrative case study. The Hangzhou case is examined through three connected layers: the evolution of the City Brain as a digital governance infrastructure, the government-service management monitoring system as a performance data mechanism, and DB3301/T 0472-2024 as an institutionalized local standard for digital evaluation of law-based government construction. The analysis shows that high-quality digital performance evaluation should not be treated as a technical dashboard. It must be embedded in legal standards, cross-department data governance, public participation, budgetary and corrective incentives, and ethical safeguards for privacy and algorithmic fairness. The paper contributes to digital economy and public administration research by shifting the focus from digitizing old performance indicators to creating dynamic, public-value-oriented, and accountable evaluation systems.

Keywords: Digital government; digital transformation; local government performance evaluation; public value; data-driven governance

I. INTRODUCTION

Digital transformation has become a defining condition of contemporary public administration. Digital government no longer refers only to transferring administrative forms and services to websites. It increasingly involves integrated data infrastructures, interoperable platforms, automated decision support, proactive service delivery, and coordination across organizational boundaries. The United Nations E-Government Survey 2024 associates digital-government development with sustainable development, institutional capacity, citizen participation, and inclusion [1]. The OECD Digital Government Index similarly emphasizes digital-by-design government, a data-driven public sector, government as a platform, openness, user-driven services, and proactive administration [2].

China has placed digital government within a broader strategy of governance modernization. The State Council's guidelines on digital government construction emphasize data sharing, coordinated administration, public-service improvement, and government capacity [3]. Subsequent national initiatives have expanded attention to artificial intelligence, integrated data markets, smart-city development, whole-city digital transformation, and the intelligent provision of public services [4], [5]. These developments create new sources of administrative information and new possibilities for continuous performance monitoring. They also create a fundamental evaluation problem. Traditional local government performance systems were designed mainly for periodic administrative control. Such systems typically depend on annual reports, manually compiled statistics, departmental targets, and rankings based on input and output indicators. They may measure the number of completed projects, processed applications, online services, or administrative inspections without determining whether digital transformation has reduced citizens' administrative burden, improved accessibility, strengthened lawful data use, or increased public trust.

Digitizing a conventional indicator does not necessarily improve the quality of evaluation. A high online-processing rate may conceal repeated submissions, failed transactions, exclusion of older adults, or pressure on citizens to use digital channels. A real-time dashboard may present precise scores derived from incomplete, inconsistent, or strategically reported data. Artificial intelligence may improve prediction and classification while introducing opacity, automation bias, privacy risks, and unequal treatment.

The central research question is therefore:



How should local government performance evaluation be reconfigured under digital transformation so that it captures administrative effectiveness, public value, data capability, organizational learning, inclusion, legality, and accountable use of algorithms?

A second, case-oriented question follows:

What institutional design mechanisms and unresolved limitations can be identified in the publicly documented development of digital performance evaluation in Hangzhou?

The paper makes three contributions. First, it develops an explicit mapping between four theoretical traditions and five evaluation dimensions. This mapping prevents the framework from becoming an unconnected list of desirable indicators. Second, it reframes the Hangzhou case as a document-based institutional analysis rather than evidence that digital evaluation has produced verified performance improvements. Third, it introduces a critical distinction between documented institutional arrangements and outcomes that cannot be established without primary data, administrative datasets, algorithmic documentation, or citizen-level evidence.

II. LITERATURE REVIEW AND ANALYTICAL FOUNDATIONS

A. Government Performance Management and Its Digital Limitations

Government performance management has been strongly influenced by New Public Management, which emphasizes measurable outputs, managerial responsibility, efficiency, target setting, and performance incentives [6]. These principles helped governments move beyond assessments based exclusively on procedural compliance. However, target-centred evaluation can produce fragmentation, gaming, tunnel vision, and displacement of objectives.

These weaknesses become more significant in digital government. Departmental indicators may conflict with integrated services organized around citizens' life events. A department may meet its own processing target while transferring costs and procedural burdens to another department or to the citizen. Evaluation systems that reward online volume may encourage agencies to maximize transactions rather than simplify or eliminate unnecessary procedures.

Digital-era governance developed partly in response to fragmented managerial reforms. Dunleavy et al. identify reintegration, needs-based holism, and digitization as defining features of digital-era governance [7]. For performance evaluation, reintegration requires attention to cross-department coordination rather than isolated departmental outputs. Needs-based holism requires evaluating complete service journeys rather than individual administrative steps. Digitization enables real-time observation, but it does not itself determine which outcomes are socially valuable.

China's experience further shows that digitalization can support both service improvement and stronger administrative control. Research on local digital governance has demonstrated that Internet-based systems may expand responsiveness in some settings while simultaneously strengthening monitoring and hierarchical oversight in others [26], [27]. Performance evaluation must therefore analyze who defines the indicators, whose behaviour becomes visible, and how evaluation results are used.

B. Public Value and Citizen-Centred Evaluation

Public value theory provides a broader normative basis for performance assessment. Moore argues that public organizations create value when outcomes are substantively valuable, politically legitimate, and operationally feasible [8]. In digital government, public value includes efficiency but cannot be reduced to it. It also includes accessibility, fairness, privacy, transparency, participation, responsiveness, legality, and trust [28].

A service may be technically efficient yet fail to create public value. An automated approval system may reduce average processing time but offer inadequate explanation when applications are rejected. A digital platform may improve convenience for experienced users while imposing barriers on older adults, people with disabilities, rural residents, migrants, or citizens with limited digital literacy.

Citizen satisfaction is relevant but insufficient. Satisfaction scores may be affected by expectations, unequal access, fear of complaining, selection into online channels, or limited knowledge of legal entitlements. Recent research in China indicates that citizens' support for local digital-government programs is mediated by trust and policy expectations rather than determined only by platform experience [29]. Evaluation should therefore combine subjective experience with evidence of accessibility, equal treatment, complaint resolution, and procedural fairness.

C. Data-Driven Government and Platform Capability

A data-driven public sector treats data as a strategic resource for service delivery, policy design, resource allocation, anticipation, and accountability [9]. Digital platforms create potentially valuable performance evidence through transaction records, processing times, workflow logs, complaint data, electronic certificates, service ratings, inspection records, and system-use patterns.

However, data availability is not equivalent to data quality. Administrative datasets are produced within organizational processes and may contain inconsistent definitions, missing values, duplicated records, reporting incentives, and access restrictions. Indicators derived from these data can reproduce institutional biases or create false precision.

Platform governance therefore has both technical and institutional dimensions. Technical capacity includes interoperability, common data standards, metadata management, cybersecurity, and system reliability. Institutional capacity includes authority to share data, responsibility for correcting errors, rules governing secondary use, procurement arrangements, vendor dependence, and mechanisms for resolving disputes among departments.

Research on open government and government information systems has repeatedly shown that technological openness does not automatically produce accountability or public value [23], [24]. Evaluation systems must assess whether data infrastructures support lawful coordination and meaningful public oversight rather than merely whether platforms exist.

D. Innovation, Learning, and Adaptive Capacity

Digital transformation is not a one-time technology project. It requires experimentation, organizational learning, redesign of work processes, development of staff capabilities, and adaptation to changing legal and social expectations. Research on digital transformation emphasizes organizational change and institutional restructuring rather than simple technology adoption [22].

An evaluation system focused only on compliance may discourage experimentation. Conversely, an evaluation system that rewards the number of digital applications may encourage project proliferation without evidence of usefulness. Innovation indicators should therefore consider whether pilots solve a defined problem, whether results are evaluated, whether unsuccessful projects are discontinued, whether successful applications are scaled responsibly, and whether lessons are transferred across agencies.

Recent research on Chinese local officials identifies different orientations toward digital government, including data-sharing, service-improvement, and risk-awareness perspectives. These orientations are shaped by practical demands and accountability pressures [30]. Evaluation systems should recognize this variation and avoid assuming that digital innovation is driven by a single set of motivations.

E. Algorithmic Accountability, Legality, and Digital Inclusion

Artificial intelligence and algorithmic systems are increasingly used in risk prediction, fraud detection, administrative inspection, resource allocation, document review, and service triage. Such systems may improve speed and consistency but can also create automation bias, opacity, discrimination, and accountability gaps [10], [11].

The EU Artificial Intelligence Act adopts a risk-based approach to the regulation of artificial intelligence [12]. China's Personal Information Protection Law establishes obligations concerning lawful processing, transparency, individual rights, and automated decision-making [13]. These developments demonstrate that digital performance evaluation is partly a legal and ethical governance problem. Algorithmic accountability is especially important when automated indicators affect organizational rankings, budget allocation, personnel appraisal, inspection frequency, or access to public services. At minimum, evaluation systems require documentation of indicator definitions, data provenance, error rates, model limitations, human review, audit trails, appeal procedures, and periodic validity testing. Digital inclusion must also be incorporated into evaluation design. Online feedback data are not representative when certain populations are unable or unwilling to use digital channels. A high digital-service utilization rate may result from the removal of offline alternatives rather than genuine citizen preference. Performance evaluation should therefore examine channel choice, assisted digital services, accessibility standards, demographic differences, and the experiences of citizens who fail to complete digital procedures.

F. Mapping Theory to Evaluation Dimensions

Table I provides the theoretical derivation of the proposed framework. Each dimension is associated with a distinct analytical problem and theoretical foundation.

TABLE I THEORETICAL FOUNDATIONS OF THE FIVE-DIMENSIONAL FRAMEWORK

Theoretical foundation	Central concern	Corresponding evaluation dimension	Main evaluative question
New Public Management and performance management	Results, efficiency, measurable responsibility	Governance effectiveness	Does digital transformation improve coordination, responsiveness, and process outcomes?
Digital-era governance	Reintegration, needs-based service, digital coordination	Governance effectiveness and data/platform capability	Are services integrated across organizational boundaries and designed around user needs?
Public value theory	Legitimacy, citizen value, fairness, trust	Public service value and accountability	Do digital systems create benefits that citizens can access fairly and regard as legitimate?
Data-driven public sector and platform governance	Data quality, interoperability, lawful reuse	Data and platform capability	Are data reliable, shareable, secure, and institutionally governed?
Digital transformation and organizational learning	Adaptation, experimentation, capability building	Innovation and learning capacity	Does government learn from digital projects and scale them responsibly?
Algorithmic accountability and administrative law	Explainability, privacy, fairness, human oversight	Trust, transparency, and accountability	Can digital and algorithmic decisions be explained, audited, challenged, and corrected?

The five dimensions are analytically distinct but mutually dependent. Governance effectiveness cannot be assessed reliably without high-quality data. Public service value depends partly on accessibility and procedural accountability. Innovation without ethical safeguards may create risk rather than value. Transparency without correction mechanisms may disclose problems without changing administrative behaviour.

III. RESEARCH DESIGN AND METHODOLOGY

A. Research Positioning

The study adopts qualitative policy-document analysis and an illustrative institutional case study. It does not claim to estimate causal effects, evaluate citizen outcomes, or validate the proposed framework statistically.

Although the framework has a design-oriented purpose, the study is not presented as formal design science research. Established design science research requires the creation of an identifiable artifact and an explicit process of demonstration and evaluation [31], [32]. The present study does not test the framework through iterative implementation, stakeholder evaluation, experimental comparison, or organizational deployment. It therefore offers a conceptually derived and policy-informed framework that requires subsequent empirical validation.

B. Document Corpus

The analysis draws on four categories of material.

First, international frameworks and scholarly literature were used to identify general principles of digital government, public value, data-driven administration, and algorithmic accountability [1], [2], [6]–[13], [22]–[30].

Second, national Chinese laws, strategies, and policy documents were examined to identify institutional expectations concerning digital government, personal information protection, data sharing, artificial intelligence, and whole-city digital transformation [3]–[5], [13].

Third, Hangzhou-level standards, plans, government reports, and administrative materials were used to reconstruct the documented institutional development of digital performance evaluation [14]–[21].

Fourth, state and local media reports were used as supplementary sources for initiatives not fully described in publicly accessible administrative documents. Media reports were treated as descriptive evidence of officially presented practices rather than independent verification of implementation outcomes.

The source hierarchy gives greater evidentiary weight to laws, formal standards, and official policy documents than to promotional reports or media summaries. Claims concerning the existence of a standard, platform, institution, or policy objective are distinguished from claims concerning actual outcomes.

C. Analytical Procedure

The analysis proceeded in three stages. First, theoretical concepts were coded into an initial analytical structure covering efficiency, integration, citizen value, data governance, innovation, legality, inclusion, transparency, and accountability. Second, overlapping concepts were consolidated into the five dimensions presented in Table I, with an emphasis on parsimony rather than comprehensive enumeration, because a framework containing excessive indicators would be difficult to interpret and could reproduce the fragmentation it seeks to address. Third, the Hangzhou documents were examined for evidence of institutional mechanisms corresponding to these dimensions, with particular attention to the conversion of digital infrastructure into evaluation evidence, the interaction between municipal standardization and district-level experimentation, the incorporation of participation and external supervision, and the connection between evaluation results and corrective action. In addition, a separate evidence-gap analysis was conducted to identify issues that could not be determined from the document corpus, including indicator validity, causal outcomes, the representativeness of citizen participation, internal algorithmic design, and privacy practices.

D. Methodological Limitations

Document analysis is appropriate for examining formal institutional design, policy evolution, indicator structures, and official governance rationales. It is not sufficient for determining whether systems operate as described, whether officials comply with formal rules, whether citizens experience improved services, or whether evaluation results influence budgets and personnel decisions.

The analysis is also affected by source-selection bias. Publicly available materials tend to emphasize innovation and institutional achievement. Internal disagreements, unsuccessful projects, implementation costs, data-quality failures, and citizen dissatisfaction may be underreported. Consequently, the Hangzhou case is used to generate propositions and identify research questions rather than to establish a best-practice model.

IV. A PUBLIC-VALUE FRAMEWORK FOR DIGITAL GOVERNMENT PERFORMANCE EVALUATION

A. Design Principles

The framework is based on six principles.

First, evaluation should be oriented toward public value rather than technology adoption. The number of platforms, applications, datasets, or artificial intelligence projects is not itself evidence of improved government performance.

Second, evaluation should be data-informed rather than data-deterministic. Administrative and platform data can improve timeliness and precision, but their meaning requires contextual interpretation and human judgment.

Third, performance should be evaluated across organizational boundaries. Citizen-facing outcomes often depend on multiple departments, shared data, and coordinated procedures.

Fourth, outcome measures should be combined with process and rights-based measures. Faster decisions are not necessarily better when procedures are unlawful, inaccessible, or opaque.

Fifth, evaluation should support learning and correction rather than function only as a ranking instrument.

Sixth, digital and algorithmic evaluation should be subject to privacy, fairness, explainability, and human-oversight requirements.

B. Five-Dimensional Architecture

TABLE II PUBLIC-VALUE FRAMEWORK FOR DIGITAL LOCAL GOVERNMENT PERFORMANCE EVALUATION

Dimension	Core question	Illustrative evidence	Main risks
Governance effectiveness	Does digital transformation improve complete administrative and service processes?	End-to-end completion time, repeated-submission rate, cross-department processing, complaint resolution, error correction	Optimizing departmental scores while transferring burden to citizens
Public service value	Are services useful, accessible, fair, and responsive?	Completion rates, user experience, accessibility audits, demographic disparities, offline support, appeal outcomes	Satisfaction bias, exclusion of digitally vulnerable groups
Data and platform capability	Are evaluation data reliable, interoperable, secure, and lawfully used?	Metadata completeness, data-quality audits, sharing records, access controls, system reliability, privacy assessments	False precision, data silos, unlawful secondary use, vendor lock-in
Innovation and learning capacity	Can government test, evaluate, adapt, and scale digital reforms responsibly?	Pilot evaluation, staff training, discontinuation of ineffective projects, cross-agency learning, post-implementation review	Application proliferation, symbolic innovation, compliance-oriented experimentation
Trust, transparency, and accountability	Are indicators and algorithmic processes legitimate, explainable, auditable, and correctable?	Indicator documentation, public disclosure, audit trails, third-party review, human oversight, appeal mechanisms	Opaque scoring, automation bias, strategic gaming, unchallengeable decisions

C. Governance Effectiveness

Governance effectiveness evaluates whether digitalization improves complete processes rather than isolated activities. Processing time remains relevant, but it should be measured from the user's initial application to final resolution. Repeated requests for the same information, transfers among departments, failed digital submissions, and unresolved complaints should be included.

Cross-department coordination is especially important. An apparently efficient department may rely on another agency to resolve data inconsistencies or may require citizens to coordinate between departments themselves. Evaluation should identify whether administrative burden has been reduced or merely relocated.

Real-time monitoring can support early warning, but short-term fluctuations require interpretation. A sudden increase in processing time may reflect system failure, increased demand, policy changes, or improved inclusion of complex cases. Performance systems should therefore link alerts to diagnostic review.

D. Public Service Value

Public service value assesses whether citizens and firms receive meaningful benefits. Indicators should include accessibility, successful completion, procedural fairness, clarity of information, complaint resolution, and channel availability.

Disaggregated analysis is essential. Aggregate satisfaction may conceal differences among older adults, people with disabilities, rural residents, migrants, low-income groups, and people with limited digital literacy. Evaluation systems should compare outcomes across groups while respecting privacy and avoiding discriminatory profiling.

Offline and assisted channels should not automatically be treated as evidence of digital failure. For some users, in-person or telephone assistance is an essential component of inclusive digital government. A mature evaluation system measures whether citizens can choose an appropriate channel and move between channels without repeating the entire process.

E. Data and Platform Capability

This dimension assesses the reliability of the infrastructure underlying evaluation. Indicators should cover common definitions, metadata, completeness, timeliness, duplication, interoperability, data lineage, access control, and correction procedures.

Data reuse for evaluation requires particular scrutiny. Information collected to provide a service may later be used to rank departments, predict risk, or evaluate employees. Secondary use should have a clear legal and administrative basis. Access should be proportionate to purpose, and sensitive information should be minimized or anonymized where appropriate.

Platform capability should also include institutional sustainability. Dependence on a single vendor, undocumented proprietary scoring systems, or limited internal technical capacity can weaken accountability. Governments should retain sufficient knowledge and contractual rights to inspect, modify, and audit evaluation systems.

F. Innovation and Organizational Learning

Innovation should be evaluated according to problem relevance, evidence of improvement, scalability, risk management, and learning. Counting applications or pilots rewards activity rather than value.

A useful digital pilot should begin with a clearly defined administrative or public-service problem. The government should identify intended users, expected outcomes, data requirements, legal constraints, and conditions for termination. Post-implementation review should examine not only success but also unintended effects.

Learning-oriented evaluation should distinguish between avoidable failure and responsible experimentation. Punishing all unsuccessful pilots can discourage innovation, while rewarding projects without evaluation encourages symbolic adoption.

The evaluation system should ask whether lessons were documented and transferred to other departments.

G. Trust, Transparency, and Accountability

This dimension addresses the legitimacy of the evaluation system itself. Departments and citizens should be able to understand what is being measured, how indicators are calculated, which data sources are used, and how errors can be corrected.

When algorithmic systems are involved, documentation should include model purpose, data provenance, performance limitations, validation procedures, human-review responsibilities, and appeal channels. Higher-risk uses require stronger safeguards.

Transparency does not necessarily require disclosure of personal data, security-sensitive information, or proprietary technical details. It requires sufficient explanation for affected actors to understand and challenge consequential decisions.

V. DOCUMENT-BASED INSTITUTIONAL ANALYSIS OF HANGZHOU

A. Case Selection and Evidentiary Boundaries

Hangzhou is an appropriate illustrative case because it has accumulated experience in urban digital infrastructure, integrated government services, local standardization, and administrative performance management. The city has also been discussed in scholarship on Chinese government performance management [33].

The case is not used to claim that Hangzhou's digital systems have produced superior citizen outcomes. Publicly available materials do not provide sufficiently detailed administrative microdata, independent evaluation, internal scoring rules, or representative citizen surveys. The analysis instead examines how the city's documented reforms have sought to institutionalize digital evaluation.

B. Institutional Development

TABLE III

DOCUMENTED DEVELOPMENT OF DIGITAL EVALUATION IN HANGZHOU

Period	Documented development	Institutional significance	Evidence not publicly established
2016–2018	City Brain developed from traffic management toward broader urban governance [14]	Created data and platform infrastructure for observing urban operations	Causal effects, full cost structure, data representativeness
2021	Xiaoshan digital intelligence index introduced for government-service evaluation [19]	Demonstrated district-level visual monitoring and experimentation	Weight validity, employee responses, citizen-level outcomes
2022–2023	Government-service management monitoring system linked operational data to reform supervision [15], [16]	Connected project implementation with continuous monitoring	Internal indicator definitions, data-quality audit procedures
2024	DB3301/T 0472-2024 standardized digital evaluation of law-based government construction [17], [18]	Formalized multidimensional indicators and supervisory procedures	Inter-rater reliability, weighting rationale, correction outcomes
2025	Assessment centre, Hangzhou Index, and closed-loop supervision were publicly reported [20]	Suggested movement from temporary projects to specialized institutional capacity	Institutional independence, external verification, measurable effects
2026 onward	Planning documents propose City Brain 3.0, digital twins, trusted data spaces, government agents, and City Brain GPT [21]	Extends evaluation challenges to generative AI and intelligent agents	Model governance, data provenance, auditability, bias testing

C. Infrastructure-to-Evaluation Conversion

The City Brain forms the infrastructural background of Hangzhou's digital governance. It was initially associated with traffic management and later expanded toward broader urban operations [14]. Its relevance to performance evaluation lies in the production of continuous operational data.

Nevertheless, infrastructure does not automatically become an evaluation system. Data must be converted into indicators, standards, responsibilities, review processes, and correction mechanisms. The Hangzhou case indicates a gradual attempt to make this conversion.

The government-service management monitoring system represents one such step. Public reports describe the use of objective operational data to track reform projects across formulation, implementation, supervision, and evaluation [15]. The system reportedly assesses government services across dimensions such as coverage, intelligent processing, efficiency, public perception, transaction transparency, and integrated governance [16].

These materials indicate an institutional intention to move beyond annual self-reporting. They do not reveal the exact data definitions, weighting formulas, missing-data procedures, or statistical validation of the indicators. The case therefore documents an evaluation architecture, not its demonstrated accuracy.

D. Standardization and Nested Experimentation

A central feature of the Hangzhou case is the interaction between municipal standardization and district-level experimentation.

The Xiaoshan digital intelligence index illustrates bottom-up experimentation. The reported system used multiple data sources and a visual interface to evaluate government-service performance [19]. Such district-level systems may detect specific operational problems more rapidly than a citywide annual assessment.

They may also produce fragmented definitions and incompatible scores. Municipal standards can improve comparability but risk suppressing local adaptation. Hangzhou's documented trajectory suggests a nested model in which district experiments operate within broader municipal rules.

DB3301/T 0472-2024 is particularly important because it extends digital evaluation into law-based government construction [17]. According to public descriptions, the standard includes legislative work, administrative decision-making, administrative law enforcement, judicial supervision, and legal publicity, with multiple second- and third-level indicators [18].

This broadens digital performance evaluation beyond speed and convenience. It recognizes that government quality includes lawful procedures, consultation, legality review, standardized enforcement, and responses to judicial or procuratorial recommendations.

However, converting legal and procedural values into indicators introduces new questions. Formal compliance can be counted more easily than substantive quality. A consultation may be recorded without meaningfully influencing a decision. A legality review may exist procedurally without challenging the proposing agency. Future research should therefore examine whether digital indicators capture substantive legality or primarily document completion of required steps.

E. Participation and External Judgment

Official descriptions of the Hangzhou standard emphasize participation by people's congress representatives, government departments, grassroots residents, experts, and third parties [18]. This is significant because digital evaluation cannot rely exclusively on administrative data.

Participation can improve interpretation and legitimacy, but its quality depends on selection, representativeness, access to information, and influence over final judgments. Public documents do not reveal how participants are recruited, how conflicting assessments are reconciled, or whether feedback changes scores.

Digital participation can also reproduce inequality. Citizens who use government platforms more frequently are more likely to appear in feedback data. Those who abandon a digital process may never reach the rating stage. People who depend on family members or frontline staff may be statistically invisible. Inclusive evaluation requires deliberate sampling beyond platform users.

F. Closed-Loop Correction

Hangzhou's public materials repeatedly emphasize collection, evaluation, tracking, feedback, rectification, and re-evaluation [18], [20]. This closed-loop orientation is potentially more important than the production of rankings.

An evaluation system creates public value only when identified problems lead to organizational response. This may involve process redesign, data correction, staff training, technical repair, budget adjustment, or revision of indicators.

Yet the existence of a corrective process does not establish its effectiveness. Future studies should examine the time between problem identification and correction, the proportion of corrective actions completed, the recurrence of identified problems, and whether departments can challenge inaccurate scores.

G. Digital Inclusion and Equity Gaps

The publicly documented Hangzhou systems appear stronger in operational monitoring and standardization than in explicit measurement of digital inequality. Available reports discuss public perception and grassroots participation, but they provide limited detail concerning demographic representativeness, accessibility audits, or differences between online and offline users.

This is an important gap. Hangzhou is digitally advanced, but a high level of urban digitalization does not eliminate differences in age, disability, income, migration status, education, or digital literacy. Performance evaluation should report service completion and failure rates by relevant groups, subject to privacy safeguards.

Digital inclusion should also be evaluated spatially. Central urban districts may benefit from stronger infrastructure and staff capacity than peripheral areas. Citywide averages may conceal these differences.

H. Algorithmic Accountability and the Next Generation of Evaluation

Hangzhou's recent planning documents propose City Brain 3.0, digital twins, trusted data spaces, government intelligent agents, and an evolution from City Brain applications toward City Brain GPT [21]. These developments increase the importance of model governance.

Generative AI and government agents can support document processing, service guidance, urban monitoring, and administrative coordination, but they may also generate inaccurate outputs, obscure information sources, expose sensitive data, and encourage excessive reliance by public officials. Future evaluation frameworks should therefore require clear documentation of model purposes and authorized uses, provenance and quality controls for training and retrieval data, testing for accuracy, robustness, and unequal effects across groups, and, where legally appropriate, logging of prompts, outputs, human edits, and approval decisions. They should also establish clear responsibility when AI-supported recommendations are incorrect, provide notification and appeal mechanisms when automated systems materially affect citizens' rights or access to services, and regularly assess whether the use of AI produces measurable public value. Accordingly, the next stage of digital government evaluation should not focus merely on how frequently public agencies adopt artificial intelligence but should determine whether such systems are necessary, lawful, accurate, explainable, proportionate, and subject to meaningful human oversight.

VI. DISCUSSION

A. Theoretical Contribution

The framework links four theoretical traditions that are often discussed separately. Digital-era governance explains the need for integrated evaluation across departments. Public value theory establishes the normative criteria of usefulness, fairness, legitimacy, and trust. Data-driven government explains the informational infrastructure required for continuous evaluation. Algorithmic accountability identifies safeguards necessary when evaluation becomes automated or AI-assisted. The framework also distinguishes performance of digital government from performance of the evaluation system. A local government may provide effective services while using an unreliable scoring model. Conversely, a technically sophisticated evaluation system may monitor services that remain inaccessible or unfair.

This distinction is increasingly important because digital evaluation systems can shape organizational behaviour. Indicators influence what departments prioritize, what data they collect, and which problems receive resources. Evaluation architecture is therefore part of governance rather than a neutral measurement layer.

B. What the Hangzhou Case Supports

The documentary evidence supports four limited conclusions.

First, digital infrastructure can create the informational conditions for continuous monitoring, but separate institutional mechanisms are needed to transform operational data into evaluation.

Second, municipal standardization and district experimentation can coexist in a layered governance structure.

Third, public and professional judgment remains necessary because administrative data cannot independently establish legality, fairness, or public value.

Fourth, closed-loop correction is a more meaningful design principle than one-time scoring or ranking.

The evidence does not establish that Hangzhou's evaluation systems have improved citizen satisfaction, reduced inequality, strengthened trust, or produced more lawful administrative behaviour. Such conclusions require primary and outcome data.

C. Institutional Trade-Offs

Digital performance evaluation involves several trade-offs.

Standardization improves comparability but may reduce sensitivity to local conditions. Real-time monitoring improves responsiveness but may increase surveillance and short-termism. Public disclosure strengthens accountability but may encourage score management. Algorithmic evaluation can process complex information but may obscure judgment and responsibility.

Participation can improve legitimacy but may become symbolic when participants lack information or influence. Inclusion measures can reveal unequal outcomes but may require collection of sensitive demographic data. Data sharing can improve coordination but expand privacy and cybersecurity risks.

A credible evaluation system should make these trade-offs visible rather than presenting all dimensions as mutually reinforcing.

D. Implications for Chinese Digital Governance Research

The analysis contributes to research on China by placing local digital performance evaluation within a hierarchical but experimentally diverse governance context. National policies define strategic priorities, municipal governments create standards and platforms, and districts develop operational applications.

This structure can accelerate diffusion, but it can also intensify compliance pressures. Recent research indicates that Chinese local officials may approach digital government through practical orientations shaped by hierarchical and public accountability [30]. Evaluation systems may therefore encourage officials to pursue measurable platform outputs while paying less attention to values that are difficult to quantify.

Future research should examine how officials interpret indicators, negotiate conflicting targets, respond to algorithmic scores, and balance upward accountability with citizen-centred outcomes.

E. Research Agenda

The proposed framework requires empirical testing.

Cross-city research could compare indicator structures, data governance, participation, and corrective mechanisms. Panel studies could examine whether reforms affect processing time, complaint resolution, public trust, and service inequality.

Interviews with officials, platform developers, frontline staff, auditors, citizens, and third-party evaluators could reveal how systems operate in practice. Observation of evaluation meetings and correction processes would help distinguish formal procedures from actual decision-making.

Technical audits could examine indicator validity, model bias, data completeness, and explainability. Survey and experimental research could test how different forms of transparency influence citizen trust and official behaviour.

Research should also investigate distributional effects. Digital evaluation cannot be regarded as successful when average service performance improves while disadvantaged users experience greater difficulty.

VII. CONCLUSION

Digital transformation requires a corresponding transformation in government performance evaluation. Traditional systems based on periodic reporting, departmental targets, and narrow output measures cannot adequately assess integrated services, data infrastructures, digital inclusion, artificial intelligence, and algorithmic accountability.

This paper proposes a five-dimensional framework comprising governance effectiveness, public service value, data and platform capability, innovation and organizational learning, and trust, transparency, and accountability. Its central

argument is that digital performance evaluation should be understood as an institutional governance system rather than a technical dashboard.

The document-based analysis of Hangzhou identifies a progression from digital infrastructure to monitoring, standardization, participation, and closed-loop supervision. It also reveals substantial evidentiary gaps. Public materials do not permit conclusions about causal outcomes, weighting validity, citizen representativeness, privacy implementation, or algorithmic fairness.

Hangzhou should therefore be treated neither as a universally transferable model nor as proof of successful digital performance reform. It is a theoretically informative case that shows how digital evaluation can become institutionally embedded while also demonstrating why stronger empirical validation and critical scrutiny are necessary.

The quality of digital government depends partly on what evaluation systems define as valuable. Systems that reward platform construction, transaction volume, and rapid processing alone may reproduce narrow administrative priorities. Systems that combine effectiveness with fairness, legality, learning, inclusion, transparency, and correction are more likely to support accountable digital transformation.

REFERENCES

- [1] United Nations Department of Economic and Social Affairs, *E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development*. New York, NY, USA: United Nations, 2024.
- [2] OECD, *2023 OECD Digital Government Index: Results and Key Findings*. Paris, France: OECD Publishing, 2024, doi: 10.1787/1a89ed5e-en.
- [3] State Council of the People's Republic of China, "Guiding opinions on strengthening digital government construction," Beijing, China, Jun. 2022.
- [4] National Data Administration of China, *Digital China Development Report 2025*. Beijing, China, 2026.
- [5] National Development and Reform Commission, National Data Administration, Ministry of Finance, Ministry of Housing and Urban-Rural Development, and Ministry of Natural Resources, "Action plan for deepening smart-city development and promoting whole-city digital transformation," *Fa Gai Shu Ju* [2025] No. 1306, Oct. 2025.
- [6] C. Pollitt and G. Bouckaert, *Public Management Reform: A Comparative Analysis—Into the Age of Austerity*, 4th ed. Oxford, U.K.: Oxford University Press, 2017.
- [7] P. Dunleavy, H. Margetts, S. Bastow, and J. Tinkler, *Digital Era Governance: IT Corporations, the State, and E-Government*. Oxford, U.K.: Oxford University Press, 2006.
- [8] M. H. Moore, *Creating Public Value: Strategic Management in Government*. Cambridge, MA, USA: Harvard University Press, 1995.
- [9] OECD, *The Path to Becoming a Data-Driven Public Sector*. Paris, France: OECD Publishing, 2019, doi: 10.1787/059814a7-en.
- [10] K. Levy, K. Chasalow, and S. Riley, "Algorithms and decision-making in the public sector," *Annu. Rev. Law Soc. Sci.*, vol. 17, pp. 309–334, 2021.
- [11] S. Alon-Barkat and M. Busuioc, "Human–AI interactions in public sector decision making: Automation bias and selective adherence to algorithmic advice," *J. Public Adm. Res. Theory*, vol. 33, no. 1, pp. 153–169, 2023, doi: 10.1093/jopart/muac007.
- [12] European Parliament and Council of the European Union, "Regulation laying down harmonised rules on artificial intelligence," *Regulation (EU) 2024/1689*, 2024.
- [13] National People's Congress of the People's Republic of China, *Personal Information Protection Law of the People's Republic of China*. Beijing, China, 2021.
- [14] Hangzhou Municipal People's Government, "Ensuring that City Brain construction remains at the forefront and building a national digital governance city," Hangzhou, China, Jan. 2020.
- [15] Hangzhou Daily, "Improving government services and the business-environment service window," Mar. 29, 2023.
- [16] Xinhua News Agency, "Hangzhou improves the business environment through accelerated government services," Oct. 25, 2023.
- [17] Hangzhou Municipal Administration for Market Regulation, *DB3301/T 0472-2024 Digital Evaluation Specification for Law-Based Government Construction*. Hangzhou, China, 2024.
- [18] Xinhua News Agency, "Hangzhou standardizes digital evaluation of law-based government construction," Nov. 30, 2024.
- [19] China News Service, "Digital technology empowers government-service evaluation," Oct. 3, 2022.
- [20] Hangzhou News, "Hangzhou digital supervision initiative for law-based government construction receives award," Jun. 17, 2025.
- [21] Hangzhou Municipal People's Government, *Outline of the 15th Five-Year Plan for National Economic and Social Development of Hangzhou*. Hangzhou, China, 2026.
- [22] I. Mergel, N. Edelmann, and N. Haug, "Defining digital transformation: Results from expert interviews," *Gov. Inf. Q.*, vol. 36, no. 4, Art. no. 101385, 2019, doi: 10.1016/j.giq.2019.06.002.
- [23] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *Inf. Syst. Manage.*, vol. 29, no. 4, pp. 258–268, 2012, doi: 10.1080/10580530.2012.716740.
- [24] J. C. Bertot, P. T. Jaeger, and J. M. Grimes, "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies," *Gov. Inf. Q.*, vol. 27, no. 3, pp. 264–271, 2010, doi:

10.1016/j.giq.2010.03.001.

- [25] S. Y. Tan and A. Taeihagh, "Smart city governance in developing countries: A systematic literature review," *Sustainability*, vol. 12, no. 3, Art. no. 899, 2020, doi: 10.3390/su12030899.
- [26] R. Han and L. Jia, "Governing by the Internet: Local governance in the digital age," *J. Chin. Gov.*, vol. 3, no. 1, pp. 67–85, 2018, doi: 10.1080/23812346.2018.1429175.
- [27] K. R. Foster and Y. Yang, "Increasing citizen access and local government responsiveness in Yichang, China," *Int. Public Manage. J.*, vol. 21, no. 3, pp. 392–414, 2018, doi: 10.1080/10967494.2017.1399945.
- [28] P. Panagiotopoulos, B. Klievink, and A. Cordella, "Public value creation in digital government," *Gov. Inf. Q.*, vol. 36, no. 4, Art. no. 101421, 2019, doi: 10.1016/j.giq.2019.101421.
- [29] Y. Zhang, "Why do citizens support digital programmes of local government? Evidence from China," *Asian J. Technol. Innov.*, early access, Apr. 2026, doi: 10.1080/19761597.2026.2651539.
- [30] Z. Wang and X. Yang, "Instrument-based or value-oriented? A Q-methodology for local public officials' perceptions on digital government," *Public Perform. Manage. Rev.*, early access, Feb. 2026, doi: 10.1080/15309576.2026.2627916.
- [31] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.
- [32] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *J. Manage. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.
- [33] B. Wu, *Government Performance Management in China: Theory and Practice*. Singapore: Springer, 2019, doi: 10.1007/978-981-13-8225-3.