Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.39078

Optimization of Transparency and Accountability Structures in Algorithm-Driven Public Governance

Jingyi Xin

The Chinese University of Hong Kong, Hong Kong, China
xinjingyi02390@outlook.com

Abstract. Governance by algorithm is widely applied in the public sector for the purposes of distributing welfare, estimating administrative risks, regulating urban behavior, rationing public services, and enabling automated decisions. However, transparency and accountability mechanisms are often scattered in the technical, legal, and organizational dimensions. In this work, we present an empirical-computational approach to improving transparency and accountability designs within algorithmic governance architectures in the public sector. We construct a governance assessment dataset based on policy guidance, procurement requirements, data inventories, model cards, decision logs, explanation logs, appeal decisions, audits, and stakeholder review annotations. Our methodology combines natural language processing, rule-based compliance coding, accountability graph construction, supervised classification, risk scoring, and intervention simulation. Using PostgreSQL, Elasticsearch, Doccano, Neo4j, NetworkX, XGBoost, LightGBM, Optuna, and PyGAD, we analyze measures of disclosure completeness, explanation quality, procedure traceability, responsibility path consistency, appeal availability, correction timeliness, and audit resolution efficiency. We show that structural interventions increase the transparency completeness score from 0.642 ± 0.071 to 0.813 ± 0.058 ($p < .001$) and decrease the aggregate Transparency- Accountability Risk Score by -0.184 ± 0.039 ($p < .001$).

Keywords: algorithm-driven public governance, transparency, governance optimization, administrative AI

1. Introduction

The research considers the application of algorithmic methods in public governance, namely in classification, selection, detection, distribution, risk prioritization, and decision support in administration [1, 2]. The systems operate by making sense of input data, producing predictions and recommendations, and, in certain cases, affecting administrative decisions. Thus, the efficiency of governance increasingly depends on the transparency, traceability, reviewability, and accountability of algorithms used in such processes [3]. In case the decision rules, data, explanations, and responsibility lines are not well understood, algorithmic governance may lead to the improvement in efficiency along with decreasing the trustworthiness and transparency of the processes [4].

The problem is that there is a disconnect between different aspects of governance. While there may be a model disclosing technical variables, it does not explain how a particular administrative

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.39078

decision was made. The agency may issue a policy document but not have full audit logs. A formal mechanism of review may be present but not linked with accountable persons, vendors, and institutions where the appeals take place. This poses problems in determining the parties accountable in case of faulty automated outputs, biases, delays, or poor explanations [5, 6]. Thus, assessment of the governance quality driven by algorithms should consider both model accuracy and completeness of procedural rules, the quality of explanations, their traceability and accessibility, and the effectiveness of corrections made [7].

In this work, we develop a framework to improve the transparency and accountability in algorithmic public governance. The data set for governance is created on the basis of administrative documents, audit logs, decision explanations, appeal documents, and review forms. The framework includes such techniques as natural language processing, compliance coding, reconstruction of accountability graphs, risk scoring, and simulation of interventions [8]. The goal is to operationalize the idea of governance quality and to see if certain interventions aimed at structure changes can increase the transparency completeness, responsibility traceability, and accountability performance.

2. Literature review

2.1. Transparency structures in algorithmic governance

The transparency of algorithmic governance is defined in terms of how open the data sources, the logic behind the algorithm, the decision-making standards, the procedures followed, the explanations, and the means of reviewing are to all the parties involved [2, 3]. Operationally, transparency implies that there are logs, explanations, and visible connections from the algorithmic output to the administrative actions [6]. The problem here is that transparency might be limited or performative. Agencies can release information on general goals and strategies while concealing information about the sources of the data, categories used, update procedures for the models, and decision-making criteria. Sometimes the explanations are created within the agency, but they are not available to the public, auditors, or appeal bodies [9, 10].

2.2. Accountability mechanisms in public-sector algorithmic systems

Accountability in public sector algorithmic systems involves the delegation, evaluation, remedy, and enforcement of responsibility [4]. Decisions made using algorithms in the public sector involve many individuals, among them those in charge of data, those involved in the development of models, vendors responsible for procurement, case workers, supervisors, and auditors. Proper accountability requires the identification of clear owners of decisions, defined points of review, defined escalations, appeals, remedies, and closure of audits [8]. Failure to have these leads to fragmented accountability. In algorithmic governance, fragmented accountability is especially detrimental since decisions could be regarded as neutral technical results by administrative entities.

2.3. Computational evaluation of governance performance

Evaluation of computational governance measures involves quantification of administrative documents. Documents including policies, audit trails, explanatory documents, appeal decisions, and case studies may be analyzed through text mining, classification, graphing, and risk assessment [10]. The advantage of using computational measures is that problems within governance tend to exhibit some kind of pattern through documents and processes. Examples include failure to include

DOI: 10.54254/2755-2721/2026.39078

disclaimer terms, broken responsibility chains, delays in corrections, unexplained automated decisions, and lack of resolution to appeal cases.

3. Experimental methods

3.1. Data construction, governance corpus building, and indicator encoding

he study constructs a governance evaluation corpus from algorithm-related administrative records, including policy documents, procurement specifications, data dictionaries, model cards, decision logs, explanation records, appeal decisions, audit reports, and stakeholder review forms. A unique identifier for each document is provided to allow association with logs, appeals, annotations, and reviews related to a single case. Input data in the form of raw PDFs, DOCX, JSON, CSV, and TXT files are preprocessed via Python 3.11 into a normalized format and then stored in a PostgreSQL 16 database and indexed in an Elasticsearch 8 instance to facilitate structured searches and full-text searches. The dataset is then annotated within Doccano based on a predefined codebook that includes completeness of disclosure, quality of explanation, continuity of tracing, clarity of responsibility, review accessibility, time to correction, audit resolution, and intervention compliance.

3.2. Transparency and accountability modeling tools

Textual transparency records are processed with spaCy, NLTK, and transformer-based classifiers. Sentence segmentation is first applied to divide policy documents, explanations, audit reports, and appeal decisions into analyzable units. Named entity recognition extracts agencies, vendors, departments, officers, data categories, system names, and review bodies. Policy clause detection identifies disclosure statements, decision-rationale expressions, contestability instructions, data-processing descriptions, and correction commitments. Explanation-type classification then separates generic explanations, model-centered explanations, case-specific explanations, and review-oriented explanations.

To measure transparency performance, this study defines a Transparency Completeness Index as shown in Equation (1):

$$TCI_c = \sum_{k=1}^{K} w_k D_{c,k} + \theta_1 E_c + \theta_2 L_c + \theta_3 R_c - \theta_4 M_c \tag{1}$$

where TCI_c denotes the transparency completeness score of governance case c, $D_{c,k}$ represents the presence or quality of disclosure component k, w_k is the component weight, E_c measures explanation depth, L_c measures log continuity, R_c measures review-instruction visibility, and M_c represents missing or contradictory disclosure elements. This formula allows the study to combine document completeness, explanation quality, and trace evidence into one standardized transparency score. Responsibility structures are reconstructed as directed graphs in Neo4j and NetworkX. Nodes represent agencies, case officers, model operators, vendors, auditors, review committees, appeal bodies, and affected citizens. Edges represent authorization, delegation, escalation, supervision, appeal submission, correction order, and audit verification. Graph metrics are then used to evaluate accountability structure. Path completion rate measures whether a decision path connects initial algorithmic output to final review or correction.

DOI: 10.54254/2755-2721/2026.39078

3.3. Evaluation metrics, validation protocols, and optimization algorithms

Transparency performance is measured through document completeness rate, explanation coverage score, feature disclosure density, decision trace continuity, and review instruction visibility. Accountability performance is measured through responsibility path depth, escalation success rate, unresolved appeal ratio, correction latency, audit closure rate, and intervention compliance score. Model evaluation uses accuracy, macro-F1, AUROC, precision-recall AUC, Cohen's kappa, and calibration curves. These indicators evaluate whether the computational pipeline can reliably classify governance quality and detect risk patterns across heterogeneous administrative cases.

To integrate transparency and accountability into one optimization target, the study constructs a composite Transparency-Accountability Risk Score as shown in Equation (2):

$$TARS_c = \alpha(1 - TCI_c) + \beta F_c + \gamma U_c + \delta A_c + \lambda Q_c - \mu C_c \tag{2}$$

where $TARS_c$ denotes the case-level governance risk score, TCI_c is the transparency completeness index, F_c measures responsibility-chain fragmentation, U_c denotes unresolved appeal ratio, A_c measures audit inconsistency, Q_c represents explanation inadequacy, and C_c captures completed correction execution. Higher scores indicate higher governance risk, while lower scores indicate stronger transparency and accountability performance.

Optimization is implemented with Optuna and PyGAD. The objective is to maximize transparency and accountability performance while minimizing procedural burden and avoiding unrealistic institutional changes. Candidate interventions include mandatory explanation templates, expanded log retention, explicit reviewer assignment, automated escalation triggers, appeal guidance publication, and audit checkpoint insertion. Validation is performed through train-test splitting at the case level, manual benchmark comparison, and sensitivity testing of indicator weights (see Figure 1).

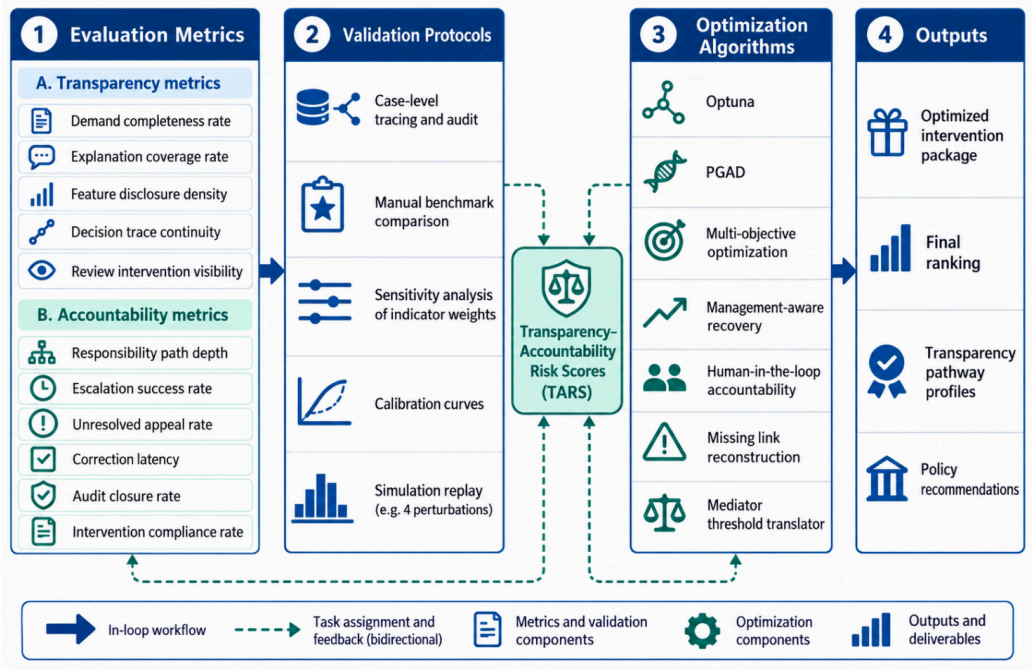


Figure 1. Evaluation metrics, validation protocols, and optimization algorithms

DOI: 10.54254/2755-2721/2026.39078

4. Experimental procedure

4.1. Data ingestion, preprocessing, and governance variable extraction

All governance records are collected into a unified repository, renamed by case identifier, and parsed through automated scripts. Metadata fields include agency name, model type, decision stage, explanation format, review level, appeal status, vendor participation, and correction outcome. Raw PDF, DOCX, JSON, CSV, and log files are normalized into UTF-8 text and tabular formats using pandas, pdfplumber, python-docx, and customized log parsers. Duplicate files are removed through hash comparison. Temporal fields are standardized into ISO format, and missing values are coded using fixed null-state categories. The cleaned corpus is processed through a variable extraction pipeline. Disclosure clauses, explanation segments, responsibility actors, decision events, oversight checkpoints, appeal triggers, and remedial actions are extracted from the corpus and linked back to case identifiers. The output is a structured governance table containing document-level, actor-level, event-level, and case-level variables.

4.2. Governance annotation, graph reconstruction, and risk scoring

Annotated governance units are labeled in Doccano under fixed categories including data source disclosure, model logic disclosure, decision rationale expression, human review presence, contestability instruction, responsibility owner, escalation node, and correction result. Labeled outputs are imported into Neo4j, where each case is converted into a governance graph arranged by execution order and institutional hierarchy. A case-level risk matrix is then computed in Python. Missing disclosure indicators, disconnected responsibility paths, absent review triggers, delayed correction events, unexplained automated decisions, and unresolved appeals are combined into the Transparency-Accountability Risk Score. Each case receives a risk score and a diagnostic profile showing which transparency or accountability component contributed most to governance weakness.

4.3. Intervention simulation, comparative testing, and output generation

The fixed intervention library is created, which includes mandatory decision explanation template creation, additional log storage, reviewer designation, automatic escalation process initiation, public access to appeals, and audit point inclusion. Interventions are modelled through updates to corresponding governance variables followed by re-computation of transparency, accountability, and risk measures by means of Python batch script. The comparison test of baseline and intervention settings is conducted using paired statistics. The outputs include governance scores tables, responsibility graphs, quality of explanations summary, intervention effect matrix, and prioritization ranking for governance improvement.

5. Experimental results

5.1. Transparency detection, explanation quality, and disclosure performance

The dataset is made up of 1,248 governance events and 18,736 document snippets. The completeness of baseline disclosures is 0.642 ± 0.071, that of explanations is 0.587 ± 0.083, and that of traces is 0.614 ± 0.076. With the implementation of the use of structured explanations and long-term logging, the completeness of disclosures increases to 0.813 ± 0.058***, that of explanations to

DOI: 10.54254/2755-2721/2026.39078

0.756 ± 0.069***, and that of traces to 0.792 ± 0.061***. Importantly, the greatest improvement is seen in review instructions transparency

5.2. Accountability graph reconstruction, responsibility path analysis, and governance risk distribution

There were 8,946 actor nodes and 21,384 responsibility edges from the accountability graph reconstruction (see Table 1). The control measure of responsibility path completeness was 0.603 ± 0.088, while the number of disconnected responsibility components was 31.7 ± 4.9 percent out of 100. Interventions using reviewer assignment and escalation triggers resulted in an increase in the completeness of responsibility paths to 0.781 ± 0.064*** and a decrease in unresolved appeals to 0.136 ± 0.037.

Table 1. Transparency and accountability performance before and after structural optimization

Indicator	Baseline	Optimized Condition	Change
Disclosure Completeness	0.642 ± 0.071	0.813 ± 0.058***	+0.171 ± 0.034
Explanation Coverage	0.587 ± 0.083	0.756 ± 0.069***	+0.169 ± 0.041
Decision Trace Continuity	0.614 ± 0.076	0.792 ± 0.061***	+0.178 ± 0.039
Review Instruction Visibility	0.531 ± 0.092	0.745 ± 0.073***	+0.214 ± 0.046
Responsibility Path Completion	0.603 ± 0.088	0.781 ± 0.064***	+0.178 ± 0.043
Audit Closure Rate	0.668 ± 0.079	0.804 ± 0.056**	+0.136 ± 0.037

5.3. Comparative effects of structural interventions on transparency and accountability optimization

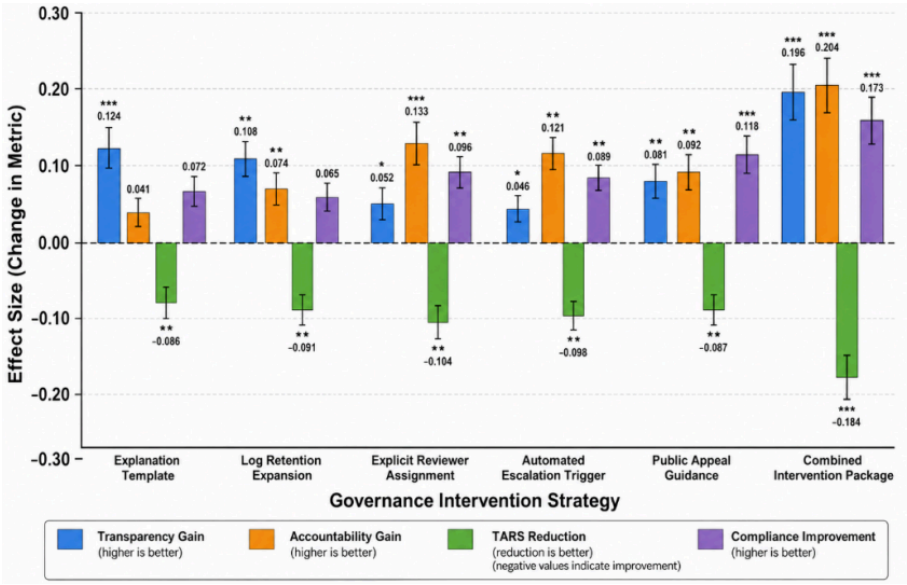


Figure 2. Comparative effects of governance intervention strategies

Simulation suggests that the application of structure optimization produces a more significant impact if it is used systematically and not selectively. In particular, explanation templates resulted in

DOI: 10.54254/2755-2721/2026.39078

an improvement in explanation deficiency by −0.142 ± 0.031***, while reviewer tasks contributed to improving the responsibility diffusion score by −0.126 ± 0.028**. Overall, the implementation of all interventions resulted in an improvement in TARS from 0.418 ± 0.066 to 0.234 ± 0.052***, i.e., a decrease of governance risks by −0.184 ± 0.039*** governance risk improvement.

6. Conclusion

This study presents an integrative computational methodological approach to optimize transparency and accountability structures for algorithmic governance. The approach integrates governance corpus formation, document retrieval, responsibility reconstruction with graphs, risk assessment, and intervention simulation into one computational workflow. By transforming policy texts, explanations, log files, appeal results, and reviews into quantitative metrics, the study provides a replicable model for assessing the effectiveness of algorithmic governance. The study's results reveal that transparency and accountability could be improved through strategic structural modifications. In particular, explanation frameworks, log extension, reviewer designation, appeal guidelines, and audit points have shown to produce positive effects, while the combination of all interventions resulted in the greatest decrease in risk.

References

[1] Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review, 81*(5), 825–836.
[2] Saldanha, D. M. F., Dias, C. N., & Guillaumon, S. (2022). Transparency and accountability in digital public services: Learning from the Brazilian cases. *Government Information Quarterly, 39*(2), 101680.
[3] Kingsman, N., Kazim, E., Chaudhry, A., Hilliard, A., Koshiyama, A., Polle, R., Pavey, G., & Mohammed, U. (2022). Public sector AI transparency standard: UK Government seeks to lead by example. *Discover Artificial Intelligence, 2*, Article 2.
[4] Bignami, F. (2022). Artificial intelligence accountability of public administration. *The American Journal of Comparative Law, 70*(Supplement_1), i312–i346.
[5] Schiff, D. S., Schiff, K. J., & Pierson, P. (2022). Assessing public value failure in government adoption of artificial intelligence. *Public Administration, 100*(3), 653–673.
[6] Oduro, S., Moss, E., & Metcalf, J. (2022). Obligations to assess: Recent trends in AI accountability regulations. *Patterns, 3*(11), 100608.
[7] Alon-Barkat, S., & Busuioc, M. (2024). Public administration meets artificial intelligence: Towards a meaningful behavioral research agenda on algorithmic decision-making in government. *Journal of Behavioral Public Administration, 7*, 1–19.
[8] Agbabiaka, O., Ojo, A., & Connolly, N. (2025). Requirements for trustworthy AI-enabled automated decision-making in the public sector: A systematic review. *Technological Forecasting and Social Change, 215*, 124076.
[9] Rizk, A., & Lindgren, I. (2025). Automated decision-making in public administration: Changing the decision space between public officials and citizens. *Government Information Quarterly, 42*(3), 102061.
[10] Hillo, J., Vento, I., & Erkkilä, T. (2025). Algorithmic governance: Experimental evidence on citizens' and public administrators' legitimacy perceptions of automated decision-making. *Public Administration*.