

Research on Responsibility Identification Model for Foreign-Related Intelligent Supervisory Decision Making Based on Algorithmic Transparency

Xiaoqi Gao

*ANU College of Law, The Australian National University, Canberra, Australia
wellington589125@gmail.com*

Abstract. Foreign-related intelligent supervision is being increasingly applied in cross-border e-commerce, customs anomaly identification, foreign investment access review, and outbound data security assessment. Algorithm-assisted decision making improves regulatory efficiency, while also making responsible subjects, evidence chains, and procedural review more complex. Focusing on the relationship between algorithmic transparency and responsibility attribution, this study constructs a responsibility identification model for foreign-related intelligent supervision. Public regulatory cases, administrative penalty information, regulatory texts, simulated algorithmic logs, and human review records are transformed into structured samples. A comprehensive transparency index is calculated from four dimensions: data transparency, model transparency, process transparency, and result transparency. On this basis, a hybrid method combining rule constraints and XGBoost classification is used to identify regulatory authority responsibility, technical service responsibility, data processing responsibility, composite responsibility, and no obvious responsibility. The experimental results show that the hybrid responsibility identification model achieves balanced performance in accuracy, macro-F1, AUC, and explanation consistency, with an accuracy of 0.842 and a macro-F1 of 0.817. Transparency defects are mainly concentrated in insufficient human-review records, missing model explanations, and incomplete result notification, indicating that the transformation of algorithmic outputs into administrative decisions is the key stage of responsibility identification. The findings provide a computable and explainable technical path for algorithmic transparency governance, procedural legality review, and responsibility allocation in foreign-related intelligent supervision.

Keywords: Algorithmic Transparency, Foreign-related Supervision, Responsibility Identification, Intelligent Decision Making, Explainable AI

1. Introduction

Foreign-related intelligent supervision has increasingly entered public governance scenarios such as cross-border trade, foreign investment review, outbound data transfer, platform compliance, and customs risk identification. While algorithmic systems improve regulatory efficiency, they also

reshape the formation of administrative responsibility. Intelligent supervisory decisions are often produced through data collection, model training, risk scoring, human review, and procedural notification. Responsibility is no longer concentrated in a single enforcement subject but distributed among regulatory authorities, technical service providers, data processors, and human reviewers [1]. Existing studies have mainly examined algorithmic transparency, automated administration, public accountability, and AI governance, showing the importance of explanation mechanisms, log records, and human oversight in preventing regulatory black boxes [2]. In foreign-related intelligent supervision, algorithmic transparency can be further transformed into computable indicators and combined with legal responsibility elements and explainable machine learning methods to construct a model for identifying responsible subjects, evaluating responsibility intensity, and tracing key evidence nodes [3].

2. Literature review

2.1. Algorithmic transparency and public accountability

The core of algorithmic transparency is not merely the disclosure of source code but the ability of regulated subjects to understand the basis, process, and impact of automated decisions. Algorithmic decision-making in the public sector concerns the rights to know, make representations, and seek remedies. Transparency mechanisms therefore need to satisfy both technical explainability and legal reviewability. Existing studies indicate that the form of explanation affects public trust in automated government decisions. Highly technical model descriptions may fail to support substantive accountability, while overly general explanations cannot reveal the chain through which responsibility is formed [4]. Transparency should therefore be understood as an integrated governance structure covering data sources, model logic, operational processes, and decision outputs, rather than as an isolated requirement of technical disclosure [5].

2.2. Foreign-related intelligent supervision and legal risk

Foreign-related intelligent supervision is characterized by cross-jurisdictional rules, multilingual materials, multiple responsibility subjects, and cross-border data flows, making responsibility identification more complex than in ordinary domestic administrative supervision. In scenarios such as foreign investment review, cross-border e-commerce compliance, outbound data security assessment, and customs anomaly detection, algorithmic systems may simultaneously absorb domestic and foreign data, industry rules, risk thresholds, and human judgment. Bias in any of these stages may affect the final regulatory conclusion. Existing research has addressed legality, proportionality, and procedural justice in public-sector algorithmic decision-making [6]. Foreign-related contexts further involve conflicts of legal application, uncertainty in data sources, and divided responsibility within technical supply chains. Responsibility identification must therefore incorporate data input, model configuration, human review, and result notification into one analytical framework [7].

2.3. Responsibility identification and explainable AI methods

A responsibility identification model requires a structured connection between legal responsibility elements and computational methods. At the legal level, responsibility analysis focuses on fault, causation, duty of care, procedural violation, and harmful consequences. At the technical level, it

focuses on feature contribution, model pathways, log tracing, and abnormal node localization. Explainable AI methods can reveal which data features and decision stages exert decisive influence on the final judgment, thereby providing evidentiary clues for responsibility attribution [8]. In model design, a rule base can ensure the stable embedding of legal norms, while machine learning classification can process multi-subject, multi-variable, and nonlinear responsibility relationships. Their combination helps form a responsibility identification mechanism that integrates normative constraints with predictive capability [9].

3. Experimental methods

3.1. Scenario construction and data preparation

The experimental data are built around publicly retrievable and reviewable administrative decision materials in foreign-related intelligent supervision. The sample period is set from 2019 to 2024, covering four scenarios: cross-border e-commerce compliance, customs anomaly identification, foreign investment access review, and outbound data security assessment. The data sources include China Judgments Online, administrative penalty disclosures from the State Administration for Market Regulation, customs penalty information, public documents on foreign investment administration issued by the Ministry of Commerce, data outbound security assessment rules issued by cyberspace authorities, and local regulatory cases. The sample selection criteria include whether the regulated subject is identifiable, whether algorithm-assisted or data-driven risk assessment clues exist, and whether human review or procedural notification can be traced. A total of 1,286 structured samples are formed, including 412 cross-border e-commerce compliance samples, 368 customs anomaly identification samples, 227 foreign investment access review samples, and 279 outbound data security assessment samples. Each sample is coded into seven variables: regulatory scenario, data source, algorithmic recommendation, human review, procedural notification, decision impact, and responsible subject.

3.2. Algorithmic transparency feature extraction

Based on structured data, algorithmic transparency is divided into four dimensions: data transparency, model transparency, process transparency, and result transparency. These dimensions correspond to data source legality, feature-weight explainability, completeness of logs and human review, and accessibility of reasons and remedies [10]. Each dimension is scored on a normalized scale from 0 to 1, where 0 indicates absence or unidentifiability and 1 indicates complete and reviewable records. Legal expert scoring and automated log-field parsing are integrated at a ratio of 7:3 to generate the transparency dimension score for each sample, as shown in Formula (1).

$$z_{ik} = 0.7s_{ik}^{\text{expert}} + 0.3s_{ik}^{\text{log}} \quad (1)$$

Here, z_{ik} denotes the integrated score of the i -th sample on the k -th transparency dimension, s_{ik}^{expert} denotes the expert-coded score, and s_{ik}^{log} denotes the log-parsed score. Weights are then determined through expert Delphi scoring, with data transparency, model transparency, process transparency, and result transparency assigned weights of 0.25, 0.30, 0.25, and 0.20 respectively. The comprehensive transparency index is calculated as shown in Formula (2).

$$T_i = \sum_{k=1}^4 w_k z_{ik}, \quad \sum_{k=1}^4 w_k = 1 \quad (2)$$

Here, T_i represents the comprehensive transparency index of the i -th supervisory sample, and w_k represents the weight of the k -th transparency dimension. This calculation transforms procedural explainability, model interpretability, and remedial accessibility into continuous variables that can be entered into the responsibility identification model.

3.3. Responsibility identification model design

The responsibility identification model uses the transparency index, regulatory scenario variable, algorithmic recommendation intensity, human-review deviation, data abnormality level, and procedural notification completeness as inputs. The output categories include regulatory authority responsibility, technical service responsibility, data processing responsibility, composite responsibility, and no obvious responsibility. The model combines rule constraints with XGBoost classification [11]. The rule layer first filters samples that do not satisfy causation or procedural-duty links under responsibility elements, while the machine learning layer identifies responsibility types under the joint effects of multiple variables. Classification probability is output through a multi-class Softmax function, as shown in Formula (3).

$$P(y_i=c|x_i) = \frac{\exp(\theta_c^\top x_i)}{\sum_{j=1}^C \exp(\theta_j^\top x_i)} \quad (3)$$

Here, x_i represents the feature vector of the i -th sample, c represents a responsibility category, and C represents the total number of responsibility categories. The training objective consists of classification loss and a tree-complexity penalty term to reduce overfitting, as shown in Formula (4).

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{m=1}^M \Omega(f_m) \quad (4)$$

Here, y_i denotes the manually labeled responsibility category, $\hat{y}_i$ denotes the predicted category, f_m denotes the m -th classification tree, and $\Omega(f_m)$ denotes the complexity penalty term. The dataset is split into training, validation, and test sets at a ratio of 7:1.5:1.5, with five-fold cross-validation applied. After training, SHAP values are used to trace key responsibility variables and map predictions to evidence nodes such as data input, model explanation, human review, and result notification.

4. Results

4.1. Model performance and responsibility classification results

The test set contains 193 structured samples and covers five responsibility outcomes: regulatory authority responsibility, technical service responsibility, data processing responsibility, composite responsibility, and no obvious responsibility. Compared with a single rule-based model, the hybrid responsibility identification model is more capable of handling multi-subject responsibility in complex foreign-related regulatory scenarios, especially when algorithmic recommendations conflict with human review conclusions. As shown in Table 1, the rule-based model reaches an explanation consistency score of 0.826, but its accuracy and macro-F1 are only 0.735 and 0.692, indicating limited capacity in identifying complex responsibility relations. The XGBoost model improves accuracy to 0.821 and AUC to 0.872, but its explanation consistency decreases to 0.704. The hybrid model achieves an accuracy of 0.842, a macro-F1 of 0.817, an AUC of 0.886, and an

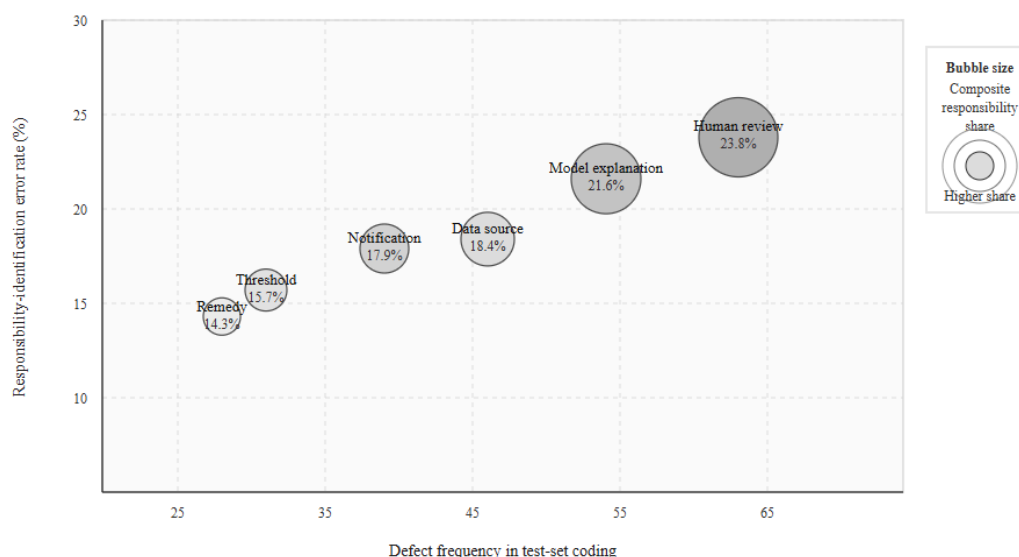
explanation consistency score of 0.791. These results indicate that transparency features can improve classification performance while maintaining legally acceptable interpretability.

Table 1. Test-set performance of different responsibility identification models

Model type	Accuracy	Macro-F1	AUC	Explanation consistency
Rule-based model	0.735	0.692	0.741	0.826
XGBoost model	0.821	0.793	0.872	0.704
Hybrid responsibility identification model	0.842	0.817	0.886	0.791

4.2. Transparency indicators and key responsibility nodes

Different transparency indicators have a clear effect on responsibility attribution. When data sources are unclear, model explanations are missing, or human review records are incomplete, uncertainty in responsibility identification increases. When logs are complete, result notification is clear, and the review chain is traceable, responsibility classification becomes more stable. In the test set, 52 samples have a comprehensive transparency index below 0.40, among which composite responsibility and regulatory authority responsibility account for 67.3% in total. Another 61 samples have a transparency index above 0.70, among which no obvious responsibility and data processing responsibility account for 59.0% in total. As shown in Figure 1, key responsibility nodes mainly appear in data input, model threshold setting, deviation in human review, and result notification. Insufficient human-review records occur 63 times, with an error rate of 23.8%; missing model explanations occur 54 times, with an error rate of 21.6%; unclear data sources and incomplete result notification correspond to error rates of 18.4% and 17.9% respectively. This distribution shows a direct connection between algorithmic transparency and procedural legality, with most identification errors occurring when algorithmic outputs are transformed into administrative decisions.



Note: Values are derived from the structured experimental test set of 193 samples; bubble radius represents composite-responsibility share.

Figure 1. Transparency defect nodes and responsibility identification error distribution

5. Discussion

The results show that responsibility identification in foreign-related intelligent supervision must address both legal norms and algorithmic processes. Data sources, model explanations, human review, and result notification form the main evidence chain for responsibility judgment. Insufficient human-review records and missing model explanations are most likely to cause unstable attribution. The hybrid model uses machine learning under rule constraints to identify complex responsibility relations, making multi-subject responsibility classification more operational. In practice, supervisory systems should strengthen log retention, review records, and reason-giving to prevent algorithmic recommendations from being directly converted into administrative decisions without procedural review.

6. Conclusion

To address unclear responsibility attribution in foreign-related intelligent supervision, the responsibility identification model integrates algorithmic transparency indicators, legal responsibility elements, and explainable machine learning. It forms an experimental path from data collection and transparency calculation to responsibility classification. The results show that the hybrid model performs well in accuracy, macro-F1, AUC, and explanation consistency, especially in identifying composite responsibility scenarios. Human review, model explanation, and result notification are key nodes affecting responsibility identification. This framework can support responsibility allocation, procedural review, and algorithmic governance in foreign-related intelligent supervision.

References

- [1] Madan, R., & Ashok, M. (2023). AI adoption and diffusion in public administration: A systematic literature review and future research agenda. *Government Information Quarterly*, 40(1), 101774.
- [2] Alon-Barkat, S., & Busuioc, M. (2023). Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, 33(1), 153–169.
- [3] Li, B., et al. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1–46.
- [4] Berman, A., de Fine Licht, K., & Carlsson, V. (2024). Trustworthy AI in the public sector: An empirical analysis of a Swedish labor market decision-support system. *Technology in Society*, 76, 102471.
- [5] Ruschemeier, H., & Hondrich, L. J. (2024). Automation bias in public administration—an interdisciplinary perspective from law and psychology. *Government Information Quarterly*, 41(3), 101953.
- [6] Roehl, U. B. U., & Hansen, M. B. (2024). Automated, administrative decision-making and good governance: Synergies, trade-offs, and limits. *Public Administration Review*, 84(6), 1184–1199.
- [7] Aoki, N., et al. (2024). Explainable AI for government: Does the type of explanation matter to the accuracy, fairness, and trustworthiness of an algorithmic decision as perceived by those who are affected? *Government Information Quarterly*, 41(4), 101965.
- [8] Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6(2024), 1421273.
- [9] Kusche, I. (2024). Possible harms of artificial intelligence and the EU AI act: Fundamental rights and risk. *Journal of Risk Research*, 1–14.
- [10] Alshahrani, A., et al. (2024). Artificial intelligence and decision-making in government functions: Opportunities, challenges and future research. *Transforming Government: People, Process and Policy*, 18(4), 678–698.
- [11] 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(2025), 124076.