

Research on the Cross-Domain Transmission Mechanism of Algorithm Ethical Risks in Enterprise AI Applications and the Construction of a Closed-Loop Governance System

Chujiao Pan¹², Bao Yuan^{13*}

¹ Renmin University of China, Beijing, 100872, China

² GCL Group Holdings Co., Limited, Suzhou, 215000, China

³ China Overseas Land & Investment Limited, Shenzhen, 518000, China

* 1094082233@qq.com

<https://doi.org/10.70695/IAAI202603A3>

Abstract

With the development of generative artificial intelligence (Gen AI), machine learning-based decision-making and intelligent agents in marketing, risk control, recruitment, auditing and customer service have started spreading algorithmic ethical risks across departments, systems and suppliers as chain-like transmissions rather than as errors in single-point models. Based on the literature from the past three years on AI risk management, audit accountability, transparency and governance frameworks, this paper focuses on risk formation and diffusion to construct a governance model of "source identification - transmission measurement - tiered handling - continuous auditing - accountability loop". The research proposes that to enhance corporate AI governance, ethical values need to be transformed into explainable, accountable and correctable institutional mechanisms, and a closed-loop control system can be established through algorithm impact assessment, model cards, data genealogy, red team testing, third-party auditing and board-level governance.

Keywords Enterprise AI; Algorithm Ethics; Risk Transmission; Responsible AI; AI Governance; Algorithm Auditing

1 Introduction

Enterprise AI Applications are changing from being auxiliary tools to becoming business infrastructure. Algorithms are employed for customer profiling, credit approval, talent selection, intelligent customer service and internal audit, etc., to adjust resource distribution, service availability and risk pricing accordingly. The NIST AI risk management framework promotes a combined governance, mapping, measurement and management system, and does not attribute all the value of AI to increased accuracy and cost reduction and efficiency improvement; it also requires addressing risks such as fairness, privacy, security, transparency and liability allocation [1].

Recently, most of the studies have shown that the governance system for artificial intelligence (AI) needs to be auditable, evidence-based and process-oriented. AI auditing should be conducted at the three levels simultaneously: technology, process and social impact [2]; algorithm accountability consists of subjects, standards, procedures and consequences [3]. If the company does not have an interface for authority and responsibility between business, data, suppliers, compliance, and the board of directors, algorithm errors may be attributed to "automatic generation by the system" and thus avoid accountability.

This paper studies the transmission paths and governance systems of ethical risks in enterprise AI algorithms. Risk transmission refers to the spread of data bias, model opacity, objective function imbalance, over-reliance on process automation and supply chain defects to other areas via business processes, organizational structures, user interaction and regulatory feedback. The three parts of the article are Introduction - Current Situation Analysis, Problem Statement, Solution Strategies; the specific questions asked are transmission carriers, measurement methods and closed-loop governance.

2 Current Status Analysis of the Research Topic

2.1 From Principled Ethics to Institutionalized Governance

Early AI ethics governance focused on fairness, transparency, privacy and human-centredness, and in the past three years, this has evolved into a set of principles. Research on transparency indicates that a major risk of black-box algorithms is that the individuals affected are unaware when the algorithm will process them, do not know the main factors that influence it, and have no way to raise objections or seek remedies [4]. When such risks occur in the company's processes, the model's output will also be spread as risks to customers' rights, employee evaluation, brand image, etc., and subject to regulatory penalties.

Based on a review of the auditing field, applying AI in audits can lead to biases at various stages, including compliance and financial judgment, training data, feature selection, model optimisation, result interpretation, and human-machine cooperation; thus, the auditee, investors and regulators are harmed [5]. Therefore, corporate AI risk does not occur in isolation within the code; rather, it is produced by algorithms, data, organisational processes and external systems.

Reform of the Regulatory System will be carried out. The EU's AI Act is built upon a risk-based system that includes a list of banned AI, high-risk AI, requirements for transparency and governance of general-purpose AI, etc [6]. As a result, the company needs to set up a definition of use, risk assessment, technical documentation, log recording, human supervision and cybersecurity controls, and no longer be managed independently by departments.

2.2 Enterprises Adopt Expansion and Risk Externalization

AI in enterprises has progressed at a much faster rate than that of the government-run institution. According to a 2024 survey by McKinsey, 72% of the surveyed companies have been using AI for some time in a certain area, and as of 2025, about three-quarters of them frequently use Gen AI [7]; data from the Stanford AI Index 2025 shows that in 2024, the number of organizations using AI had reached approximately 78% [8]. AI has moved from experiments by the technical team to daily work in the company.

Responsible AI governance needs to build institutional arrangements for role assignment, interdepartmental cooperation, full-life management and impact assessment [9]. According to the AI Index 2026, the proportion of companies that have adopted AI is 88 per cent, and the number of AI accidents has been rising. Policy Gaps, Knowledge Gaps and Budget Constraints still exist [10]. The actual problem has moved from "whether to use AI" to "whether the governance capacity can keep up with the pace of AI adoption".

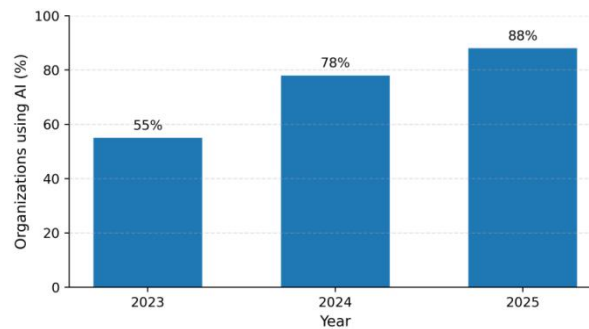


Fig. 1. Changes in enterprise AI adoption rate

Figure 1 Explanation: Based on the public data released by Stanford AI Index 2025 and 2026, it can be seen that by 2025, the percentage of organisations using AI will have risen to 88% from 55% in 2023. The trend is that AI has gone from being a one-off experiment to large-scale application, and therefore, the focus of governance should also shift from pre-launch model review to continuous control of cross-departmental processes, supplier interfaces and user interactions.

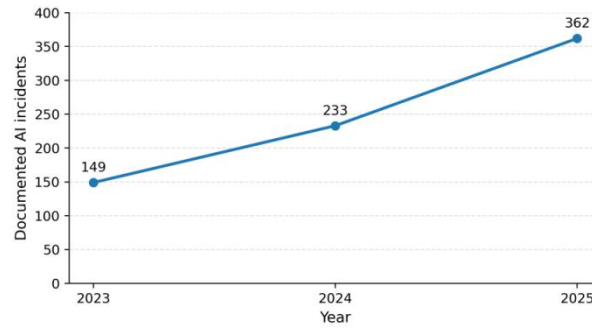


Fig. 2. Changes in the number of AI events

Figure 2 Explanation: The contents of this figure are derived from the AI Incident Database and the Stanford AI Index. In 2024, there were 233 AI incidents; by 2025, this figure will reach 362, and based on the growth rate, it was approximately 149 in 2023. An increase in the number of cases indicates that ethical risks are not random events but are getting more prominent as the scale of deployment, complexity and social dependence rise.

Table 1. Dimensions of ethical risk transmission in enterprise AI algorithms

Layer	Ethical risk	Transmission carrier	Enterprise Symptom	Evidence artifact
Data	Bias; privacy leakage	Dataset; feature pipeline	Unequal output; data complaint	Data lineage; consent log
Model	Opacity; hallucination	Model weights; prompt chain	Unexplained decision; false content	Model card; evaluation report
Process	Automation over-reliance	Workflow rule; API trigger	Rubber-stamp approval	Human review record
Organization	Accountability gap	RACI boundary; escalation path	No owner after failure	Governance charter
Ecosystem	Vendor and regulatory spillover	Third-party model; cross-border data	Contract dispute; compliance breach	Vendor audit file

Table 1 shows that the five layers of ethical risks are: data, model, process, organisation and ecosystem, and they are linked to transmission carriers and evidentiary materials. This division does not miss examining the entire process by only looking at the model, but rather requires the inclusion of data genealogies, model cards, manual review records and supplier audit documents in the same chain of evidence.

3 The Issue of Ethical Risk Transmission in Enterprise AI Algorithms

The first is that the origins of the problem are often unknown. Data bias, sample gaps, labeling errors, proxy variable discrimination, prompt word injection and model illusions may not be severe in the development stage, but they can be associated with differences in users, process rules and performance indicators in a business scenario. A recruitment model may be limited to traditional worksite cultures, and a credit risk control mechanism may also reduce the social opportunities for certain groups somewhat.

Secondly, there is chain coupling. Generally speaking, the components of enterprise AI systems include a data warehouse, feature platform, base model, fine-tuning model, rule engine, RPA process, third-party interfaces, and manual approval. At any stage, ethical deficiencies will be carried forward in API calls, automatic approvals, recommendation rankings, customer outreach and internal performance assessments. Therefore, the governance will also cover pre-deployment testing, drift monitoring and anomaly explanations, as well as appeals and accountability.

The third is a loss of institutional memory. Although AI projects are frequently led by the Digital or Business departments, the scope of the risks associated with them also includes legal and regulatory risks, compliance liabilities, personnel planning challenges, damage to brand image, data security issues, and others. If there is no unified language for risk and escalation at all levels, it is likely that the divisions

among technical, business and compliance teams will result in a buck-passing problem and a failure to control ethical risks.

To improve the identification efficiency, the ethical risk of a single AI application can be expressed as the product of probability, impact, exposure range and sensitivity coefficient:

$$R_i = P_i \times I_i \times E_i \times S_i \quad (1)$$

The left side of the formula is the total ethical risks of a single AI application, and the four items on the right side are the probability of occurrence, the severity of impact, the scope of exposure, and the sensitivity of the scenario; they are used to quickly classify high, medium and low risks.

Risk transmission from department i to department j can be classified into business dependence, data-sharing intensity and gaps in manual supervision.

$$TC_{ij} = B_{ij} \times D_{ij} \times H_{ij} \quad (2)$$

The left side of the formula is the risk transmission coefficient among departments, and the three terms on the right side are business dependence, intensity of data sharing, and human supervision gap, respectively; they are used to identify areas with high diffusion links.

Table 2. Key issues in the transmission of ethical risks in corporate AI

Problem	Trigger	Impact	Diagnostic indicator	Priority
Hidden source risk	Historical bias; weak consent	Discrimination; privacy harm	Bias test failure; complaint rate	High
Black-box decision	Complex model; weak documentation	Low contestability	Missing explanation log	High
Workflow amplification	Auto-approval; API chaining	Fast error scaling	Override ratio; exception volume	High
Vendor opacity	Third-party model; SaaS lock-in	Audit difficulty	Missing SOC/AI audit report	Medium
Accountability vacuum	Unclear RACI; no escalation	Delayed remediation	Unassigned incident ticket	High

As shown in Table 2, the problems listed correspond to the main breaks in the risk transmission path. Opaque Hidden Sources, Non-transparent decision-making processes, Strong amplification mechanisms, A lack of supplier disclosure, and Deficient accountability will each reduce one or more links in the capacity to identify, explain, monitor, audit and correct. Prioritisation indicates that the company should first tackle high-impact, high-spread and low-accountability issues.

4 Problem-solving strategies and governance system construction

4.1 Establish a Risk Classification and Algorithm Impact Assessment Mechanism

Group risks by use and impact, not only by model type. General-purpose models are relatively low-risk for use in internal fine-tuning, but their risks are significantly higher when applied to customer complaint handling, credit assessment, or employee performance evaluation. The scope of the algorithm's impact assessment should include the start of the project, acquisition, development, deployment, significant modifications and shutdown, focusing on legality, data sources, objectives, affected parties, human intervention, appeal paths and exit plans.

Link impact assessment to the budget, go-live approval and performance accountability. High-risk AI projects should submit model cards, data cards, red team tests, security assessments, deviation assessments and business rollback plans; medium- and low-risk projects can be simplified in process but must still keep their intended use, data, responsible persons and monitoring indicators, so that governance resources are allocated to high-impact situations.

4.2 Embedding Ethical Constraints into Model Development and Operation

Algorithmic Ethics Governance should not be based on manual review post-deployment. Fairness, privacy protection, robustness, and alert word attack testing should be implemented in the training, optimisation, acceptance and deployment stages instead. High-risk situations should be designated as "ethical metric thresholds for deployment", and even low- and medium-risk situations need to maintain sampling testing, logging and manual review mechanisms to prevent performance indicators from concealing rights violations.

4.3 Construct an organizational structure of "three-tiered defense + board oversight"

The first line of defence is under the business department and the model development team; it sets goals for the application, ensures data accuracy, conducts model testing, performs regular monitoring, etc. The second line of defence includes the legal and compliance office, information security centre, privacy department, risk management unit and so on, which issue guidelines, evaluate high-risk applications, arrange vendor risk assessments, and review significant modifications to establish a responsibility mechanism. The third line of defence is independent internal and external auditors who independently examine documents, data and log records of the audit results, address irregularities at previous links, and build a feedback system for accountability.

The Board of Directors or senior management can establish an AI governance committee to set the risk appetite for AI, allocate resources, list high-risk scenarios, devise external compliance strategies, and specify incident accountability rules. Only when governance responsibility is given to the high-level management will the ethical risks of algorithms not be confined to the technical team.

4.4 Establish a Quantifiable Assessment of Governance Maturity

To avoid the weak governance system being only in name, build corresponding, measurable and attributable indicators for the level of governance in the company.

$$GMI = \sum_{k=1}^5 w_k C_k \quad (3)$$

The formula is as follows: The left side is the AI governance maturity index, and the right side is a weighted sum of the five capabilities: strategic responsibility, data model control, transparency and explanation, monitoring and auditing, and incident handling.

Table 3. Enterprise AI closed-loop governance system

Governance layer	Core control	Responsible role	Evidence artifact	Review cycle
Strategy	AI risk appetite; use-case inventory	Board committee	AI governance charter	Annual
Design	Impact assessment; data lineage	Product owner; DPO	AIA report; data card	Project Gate
Development	Bias test; red teaming; security test	Model team; security team	Model card; test log	Before release
Operation	Drift monitoring; human override	Business owner	Monitoring dashboard	Monthly
Audit	Independent review; vendor audit	Internal audit	Audit report; remediation plan	Semiannual
Incident	Escalation; remedy; disclosure	Risk office; legal team	Incident file; lessons learned	Event-driven

Table 3 shows that the governance system in the table forms a closed-loop structure of strategy, design, development, operation, audit and incident handling. The goal is not to increase the amount of documents, but rather to ensure that at each stage there is a person or department responsible for the work, materials, and a review period; thus, the system is continuously improved through audit reports and corrective plans.

4.5 Form a Closed-Loop Operation of "Monitoring-Early Warning-Response-Learning"

A system of government cannot operate normally in practice. Build AI monitoring dashboards in the company to integrate model performance, deviation indicators, abnormal calls, unauthorised access, user complaints, human coverage, and supplier service status into a single monitoring platform. Set thresholds and automatic alarm rules for critical systems to automatically trigger review processes when a certain group exhibits an abnormal rejection rate, continuously declining confidence levels, an increase in attack hit rate, or insufficient human coverage.

Design tiers for incident response. General incidents should be handled by the business team within a certain period; serious incidents will require cooperation among the compliance, legal and information security departments; and very serious incidents must be reported to the AI Governance Committee and communicate externally. Conduct a root cause analysis for each incident, implement corresponding countermeasures such as repairing the model, modifying the process, correcting user behaviour, etc., and issue policy adjustments based on what has been learned from such incidents. Supplier governance should also be included in procurement, contracts and service level agreements to specify their scope of application, evaluation reports, deviation tests, log retention requirements and exit procedures.

5 Conclusion

Ethical risks of enterprise AI applications are not only technological risks but also socio-technical risks that stem from data sources, model structures, business objectives, organisational division of labour, user interactions and external institutional mechanisms. This study proposes a cross-domain transmission mechanism for enterprise AI ethical risks and constructs a closed-loop governance system based on risk classification, impact assessment, model constraints, three-tier defense, board oversight, audit evidence and event learning.

The first is that as the adoption rate of AI and the number of AI-related incidents increase, governance capabilities should also be strengthened in line with the spread of applications; the second is that the risk transmission mechanism includes data, models, processes, organisations and ecosystems, and governance targets should not be limited to algorithm code; finally, effective governance needs to turn ethical principles into quantifiable indicators, auditable evidence and accountable processes. In the future, real-world enterprise data will be used to verify the risk transmission matrix and governance maturity index, and the governance costs and benefits across different industries and regions will be compared.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).
2. Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society*, 2, 49.
3. Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. *AI & Society*, 39, 1871-1882.
4. Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273.
5. Murikah, W., Nthenge, J. K., & Musyoka, F. M. (2024). Bias and ethics of AI systems applied in auditing: A systematic review. *Scientific African*, 25, e02281.
6. European Parliament and Council. (2024). Regulation (EU) 2024/1689 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.
7. Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2024). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. *McKinsey & Company*.

8. Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., et al. (2025). Artificial Intelligence Index Report 2025. Stanford Institute for Human-Centered Artificial Intelligence. arXiv.
9. Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *Journal of Strategic Information Systems*, 34(2), 101885.
10. Sajadieh, S., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Santarlasci, L., et al. (2026). Artificial Intelligence Index Report 2026. Stanford Institute for Human-Centered Artificial Intelligence. arXiv.

Biographies

1. **Chuoqiao Pan** Postgraduate student at Renmin Business School, Renmin University of China. She graduated from Wentian College of Hohai University with a Bachelor of Engineering degree in Port, Waterway and Coastal Engineering. She currently serves as Training R&D Manager at GCL Group Holdings Co., Limited.
2. **Bao Yuan** Postgraduate student at Renmin Business School, Renmin University of China. He graduated from Shenyang Jianzhu University with a Bachelor of Management degree in Engineering Management. He currently works as an Investment Management Manager at China Overseas Land & Investment Limited.

企業 AI 應用中演算法倫理風險跨域傳導機制與閉環治理體系建構研究

潘楚翹¹²，袁寶¹³

¹中國人民大學商學院，北京，中國，100872

²協鑫（集團）控股有限公司，蘇州，中國，215000

³中國海外發展集團有限公司，深圳，中國，518000

摘要：隨著生成式人工智慧的發展，基於機器學習的決策系統與智慧體已廣泛進入行銷、風控、招募、審計和客戶服務等場景，演算法倫理風險也不再表現為單點模型錯誤，而是沿部門、系統與供應商形成鏈式傳導。基於近三年有關 AI 風險管理、審計問責、透明度與治理框架的研究，本文從風險生成與擴散過程出發，建構「源頭識別—傳導測度—分級處置—持續審計—問責閉環」治理模型。研究認為，提升企業 AI 治理能力，需要將倫理價值轉化為可解釋、可問責、可糾正的制度機制，並透過演算法影響評估、模型卡、數據血緣、紅隊測試、第三方審計和董事會治理等措施建立閉環控制體系。

關鍵詞：企業AI；算法伦理；风险传导；AI治理；算法审计

1. 潘楚翹，中國人民大學商學院在讀研究生，本科畢業於河海大學文天學院，取得港口航道與海岸工程工學學士學位，供職於協鑫（集團）控股有限公司，擔任培訓研發經理；
2. 袁寶，中國人民大學商學院在讀研究生，畢業於瀋陽建築大學，取得工程管理學士學位，任職於中國海外發展集團有限公司，擔任投資管理經理。