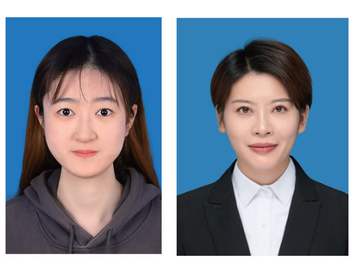
Governance of Fairness in Data Annotation:A Study of Dual Paths in Law and Ethics

Shen Rongrong, Zhao Yaqin

(School of Law, Shandong Jianzhu University, Jinan 250101)

Author Biography



Shen Rongrong, female, associate professor at the School of Law, Shandong Jianzhu University, Doctor of Economic Law, master's supervisor, former assistant professor in the Department of International Studies at Silla University, South Korea, currently serving as the academic deputy dean of the School of Law. Her main research areas are economic law and intellectual property law.

Zhao Yaqin, female, master's student at the School of Law, Shandong Jianzhu University.

Email:13916@sdjzu.edu.cn

Thesis structure framework

- Introduction of the problem
  - Research Background: Trusted Artificial Intelligence Consensus and National Security Requirements
  - The core role of data annotation: the “primary labor” of AI model training
  - Real-world Issue: Algorithmic Bias Cases Caused by Data Labeling
- An Examination of the Current Regulations on the Fairness of Data Annotation Domestically and Abroad
  - Comparison of Legislative Models in Major Countries and Regions
    - EU: Risk Classification of the Artificial Intelligence Act and Bias Mitigation Mechanisms
    - United States: Federal Voluntary Guidelines, State-Level Mandatory Regulations
    - China: Principle-Based Policy Guidance, Insufficient Legal Hierarchy and Enforceability
  - Research Status of Fairness Governance in Data Annotation
    - Domestic scholars suggest: standardization of labor, integration of ethics, and hierarchical and categorized management
    - Foreign scholars suggest: international collaboration, diverse samples, third-party audits
    - Research Trends and Limitations
  - Theoretical Blind Spots in Existing Research: Formalizing Value Embedding
- Mechanisms of Algorithmic Bias in Data Annotation and Legal Dilemmas
  - The Dual Pathways of Algorithmic Discrimination
    - Embedding of Subjective Cognitive Bias
    - Object Data Structure Exclusion
  - Specific manifestations of fairness imbalance
    - Lag in Governance Focus
    - Technical fixes have inherent limitations
    - The Deep Paradox of Algorithmic Discrimination
- Legal and Ethical Dual Governance Path for Ensuring Fairness in Data Labeling
  - Regulatory Embedding and Ethical Review
    - Planning Stage: Mandatory Fairness Guidelines Ethics Hearing
    - Implementation Phase: Trusted Annotation Certification and Ethical Training for Annotators
    - Before deploying applications: fairness testing and data usage boundary legislation
  - Theoretical Innovation in Dual Governance
    - Governance Object Downward Shift: From the Application End to the Data Production End
    - Integration of regulatory systems: ethics → soft law → hard law
    - Restructuring Main Responsibilities: A Full-Chain Accountability Mechanism
- Conclusion
  - Research Significance: Providing a Source Governance Path for Trustworthy Artificial Intelligence
  - Future Challenges: Technological Iteration and International Coordination

Introduction

As the foundational "meta-labor" for training artificial in-telligence models, data annotation processes exhibiting fairness deficits and bias implantation have become sig-nificant sources of algorithmic discrimination, directly impeding the implementation of trustworthy artificial in-telligence. This study systematically analyses the genera-tion mechanisms of algorithmic bias within data annota-tion, revealing governance dilemmas arising from dual pathways: cognitive embedding by agents and structural exclusion by data objects. By comparing legislative ap-proaches to data annotation fairness governance across the EU, US, and China, it identifies theoretical blind spots in current regulations concerning the formalization of value embedding. Building upon this, a dual governance framework of "rigid legal constraints coupled with flexi-ble ethical guidance" is proposed. This framework out-lines pathways for bias mitigation through dual dimen-sions of rule embedding and ethical review, offering a systematic solution to address the "inherent flaws" of al-gorithmic discrimination and achieve fairness in data an-notation. It further drives the paradigm shift in AI govern-ance from "algorithm explanation" towards "data tracea-bility".

Research Methods

Literature Research Method

Systematically collect, organize, and analyze existing literature

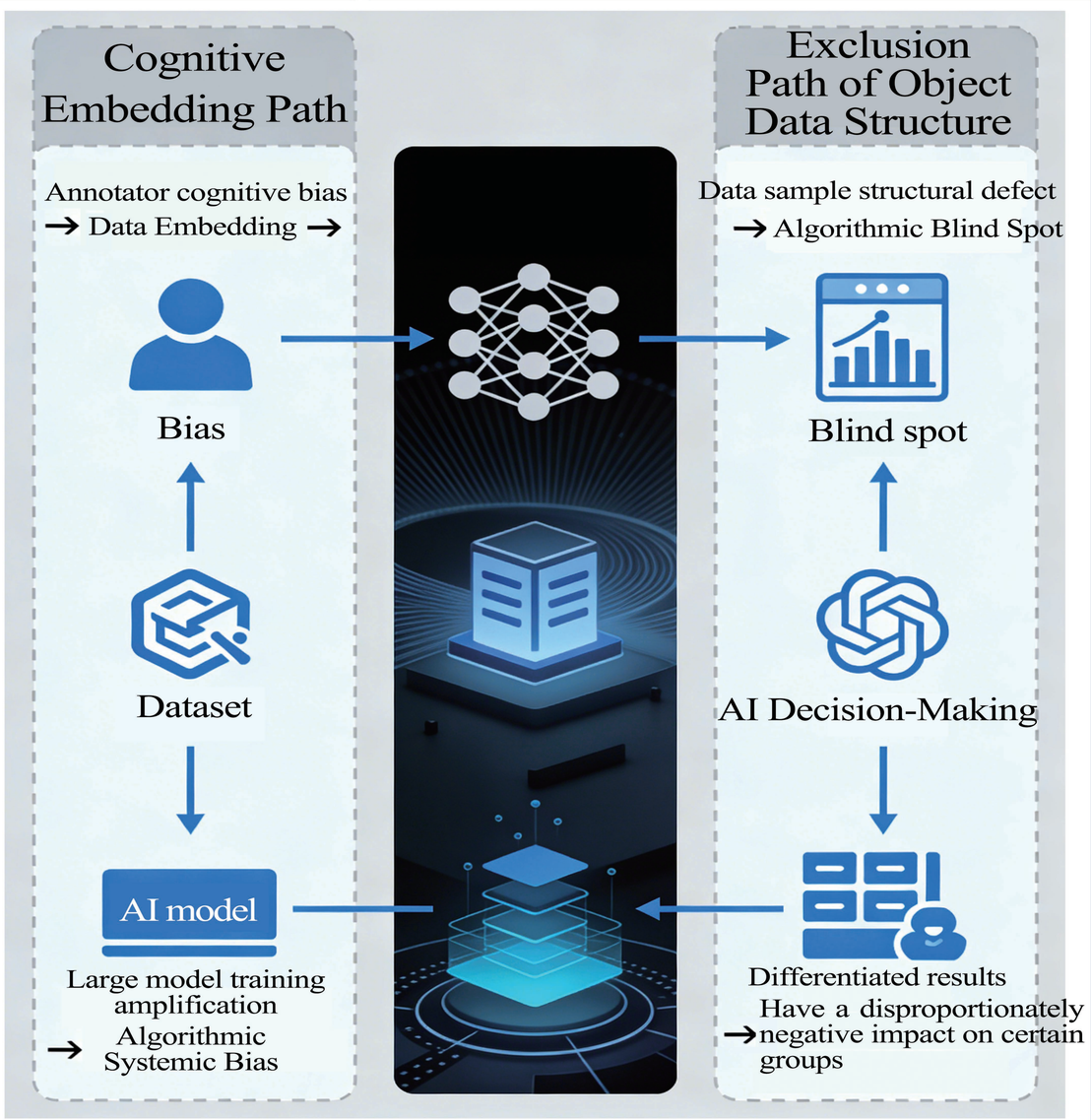
Comparative Research Method

By comparing the characteristics of different objects, reveal their similarities, differences, and patterns.

Framework Construction Method

Establish a systematic analytical model and theoretical framework

Mechanisms of Algorithmic Discrimination



Theoretical Innovations

Governance Focus Shifts Downstream

From algorithmic application to data production , achieving "source governance".

Integration of Normative Systems

Overcoming the dualistic opposition between law and ethics by translating ethical principles such as fairness and non-discrimination into operational legal rules, forming a normative hierarchy of "ethics → soft law → hard law".

Reconfiguration of Principal Responsibilities

Transcending the traditional "platform-user" dual re-sponsibility framework by incorporating annotators, annotation platforms, algorithm developers, and data users into a unified system of rights and obligations, thereby establishing a "full-chain accountability mech-anism".

Conclusion

As the bedrock of AI trust systems, fairness governance in data annotation concerns not only technological credibility but also social justice and legal order. By an-alyzing the mechanisms generating algorithmic bias and the legal dilemmas in data annotation, this paper pro-poses a governance framework centered on "legal-ethical" synergy, shifting the focus from back-end algo-rithmic explanation to front-end data traceability. This framework institutionalizes fairness values within data annotation through rule embedding and ethical review pathways, providing foundational governance for trust-worthy AI. Looking ahead, advancements in generative AI and multimodal large models will pose dual chal-lenges of technological iteration and international coor-dination. Continuous refinement of institutional design is essential to ensure AI evolves fairly, reliably, and be-nevolently within the rule of law.