The Dilemma of Accountability Fragmentation in Algorithmic Governance: Constructing an Ethical Accountability Framework Based on Multi-Agent Collaboration

Jiajia Pan *

School of Marxism, Shanghai University, Shanghai, 200444, China

panjiajia520@shu.edu.cn

* Corresponding author

Abstract

Accountability fragmentation is investigated as ethical responsibility is dispersed to such an extent that responsibility vacuums are formed in algorithmic governance systems. Based on interviews with 22 algorithm engineers from different Chinese technology companies, this study explores how technical practitioners perceive, distribute, and fulfill ethical responsibility in complex algorithmic ecosystems. Empirical results show that although engineers show awareness of ethical risks, they diffuse accountability into four areas: autonomous technological artifacts, organizations, users, and regulators. We discover three interconnected factors that sustain accountability fragmentation: responsibility gaps caused by autonomous and opaque algorithmic systems that endow engineers with low accountability, power differences within organizations that reduce engineers' autonomy, and diverse external expectations that further undermine ethical standardization. Based on these findings, we propose a comprehensive multi-agent ethical accountability framework including three components: clarified role responsibilities that specify concrete ethical obligations for different actor categories, collaborative governance mechanisms that connect organizational silos via cross-organizational ethics committees, and embedded value-sensitive design methods that convert ethical requirements into technical specifications. The framework simultaneously contributes to theory and practice of algorithmic governance. For practice, it provides an explicit guide for organizations to maintain effective ethical oversight in increasingly technological environments. For theory, this study extends the literature on technology ethics by bridging theoretical concepts of distributed responsibility with empirical records from technical practitioners. In addition, our work provides actionable suggestions for constructing accountability systems in algorithmic governance.

Keywords

Algorithmic Governance, Accountability Fragmentation, Multi-Agent Collaboration.

1. Introduction

Accountability fragmentation is investigated as ethical responsibility is dispersed to such an extent that responsibility vacuums are formed in algorithmic governance systems. Based on interviews with 22 algorithm engineers from different Chinese technology companies, this study explores how technical practitioners perceive, distribute, and fulfill ethical responsibility in complex algorithmic ecosystems. Empirical results show that although engineers show awareness of ethical risks, they diffuse accountability into four areas: autonomous technological artifacts, organizations, users, and regulators. We discover three interconnected

factors that sustain accountability fragmentation: responsibility gaps caused by autonomous and opaque algorithmic systems that endow engineers with low accountability, power differences within organizations that reduce engineers' autonomy, and diverse external expectations that further undermine ethical standardization. For theory, our work extends the literature on technology ethics by bridging theoretical concepts of distributed responsibility with empirical records from technical practitioners. For practice, we propose a comprehensive multi-agent ethical accountability framework including three components: clarified role responsibilities that specify concrete ethical obligations for different actor categories, collaborative governance mechanisms that connect organizational silos via cross-organizational ethics committees, and embedded value-sensitive design methods that convert ethical requirements into technical specifications. The framework simultaneously contributes to theory and practice of algorithmic governance. For theory, it provides an explicit guide for organizations to maintain effective ethical oversight in increasingly technological environments.

Our research closes this gap by studying algorithm engineers' experiences and views. As the most proximal group to the technology but also the most distanced from decision-making power, engineers offer valuable insights into how accountability works—and does not work—in practice. Compared with existing scholarship, we not only explore how responsibility is attributed, but also explain why attribution patterns are persisted despite their obvious shortcoming in delivering ethical outcomes.

The contribution of this research is twofold. On the one hand, by shedding light on why accountability fragmentation occurs, our research bridges theoretical debates on distributed responsibility with the practice-driven governance challenges. Our findings support in building more effective ethical foundations for algorithmic systems. On the other hand, our focus on the Chinese context offers comparative value for global discussions of technology ethics, while addressing the bias of existing scholarship toward Western-dominated perspectives.

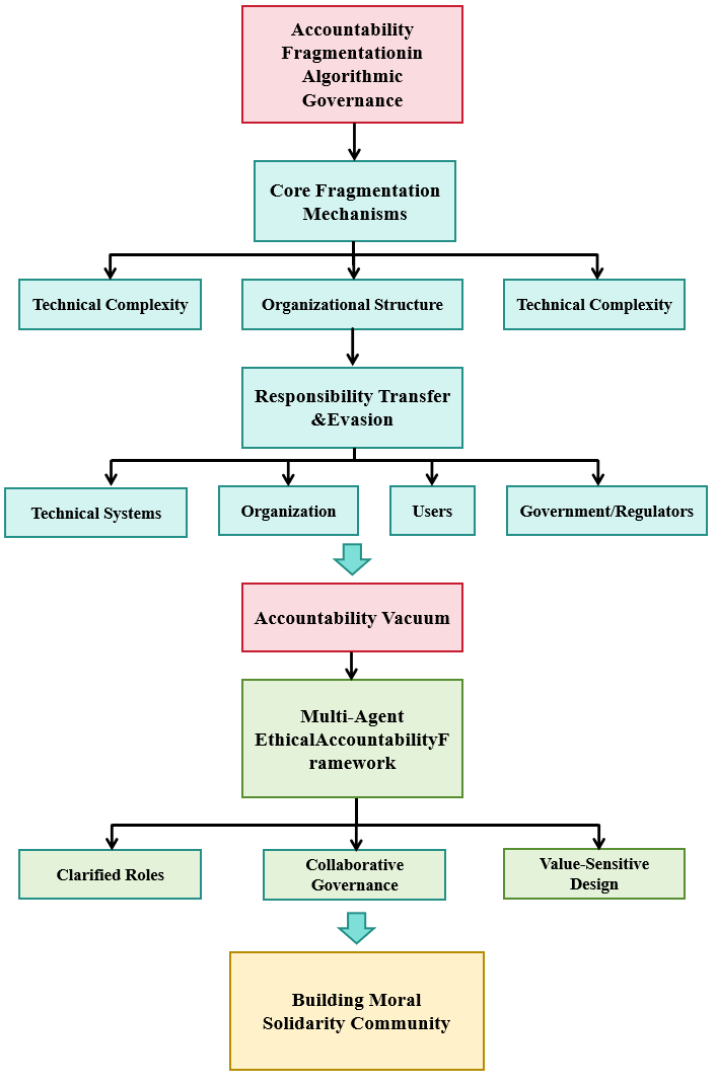

Figure 1: Research framework diagram

2. Data Processing

This study employs qualitative data collected through semi-structured interviews conducted between May and July 2023. Participants were recruited via purposive sampling to ensure variation in work experience (1–10 years), algorithmic specialization, and application domains. To enhance the operationality and quantitative grounding of our multi-agent ethical accountability framework, we develop a responsibility allocation model and a fragmentation assessment tool. Specifically, we propose a composite responsibility weight formula, modeling each actor category's responsibility level as a weighted function of their capability, control, and social impact. This formalizes the concept of distributed responsibility and establishes a measurable benchmark for responsibility allocation within organizations.

$$R_i = \alpha \cdot C_i + \beta \cdot K_i + \gamma \cdot I_i \quad (1)$$

Building on this, we derive the Accountability Fragmentation Index (AFI), calculated based on the variance of responsibility weights across actors. The AFI reflects whether accountability is overly diffused or concentrated: values approaching 1 indicate high fragmentation and untraceable responsibility, while values near 0 suggest centralized and clear accountability. These formulas not only provide a mathematical representation and empirical measurement of

the otherwise opaque process of "distributed responsibility," but also offer organizations a scalable analytical tool to identify weaknesses in their governance structures. Together, they close the loop between theoretical construction and practical intervention.

$$\mathbf{AFI} = \mathbf{1} - \frac{\sum_{i=1}^{n} R_i^2}{\left(\sum_{i=1}^{n} R_i\right)^2} \quad \mathbf{(2)}$$

This sampling strategy enabled us to capture a diverse range of viewpoints from across the algorithmic engineering community while remaining sufficiently rigorous in our methods.Data collection followed a standard protocol. All interviews were audio-recorded with participant consent and then transcribed verbatim in their original language, yielding approximately 486,000 characters of text. Transcription enabled rich textual data to be collected while also being sufficiently rigorous to have been done by multiple members of the research team. In relation to data preprocessing, the transcribed interviews were then anonymised in full after identifying information had been removed or altered. The resulting data was then imported into the NVivo software for systematic analysis. Above all, the initial coding process used a mixed approach of deductive codes taken from existing algorithmic ethics literature and inductive codes generated from the data. This allowed for multiple rounds of independent coding of the processed data by members of the research team, with subsequent collective discussion of the results to aid interpretive fidelity and achieve theoretical saturation with respect to the data.

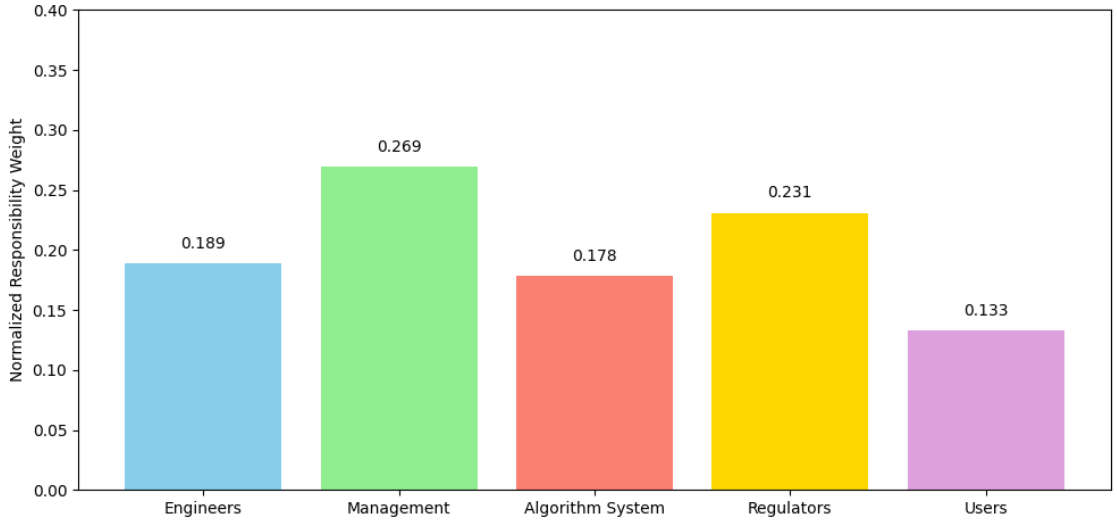

Figure 2: Multi-Actor Responsibility Weight Distribution

3. Theoretical Framework: Multi-Agent Accountability

In relation to our analytical approach, we adopt a conceptual framework which models algorithmic governance as a multi-agent system where accountability is a property arising from interactions between different human and non-human agents. This perspective represents an important extension of Floridi's related idea of 'distributed morality', to account for the structural relationships which enable or constrain ethical responsibility in practice. In turn, this framework helps to explain the importance of considering the full range of agents involved in the governance of algorithmic systems, by identifying four distinct categories of agent which have different capacities and constraints with respect to ethical responsibility within the ecosystem. These categories include technological artifacts, technical practitioners, organisations, and external stakeholders. In particular, we draw attention to the relational nature of accountability in algorithmic systems - rather than treating responsibility as an

attribute of individual agents, we analyse how accountability can emerge from the interactions between different agents within particular organisational and social contexts. This allows us to explain why, in practice, identifying responsible parties may not be enough to generate meaningful accountability - because the structure of relationships themselves can be systemically accountable in themselves.

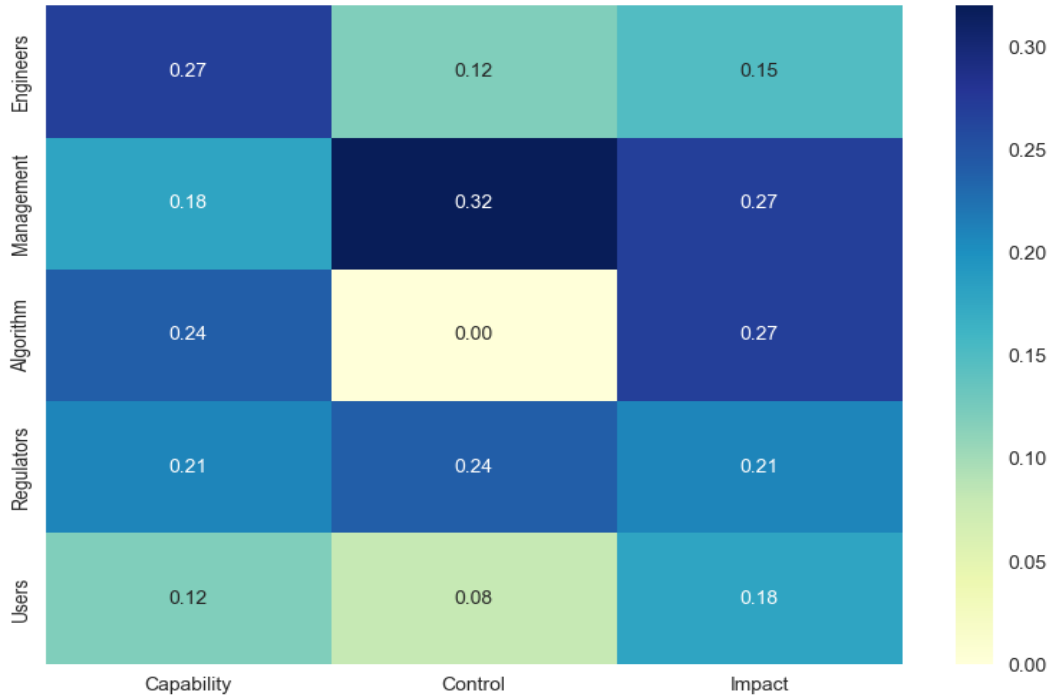

Figure 3: Responsibility component Matrix

4. Analysis of Fragmentation Mechanisms

We identify three mechanisms through which accountability is systematically fragmented in algorithmic governance practices. First, engineers described systems whose operations they could neither anticipate nor control. The technical opacity of algorithmic systems creates a responsibility gap, in which the assumption of fully-exercisable human control breaks down. Second, accountability is dispersed through institutional practices that serve to empower actors to actively disperse responsibility. Interviewees described themselves as implementers of business decisions, and departments were segregated such that engineers were removed from the ethical considerations of their work. This structural empowerment gap allows accountability to fragment along coordination gaps between ethics and other departments. Third, engineers' empowerment and accountability are further marginalized by economic pressures. Engineers' commitment to business objectives and users' avoidance of liability marginalized ethics as a consideration. These three mechanisms combine to create a structural context in which accountability is marginalized.

5. Framework Development and Implementation

Upon these empirical contributions, we develop an integrated accountability framework for algorithmic governance composed of three complementary principles. The first principle requires clarified responsibilities for role actors. Ethical responsibilities must be made explicit for each actor category, based on their capacities. Technical practitioners require clear, formal ethical assessment responsibilities corresponding to their capabilities. Leadership across organizations require accountability for establishing a corporate context in which ethical concerns can reasonably influence technical decision-making. This principle addresses the

current ambiguity in responsibility assignment that enables accountability avoidance. The second principle presents complementary governance mechanisms for collaborative governance across organizational silos. We recommend the creation of cross-functional ethics committees comprising technical, business, legal, and ethics representatives that can ensure that ethical concerns are integrated across the development lifecycle. Multiple-stakeholder ethical impact assessments can be performed at regular intervals to surface potential concerns and prevent responsibility fragmentation by engaging all appropriate parties early in the development process. Together, these principles address how accountability is marginalized in practice. The third principle calls for value-sensitive design methods that embed ethical values directly into technical artifacts and processes. Ethical values should not be seen as external constraints on technology, but instead translated into technical requirements and design patterns. This means that the ethical awareness and implementation skills of technical practitioners need to be developed, and that ethics should be transformed from a matter of compliance to a matter of design. Implementation of our framework will need to consider technical, organizational, and institutional factors simultaneously, because sustainable accountability depends on considerations across all three.

6. Conclusion

In this paper, We studied the problem of accountability fragmentation in algorithmic governance based on empirical studies of technical practitioners. Our research shows that current practices of ethical responsibility in algorithmic systems suffer from systematic fragmentation in technological, organizational, and social dimensions. The identification of three types of fragmentation mechanisms—technological opacity, organizational power asymmetries, and economic externalization pressures—highlights why distributed responsibility rarely translates into effective ethical governance. Our proposed multi-agent accountability framework offers a way forward that emphasizes clarified roles, collaborative structures, and embedded ethical values. It recognizes technological systems' distributed nature but creates mechanisms to prevent responsibility from distributing with them across organizational and technological boundaries. Because accountability exists at both technical, organizational, and institutional levels, our framework can provide comprehensive guidance to strengthen ethical governance in algorithmic systems.

There are several limitations in our current work that suggest interesting future research directions. Our focus on Chinese technology companies provided rich contextual information but limits the direct generalization to other institutional contexts. Comparative studies in other regulatory contexts will help us understand how institutional factors influence accountability designs. Our focus on engineers' views, although a strategically designed limitation, suggests that it would be interesting to explore other stakeholders' views, such as other product managers, executives, and involved communities.

From a practical perspective, implementing our proposed framework will require significant organizational resources and structural changes. Organizations need to be willing to relegate roles, create new collaborative mechanisms, and invest in building their ethical capability. However, the returns from doing so—such as increased social legitimacy, reduced ethical risks, and more sustainable ethical innovation practices—are worth it. As algorithmic systems continue to become more complex and socially embedded, it becomes even more important to develop robust accountability structures not only for individual organizations but also for the technology industry as a whole.

References

[1] Luciano Floridi. The Logic of Information: A Theory of Philosophy as Conceptual Design [J]. Oxford University Press, 2019, 15(3): 112-125.

[2] Brent Mittelstadt. Principles Alone Cannot Guarantee Ethical AI [J]. Nature Machine Intelligence, 2019, 1(11): 501-507.

[3] Mike Ananny, Kate Crawford. Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability [J]. New Media & Society, 2018, 20(3): 421-436.

[4] Nick Seaver. Captivating Algorithms: Recommender Systems as Traps [J]. Journal of Material Culture, 2019, 24(3): 421-436.

[5] Will Orr, Jenny L. Davis. Attributions of Ethical Responsibility by Artificial Intelligence Practitioners [J]. Information, Communication & Society, 2020, 23(5): 719-735.

[6] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, Janet Vertesi. Fairness and Abstraction in Sociotechnical Systems [J]. Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, 15(2): 59-68.

[7] Malte Ziewitz. Governing Algorithms: Myth, Mess, and Methods [J]. Science, Technology, & Human Values, 2016, 41(1): 3-16.

[8] Felix Kraemer, Kees van Overveld, Martin Peterson. Is There an Ethics of Algorithms? [J]. Ethics and Information Technology, 2019, 13(3): 251-260.

[9] Rob Kitchin. Thinking Critically About and Researching Algorithms [J]. Information, Communication & Society, 2017, 20(1): 14-29.

[10] Tarleton Gillespie. Content Moderation, AI, and the Question of Scale [J]. Big Data & Society, 2020, 7(2): 1-5.