DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Metadata-Conditioned Moral Grammars: How AI Companies Define Responsibility, Moral Language, Value Hierarchies, and Governance in AI Policy Documents

Seul Lee [1]

[1] Florida State University

Abstract

This study examines how leading AI companies, including Anthropic, Google, Meta, and OpenAI, construct ethical commitments in their governance documents, and how these constructions are systematically shaped by metadata conditions such as document type, organizational structure, and temporal context. While prior work has increasingly examined AI ethics documents as governance instruments shaped by institutional interests, sectoral differences, and political motivations, these studies still primarily approach AI ethics documents at the level of themes, principles, stakeholder representation, and policy discourse. This research extends that literature by arguing that AI ethics documents function as metadata-conditioned moral grammars: dynamic linguistic architectures that distribute responsibility, prioritize values, structure anticipatory governance, and guide decision-making under conditions of uncertainty. It investigates how AI governance documents by major technology corporations function as discursive infrastructures for the construction of ethical meaning. It analyzes how ethical frames including safety, innovation, autonomy, responsibility, and societal benefit are prioritized, sequenced, contextualized, and relationally embedded within these documents. Rather than understanding such principles as static normative commitments, the study examines how their linguistic organization operates as a metadata-conditioned moral grammar that distributes responsibility, structures anticipatory governance, and legitimizes institutional authority in AI governance. It examines how responsibility is both explicitly and implicitly attributed across various agents, including the organization, the AI system, users, developers, and abstract institutional entities, through patterns of grammatical agency and semantic positioning, while also identifying the linguistic mechanisms, such as passive constructions, abstract nominalizations, hedging expressions, and conditional qualifiers that contribute to the diffusion or attenuation of responsibility.

1. Extended abstract

This study examines how leading Artificial Intelligence (AI) companies, including Anthropic, Google, Meta, and OpenAI, construct ethical commitments in their governance documents, and how these constructions are systematically shaped by metadata conditions such as document type, organizational structure, and temporal context. Existing scholarship has moved beyond viewing AI ethics documents as merely collections of abstract principles, emphasizing instead their governance functions, institutional motivations, and political implications (Schiff et al., 2020 [1] Schiff et al., 2021 [2] Schiff et al., 2022 [3]). These studies demonstrate that AI ethics documents are produced by governments, corporations, NGOs, and transnational organizations as mechanisms for establishing legitimacy, influencing regulatory agendas, and articulating institutional responsibilities in the governance of AI. They highlight how AI ethics documents often function as governance artifacts that mediate tensions between economic competitiveness, technological innovation, public accountability, and social responsibility. At the same time, scholars have raised concerns that many of these frameworks remain ab-

Project Report

Available Aug 01, 2026

Correspondence to
Seul Lee
seul.lee@fsu.edu

stract, weakly enforceable, and susceptible to "ethics washing," where ethical discourse substitutes for substantive institutional change (Gijs, 2022 [4]). However, despite this growing attention to governance, motivation, and institutional power, prior work still primarily approaches AI ethics documents at the level of themes, principles, stakeholder representation, and policy discourse. This research extends that literature by arguing that AI ethics documents function as metadata-conditioned moral grammars: dynamic linguistic architectures that distribute responsibility, prioritize values, structure anticipatory governance, and guide decision-making under conditions of uncertainty.

In particular, this study investigates how AI governance documents produced by major technology corporations such as Google, Meta, Microsoft, OpenAI, and Anthropic function as discursive infrastructures for the construction of ethical meaning. It analyzes how ethical frames including safety, innovation, autonomy, responsibility, and societal benefit are differentially prioritized, sequenced, contextualized, and relationally embedded within these documents. Rather than understanding such principles as static normative commitments, the study examines how their linguistic organization operates as a metadata-conditioned moral grammar that distributes responsibility, structures anticipatory governance, and legitimizes institutional authority in AI governance. It examines how responsibility is both explicitly and implicitly attributed across various agents, including the organization, the AI system, users, developers, and abstract institutional entities, through patterns of grammatical agency and semantic positioning, while also identifying the linguistic mechanisms, such as passive constructions, abstract nominalizations, hedging expressions, and conditional qualifiers that contribute to the diffusion or attenuation of responsibility.

In addition, the study aims to analyze whether and how these documents encode explicit or implicit hierarchies among competing values, and how such hierarchies are articulated through discourse markers and structural cues, as well as whether these hierarchies can be operationalized into consistent and context-sensitive decision rules in situations involving ambiguity or value conflict, thereby assessing the extent to which ethical principles are actionable versus underdetermined. Finally, the study investigates how patterns of ethical framing, responsibility attribution, and value prioritization vary across document-level and organizational metadata, including document type, organizational structure, business model, and temporal context. Through comparative discourse analysis, it examines how these contextual variables correspond to recurring differences in the construction of ethical meaning and governance priorities within AI governance documents. Based on these patterns, the study conceptualizes AI governance documents as metadata-conditioned moral grammars—structured linguistic systems in which ethical language is systematically organized in relation to institutional context, governance objectives, and organizational incentives.

To operationalize the research questions, this study develops a multi-layered measurement framework combining sentence-level annotation, computational linguistic features, and document-level metadata variables. The analysis begins by segmenting

corporate AI governance documents into fine-grained linguistic units, such as sentences, clauses, modal constructions, and statements of obligation or responsibility. Using a structured coding scheme, the study identifies ethical frames, actor-responsibility relations, value prioritization, temporal orientation, and governance logics within each unit. Ethical framing can be operationalized as the distribution and co-occurrence of ethical categories at the sentence level. Responsibility attribution can be measured through grammatical agency and semantic role assignment, identifying who is positioned as the responsible actor. Each sentence can be coded using a multi-label classification scheme capturing dominant ethical themes. Large Language Models (LLMs) are employed to support systematic annotation and pattern recognition across a large corpus, while human-led discourse analysis is used to interpret how these recurring linguistic patterns construct institutional legitimacy, distribute responsibility, and organize moral authority. This mixed approach makes it possible to examine both the measurable distribution of ethical language and the broader discursive formations through which technology companies define responsible AI governance. By combining LLM-assisted fine-grained textual analysis with discourse analysis, this study contributes a new methodological and theoretical framework for examining how AI governance documents construct ethical meaning across institutional contexts. Rather than treating ethics principles as universally stable commitments, the study demonstrates how responsibility, legitimacy, and governance priorities are differentially organized through linguistic and discursive structures, with implications for AI accountability, regulation, and alignment practices.

References

[1] D. Schiff, J. Biddle, J. Borenstein, and K. Laas, *What's next for AI ethics, policy, and governance? A global overview*. in Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20). Association for Computing Machinery, 2020, pp. 153–158. doi: 10.1145/3375627.3375804.

[2] D. Schiff, J. Biddle, J. Borenstein, and K. Laas, "AI ethics in the public, private, and NGO sectors: A review of a global document collection," *IEEE Transactions on Technology and Society*, vol. 2, no. 1, pp. 31–42, 2021.

[3] D. Schiff, K. Laas, J. Biddle, and J. Borenstein, "Global AI ethics documents: What they reveal about motivations, practices, and policies," in *Codes of Ethics and Ethical Guidelines*, vol. 23, K. Laas, M. Davis, and E. Hildt, Eds., Springer, pp. 121–143. doi: 10.1007/978-3-030-86201-5_7.

[4] G. Van Maanen, "AI ethics, ethics washing, and the need to politicize data ethics," *Digital Society*, vol. 1, no. 9, 2022.