

Relational accountability in AI governance: A Confucian framework

Jooya Wong¹, Shumin Ren¹

¹ *UCSI University, Malaysia*

Corresponding author: Jooya Wong

Email: jooyawong@gmail.com

Abstract

Artificial intelligence increasingly operates through organisational and technical relationships that distribute control across model providers, platform providers, deploying institutions, professionals, users, and regulators. Existing accountability research clarifies answerability, audit, documentation, oversight, and institutional duties, while relational approaches in AI ethics show how roles and dependencies shape moral responsibility. A remaining problem concerns the translation of relational ethical concepts into a structured form of institutional accountability. This exploratory theory and model development study addresses that problem through the Confucian Relational Responsibility Framework (CRR). The paper first synthesises four patterns in relevant literature: the relational turn in AI and robot ethics, the translation of Confucian concepts into AI ethics, semantic compression during governance translation, and the continuing institutional accountability gap. It then uses a bounded conceptual corpus, explicit concept selection criteria, a four-step governance translation protocol, and framework construction to develop CRR. Five selected concepts are translated into responsibility categories: role into role responsibility, relationship into relational responsibility, ritual propriety into boundary-setting responsibility, trustworthiness into trust responsibility, and coordination into coordinated accountability. These categories are organised through five analytical dimensions: actors, relationships, scenarios, responsibilities, and mechanisms. A worked example involving generative AI search in a public knowledge institution illustrates how the framework traces duties across providers, institutions, professionals, and users and connects them to disclosure, human review, documentation, monitoring, audit, complaint procedures, and institutional review. CRR is presented as an exploratory analytical layer between relational ethics and institutional accountability. Empirical, legal, organisational, and cross-cultural evaluation remains necessary before broader claims of practical validity can be made.

Keywords: AI governance, accountability, Confucian ethics, relational responsibility, institutional accountability

1. Introduction

Artificial intelligence systems are increasingly embedded in information access, education, public services, organisational decision making, and digital platforms. Their operation depends on a chain of technical and institutional choices involving model development, data practices, interfaces, deployment settings, professional workflows, user practices, and oversight. Research on sociotechnical systems has shown that ethical analysis becomes incomplete when abstraction removes the social actors and institutional settings that shape an automated system in practice (Selbst et al., 2019). AI accountability research reaches a similar conclusion by treating responsibility as an institutional problem of answerability,

documentation, review, and control across the system lifecycle (Lechterman, 2022; Novelli et al., 2024; Raji et al., 2020).

This problem is especially visible when several organisations participate in one AI-enabled service. A model provider controls model architecture and documentation. A platform provider may control interfaces, retrieval settings, or monitoring. A deploying institution determines the purpose of use and local safeguards. Professionals interpret outputs within domain practice. Users make decisions about reliance and verification. Regulators and institutional leaders define legal and organisational constraints. When an outcome is unreliable, biased, privacy-intrusive, or difficult to contest, accountability depends on how these roles and relationships fit together. Existing governance frameworks provide strong principles and increasingly detailed requirements for transparency, safety, privacy, human oversight, risk management, and accountability (European Union, 2024; Hagendorff, 2020; Jobin et al., 2019; OECD, 2024; UNESCO, 2021). The remaining analytical challenge is to trace how duties arise and remain connected across multi-actor systems.

Confucian relational ethics offers a useful theoretical resource for this challenge because it places moral agency within roles, relationships, appropriate conduct, trustworthiness, and coordinated social practice (Ames, 2011; Ames & Rosemont, 1998; Rosemont, 2015). Contemporary work has already connected these resources to technology and AI. Confucian role ethics has informed robot ethics and human–robot interaction (Kim & Strudler, 2023; Zhu et al., 2019; Zhu et al., 2020). Ritual propriety has been developed as a resource for technology ethics and interaction design (Wang, 2021; Wong, 2019). Relational decision making and consent have been examined through the Confucian idea of oneness (Wong, 2021). Trustworthiness has been connected directly to trustworthy AI and governance (Wong, 2025). Privacy scholarship has also explored relational and intimacy-based Confucian perspectives (Chan, 2025). These contributions establish a substantive foundation for relational AI ethics, while their governance implications remain dispersed across separate concepts and application domains.

The present study focuses on a narrower problem: the movement from relational ethics to institutional accountability. It develops the Confucian Relational Responsibility Framework (CRR) as an exploratory framework for tracing responsibility across organisations, roles, dependencies, use contexts, and governance procedures. The paper asks one research question:

RQ. How can Confucian relational ethics be translated into a structured approach to institutional accountability in multi-actor AI systems?

The contribution is intentionally focused. CRR provides an analytical layer between relational ethical concepts and established accountability mechanisms. Its five responsibility categories identify differentiated duties, dependency-generated duties, procedural boundaries, conditions of trustworthy conduct, and mechanisms for coordinating accountability. The categories are then organised through actors, relationships, scenarios, responsibilities, and mechanisms. The framework is conceptual and exploratory. Its purpose is to make responsibility structures more analysable and to provide a basis for later empirical, legal, and institutional evaluation.

2. Literature Review

2.1 The relational turn in AI and robot ethics

A relational turn in robot ethics has shifted attention from the internal properties of artificial agents toward the social relationships through which technological systems acquire ethical meaning. From this perspective, moral consideration can be understood as emerging partly through the relations between humans and technological entities rather than solely from the intrinsic capacities of artificial agents (Coeckelbergh, 2010, 2020). This relational orientation has also been developed through Confucian accounts of roles, rituals, and socially situated moral practices. Kim and Strudler (2023), for example, distinguish conventional rights-oriented approaches from a rites-oriented understanding of robots embedded within social practices. Ritualised human–robot interaction may consequently be evaluated in terms of its contribution to human moral cultivation and the maintenance of appropriate social relations (Wong, 2019). Extending this perspective to design, Zhu et al. (2019) draw on Confucian role ethics to conceptualise morally significant human–robot relationships, while practices such as moral rebuke further illustrate how normative expectations can be incorporated into human–robot interaction (Zhu et al., 2020). Taken together, this literature marks a broader movement away from isolated questions of moral status toward roles, relationships, interaction norms, and situated forms of responsibility.

The relational turn is important for AI governance because contemporary systems are also embedded in networks of organisations and users. The relevant relation may connect a developer to a model provider, a provider to a platform, a platform to a deploying institution, or an institution to professionals and users. This wider sociotechnical view is consistent with work that warns against treating algorithmic systems as detached technical objects (Selbst et al., 2019). Relational ethics therefore offers a conceptual vocabulary for identifying dependencies that accountability analysis can then examine institutionally.

2.2 Translating Confucian concepts into AI ethics

A second pattern concerns the translation of specific Confucian concepts into contemporary AI ethics. Confucian role ethics offers a vocabulary for understanding differentiated obligations that arise from social position, relational capacity, and situated responsibilities (Ames, 2011; Rosemont, 2015). Within this broader framework, ritual propriety, or *li*, has been interpreted as a normative resource for organising technological practices and embodied forms of interaction (Wang, 2021). Wong (2021) extends this relational approach through the Confucian concept of oneness, showing how AI-mediated decision making and consent may generate responsibilities that exceed the boundaries of an isolated individual. Questions of trustworthy AI have likewise been reframed through *xin*, with ethical attention directed not only toward technical systems but also toward the trustworthiness of individuals, institutions, regulators, and communicative practices (Wong, 2025). In the context of privacy, Chan (2025) demonstrates how Confucian relational thought can foreground intimacy, contextual relationships, and relational boundaries as ethically significant considerations. Together, these studies illustrate how specific Confucian concepts can be reconstructed as analytical resources for addressing contemporary problems in AI ethics and governance.

This translation activity demonstrates that Confucian resources can enter AI ethics through recognisable governance problems. Comparative analysis also shows that shared ethical terms can carry different normative meanings across cultural contexts (Fung & Etienne, 2023). A term such as trustworthiness may refer to technical reliability in one framework and

to the credibility of persons and institutions in another. CRR treats this difference as analytically productive. It uses translation to connect philosophical resources with governance problems while preserving the relational structure that gives those resources their distinctive value.

2.3 Semantic compression and the translation problem

Translation also creates a methodological risk. Philosophical concepts can become semantically compressed when they are converted too quickly into policy labels, design requirements, or checklist items. Ritual propriety contains dimensions of situated conduct, social formation, and relational appropriateness that exceed a simple interaction rule (Wang, 2021). Confucian trustworthiness concerns patterns of credible conduct and relational practice that extend beyond system-level reliability indicators (Wong, 2025). Relational privacy perspectives similarly contain contextual and interpersonal dimensions that cannot be reduced to a single consent event (Chan, 2025). Comparative AI ethics reinforces the same caution by showing that apparent vocabulary convergence may conceal substantive normative divergence (Fung & Etienne, 2023).

For CRR, semantic compression is treated as a design constraint on theory construction. Governance translation needs enough abstraction to create usable responsibility categories while retaining the ethical function of the original concept. This requirement motivates an explicit translation protocol in Section 3. The protocol separates semantic interpretation, accountability alignment, responsibility formulation, and mechanism traceability. The separation makes each interpretive move visible and reduces the appearance of a direct or automatic equivalence between classical vocabulary and contemporary governance language.

2.4 Institutional accountability as the missing layer

The third body of literature concerns accountability itself. AI accountability has been described as an answerability relation involving agents, forums, standards, processes, and consequences (Novelli et al., 2024). Lechterman (2022) highlights an accountability gap created by both technical characteristics and the social context of AI deployment. Auditing research further connects accountability to documentary evidence, organisational processes, review procedures, and lifecycle governance (Mökander, 2023; Raji et al., 2020). These approaches provide strong institutional mechanisms, yet they generally take governance structures and accountability requirements as their starting point, leaving the relational ethical account of how duties arise among interdependent actors largely unaddressed.

The Confucian AI ethics literature approaches the problem from the opposite direction. It provides rich concepts for roles, relations, ritual conduct, trust, consent, privacy, and moral interaction, while institutional accountability remains less systematically integrated. The gap addressed by CRR lies between these literatures. The framework asks how relational ethical concepts can be translated into responsibility categories that remain connected to concrete governance mechanisms. This gap is narrower than the broad project of developing a complete Confucian theory of AI governance and gives the present study a more precise analytical target.

Table 1

Literature synthesis supporting the CRR problem definition

Literature pattern	Representative sources	What the literature	Implication for
--------------------	------------------------	---------------------	-----------------

		establishes	CRR
Relational turn	Coeckelbergh, 2010; Kim & Strudler, 2023; Wong, 2019; Zhu et al., 2019, 2020	Ethical analysis can focus on roles, relations, interaction norms, and situated practice.	Accountability analysis should begin with relational position and dependency.
Governance translation	Wang, 2021; Wong, 2021, 2025; Chan, 2025	Confucian concepts can illuminate ritual boundaries, consent, trustworthiness, and privacy in AI contexts.	CRR can derive responsibility categories from established conceptual bridges.
Semantic compression	Fung & Etienne, 2023; Wang, 2021; Wong, 2025; Chan, 2025	Cross-cultural translation can reduce philosophical density when concepts become governance labels.	Translation steps should remain explicit and traceable.
Institutional accountability gap	Lechterman, 2022; Novelli et al., 2024; Raji et al., 2020; Mökander, 2023	AI accountability requires answerability, records, oversight, audit, and organisational processes.	CRR should connect relational duties to institutional mechanisms.

3. Methodology

3.1 Exploratory conceptual model development

This study uses an exploratory conceptual model development design. Conceptual research requires an explicit research design, clear theoretical inputs, and a transparent account of how concepts are combined into a model (Jaakkola, 2020). Jabareen (2009) similarly treats a conceptual framework as a network of linked concepts developed through a traceable process of interpretation and synthesis. Whetten (1989) emphasises the importance of specifying what elements belong in a theoretical explanation, how they are related, and why those relations matter. These principles guide the construction of CRR.

The conceptual corpus is purposive and theory-driven. It contains three source domains. The first is foundational Confucian role and relational ethics, used to interpret role, relationship, ritual propriety, trustworthiness, and coordination. The second is contemporary Confucian technology and AI ethics, used to identify established bridges between these concepts and technological practice. The third is AI governance and accountability research, used to identify recognised problems and mechanisms involving answerability, sociotechnical dependency, documentation, audit, oversight, and institutional responsibility. The corpus is bounded by the research question and functions as a selective source base for the present conceptual model.

3.2 Concept selection

Five concepts are selected through purposive conceptual sampling: role, relationship, ritual propriety, trustworthiness, and coordination. Each concept must satisfy three inclusion criteria. First, it must have a defensible basis in Confucian relational or technology ethics. Second, it must illuminate a distinguishable accountability problem in multi-actor AI systems. Third, it must support a governance translation that can be linked to observable institutional

practices. The selection therefore privileges structural concepts that help organise responsibility. Other Confucian concepts, including benevolence, righteousness, and wisdom, remain important ethical resources and are reserved for later extensions because their functions in the present accountability model would require additional interpretive steps.

The concept of coordination requires clarification. CRR uses coordination as an analytical rendering of relational harmony and coordinated role conduct within the governance problem examined here. Confucian technology ethics treats harmony and relational ordering as processes that connect persons, roles, practices, and technological environments (Wong, 2012; Wong & Wang, 2021). CRR narrows this function to the task of keeping dispersed responsibilities connected across organisations and time.

3.3 Governance translation protocol

The governance translation protocol contains four steps. Step 1, semantic function, identifies the ethical function of a selected concept in the supporting philosophical literature. Step 2, accountability alignment, identifies a recurring institutional problem in AI governance that has a plausible structural relation to that ethical function. Step 3, responsibility formulation, expresses the relation as a distinct responsibility category. Step 4, mechanism traceability, connects the category to governance practices that can be assigned, documented, monitored, reviewed, or challenged. The protocol creates a visible chain from philosophical interpretation to governance use.

Four analytical checks are applied to each translation. Conceptual fidelity asks whether the governance category retains the central relational function used to justify the concept. Governance relevance asks whether the category addresses a recognised AI accountability problem. Category distinctiveness asks whether the category performs a function that can be separated analytically from the other four categories. Mechanism traceability asks whether the category can be connected to institutional practices that leave evidence for review. These checks serve as conceptual quality controls. Empirical validation and claims of cultural exhaustiveness remain outside the evidentiary scope of the present study.

3.4 Framework construction

The five translated categories are organised into a responsibility mapping sequence composed of actors, relationships, scenarios, responsibilities, and mechanisms. Actors identifies participants and relevant capacities. Relationships identifies provision, authority, dependency, data flow, professional mediation, and reliance. Scenarios specifies the purpose and institutional context of AI use. Responsibilities applies the five CRR categories to the mapped relations. Mechanisms connects identified duties to disclosure, human review, documentation, monitoring, audit, complaint procedures, escalation, and institutional review. The sequence is designed to reveal responsibility gaps that can arise between organisations or stages of deployment.

Framework construction therefore proceeds from literature synthesis to concept selection, from concept selection to governance translation, and from translated categories to a reusable accountability mapping procedure. The output is a conceptual framework whose current evidentiary claim is analytical coherence and traceability. Practical effectiveness, stakeholder acceptance, and legal compatibility require later evaluation.

Table 2*Concept derivation and governance translation within CRR*

Selected concept	Ethical function	AI accountability problem	CRR category	Representative support
Role	Differentiated obligations linked to position, capacity, and social expectation.	Role ambiguity and unclear division of duties across the AI lifecycle.	Role responsibility	Ames, 2011; Rosemont, 2015; Zhu et al., 2019
Relationship	Duties generated through interdependence and effects on others.	Responsibility fragmentation across providers, institutions, professionals, and users.	Relational responsibility	Wong, 2021; Selbst et al., 2019
Ritual propriety	Appropriate conduct, boundaries, procedures, and situated norms.	Boundary crossing, weak disclosure, inadequate review, and inappropriate use.	Boundary-setting responsibility	Wang, 2021; Wong, 2019
Trustworthiness	Credible conduct, reliability, sincerity, and conditions of justified trust.	Unreliable information, weak verifiability, opacity, and limited routes for challenge.	Trust responsibility	Wong, 2025; Floridi & Cowls, 2019
Coordination	Relational ordering and the maintenance of harmony across connected roles and practices.	Fragmented records, monitoring, oversight, and accountability across organisations.	Coordinated accountability	Wong, 2012; Wong & Wang, 2021; Raji et al., 2020; Mökander, 2023

4. The Confucian Relational Responsibility Framework

4.1 Five responsibility categories

Role responsibility identifies duties that follow from position, capability, authority, knowledge, and control. Confucian role ethics provides the ethical basis for differentiated obligations, while AI accountability research shows why the location of authority and capacity matters for answerability (Ames, 2011; Lechterman, 2022; Rosemont, 2015; Novelli et al., 2024). In practice, a model provider may hold duties concerning model documentation and known limitations. A deploying institution may hold duties concerning procurement, local configuration, supervision, and appropriate use. Professionals may hold duties concerning verification and user guidance. Role responsibility prevents shared participation from being interpreted as equal responsibility.

Relational responsibility identifies additional duties generated through dependency, reliance, and mutual influence. Wong (2021) shows how interconnected AI environments can transform apparently personal choices into acts with effects on others. Sociotechnical research similarly demonstrates that system outcomes arise through relations among technical components, institutions, and social actors (Selbst et al., 2019). Relational

responsibility therefore asks what one actor owes another because the second actor depends on information, controls, documentation, or decisions provided by the first.

Boundary-setting responsibility concerns the limits and procedures that make AI use appropriate within a specific institutional context. Ritual propriety provides a resource for understanding situated norms and appropriate forms of conduct in technological practice (Wang, 2021; Wong, 2019). In governance terms, this category covers disclosure, consent, human review, sensitive use restrictions, escalation thresholds, data access rules, and procedures for contesting outputs. The central question is whether an institution has defined where AI authority begins, where it ends, and which human or organisational intervention is required at the boundary.

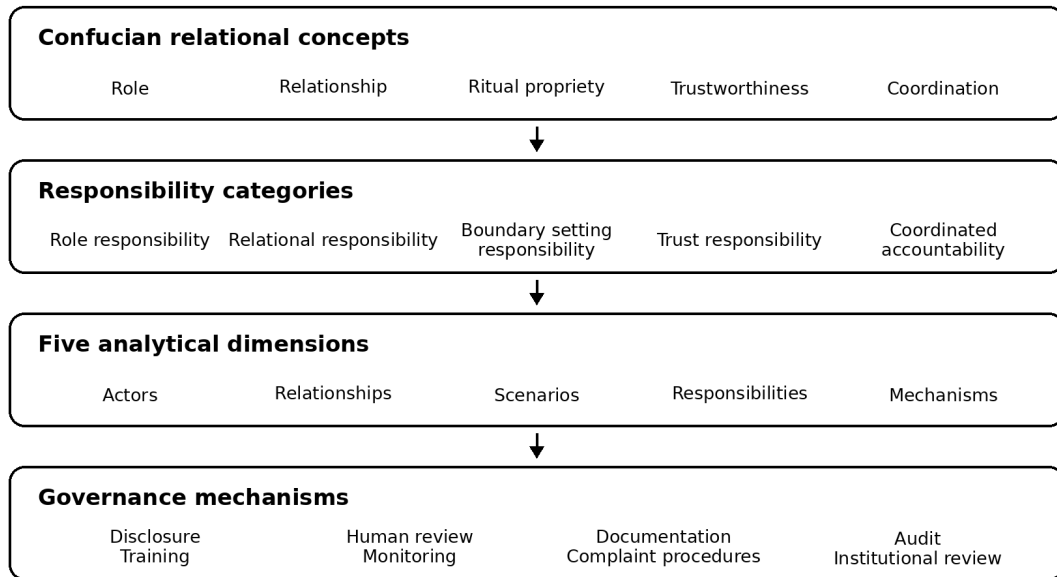
Trust responsibility concerns the conditions under which reliance on an AI-supported service is justified and open to challenge. Confucian analysis of trustworthy AI broadens trustworthiness from a property of a system to the conduct of developers, institutions, regulators, and communicative practices (Wong, 2025). CRR therefore connects reliability, verifiability, transparency, correction, uncertainty communication, and human challenge to the quality of the relationship between institutions and users. Trust responsibility asks which practices make an actor worthy of reliance in a given role.

Coordinated accountability connects the previous categories across organisational boundaries and across time. Its relational basis draws on Confucian discussions of harmony and coordinated role conduct (Wong, 2012; Wong & Wang, 2021). Its governance expression draws on auditing and accountability research that emphasises records, lifecycle processes, internal review, monitoring, and answerability (Mökander, 2023; Raji et al., 2020). Coordinated accountability asks how duties remain visible when responsibility is distributed among several organisations and when systems change after deployment.

4.2 Five analytical dimensions

CRR organises these responsibility categories through five analytical dimensions. Actors identifies participants and relevant capacities. Relationships identifies the connections through which duties and dependencies arise. Scenarios specifies the institutional purpose, affected users, and risk context. Responsibilities classifies the duties that become visible through role responsibility, relational responsibility, boundary-setting responsibility, trust responsibility, and coordinated accountability. Mechanisms connects those duties to institutional procedures and evidence. The dimensions form a sequential mapping procedure and carry no claim of empirically validated causal relations.

The practical value of the sequence lies in traceability. An analyst can begin with a concrete AI service, identify the actors and dependencies, specify the use context, classify the responsibilities that arise, and then examine whether governance mechanisms exist to carry those responsibilities. A final residual responsibility check asks whether any important duty remains unassigned, undocumented, or unreviewed because it falls between organisational boundaries. This check is especially important for externally supplied models and platforms where technical control and public responsibility may be located in different organisations.



CRR connects relational ethical concepts to responsibility categories and organises them through a sequential accountability analysis.

Figure 1

The CRR framework linking selected Confucian relational concepts, responsibility categories, analytical dimensions, and governance mechanisms.

5. Illustrative Application

5.1 Worked example: generative AI search in a public knowledge institution

A generative AI search service in a public knowledge institution provides a useful worked example because responsibility is distributed across several technical and institutional layers. The service may use a foundation model supplied by one organisation, a search or retrieval layer supplied by another, local configuration determined by the institution, professional review performed by information staff, and user reliance shaped by interface design and guidance. A hallucinated answer or misleading citation can therefore emerge from several connected decisions. CRR treats the problem as an accountability mapping exercise.

The mapping begins by identifying participants and dependencies. The model provider controls model capabilities and documentation. The platform provider controls retrieval, interface behaviour, updates, and monitoring tools. The institution selects the service, defines local use conditions, and establishes safeguards. Information professionals mediate use, evaluate outputs, and guide users. Users make decisions about verification and reliance. Each relationship creates a potential transfer of information, authority, or risk. The responsibility categories then clarify which duties arise and how they should remain connected.

Table 3

Worked CRR mapping for generative AI search

CRR element	Worked mapping
Actors	Model provider; platform provider; deploying institution; information professionals; users; institutional leaders or regulators where

	applicable.
Relationships	Model provision; platform integration; institutional deployment; professional mediation; user reliance; reporting and oversight.
Role responsibility	Provider documentation; platform configuration and update information; institutional procurement and safeguards; professional verification and guidance; user verification within reasonable capacity.
Relational responsibility	Communication of limitations across provider and institution; escalation when local risks depend on vendor behaviour; correction duties when one actor's output affects another actor's decision.
Boundary-setting responsibility	Disclosure that generative AI is used; source-checking requirements; limits for sensitive or high-consequence queries; human review and escalation rules.
Trust responsibility	Verifiable sources; uncertainty communication; correction routes; visible explanation of service limits; opportunities for human challenge.
Coordinated accountability	Incident records; vendor documentation; monitoring; audit; complaint procedures; periodic institutional review; documented response to model or platform changes.
Residual responsibility check	Identify duties that remain unclear when model updates, retrieval changes, or vendor decisions alter the service without a corresponding institutional review.

The worked example shows the distinction between responsibility categories and governance mechanisms. A disclosure notice is a mechanism. Its analytical basis may involve role responsibility because an institution has a duty to communicate, boundary-setting responsibility because disclosure defines the conditions of appropriate use, and trust responsibility because disclosure supports informed reliance. CRR therefore avoids treating a single governance mechanism as belonging exclusively to one ethical category. The categories explain why a mechanism is required and which relationships it is intended to govern.

5.2 Additional illustrations

Automated knowledge organisation provides a second illustration. Classification outputs may depend on training data, inherited metadata, model behaviour, professional standards, and local institutional policy. CRR can distinguish vendor duties concerning system documentation from professional duties concerning validation and correction. Boundary-setting responsibility can identify categories that require human review. Trust responsibility can require correction routes and visible explanations of automated assistance. Coordinated accountability can connect vendor records, local review decisions, and institutional policy when culturally inappropriate or biased classifications are detected.

Recommendation systems provide a third illustration. Influence over information exposure can be distributed across ranking models, platform objectives, institutional settings, and user behaviour. Relational responsibility maps how provider choices and institutional goals combine to shape exposure. Boundary-setting responsibility can identify restrictions for children, sensitive topics, or public service contexts. Trust responsibility can support meaningful explanation and challenge. Coordinated accountability can connect monitoring of recurring bias to review, reporting, and policy change. These applications remain conceptual probes whose purpose is to demonstrate analytical use. Evidence of framework effectiveness requires later empirical study.

6. Discussion

6.1 From relational ethics to institutional accountability

The central contribution of CRR is the construction of an explicit bridge between relational ethical concepts and institutional accountability. Existing Confucian AI ethics has already established substantial work on roles, rites, trust, consent, privacy, and moral interaction. Existing AI governance research has developed strong accounts of answerability, sociotechnical context, documentation, audit, and oversight. CRR brings these literatures into one analytical sequence. The framework begins with the relational source of a duty and follows that duty into institutional mechanisms that can make it visible and reviewable.

This bridge also clarifies the function of Confucian concepts in AI governance. CRR positions Confucianism as a complementary analytical resource alongside existing governance frameworks. The selected concepts help identify differentiated and relational duties. Legal rights, safety requirements, privacy protections, risk classification, and procedural safeguards continue to define binding requirements in regulated settings. CRR adds relational traceability by making dependencies and responsibility gaps easier to examine.

6.2 What CRR adds to accountability analysis

Accountability models often begin with institutional requirements and ask which actors must provide accounts, evidence, oversight, or enforcement (Lechterman, 2022; Novelli et al., 2024). CRR adds an earlier analytical question: what relational condition generated the duty that accountability mechanisms are expected to carry? Role responsibility links duties to position and capacity. Relational responsibility identifies duties created by dependency. Boundary-setting responsibility identifies limits and review points. Trust responsibility identifies conditions of justified reliance and challenge. Coordinated accountability identifies the organisational work needed to keep these duties connected. The result is a responsibility vocabulary that can support the design and interpretation of accountability processes.

CRR also distinguishes shared influence from equal responsibility. Multi-actor systems frequently produce outcomes through combined technical and institutional choices. A relational approach can acknowledge this shared influence while role responsibility preserves differences in authority and capability. This distinction is useful for vendor-dependent services because a deploying institution may retain public obligations while possessing limited technical control. The framework makes that asymmetry visible and directs attention to contracts, documentation, escalation, monitoring, and review as mechanisms for managing the relationship.

6.3 Translation limits and rights-based governance

The governance translation itself requires caution. Confucian concepts are historically and philosophically richer than the responsibility categories developed here. The protocol offers functional correspondences for a contemporary governance problem. Semantic equivalence remains outside its claim. Cross-cultural AI ethics has shown that common vocabulary can conceal divergent normative commitments (Fung & Etienne, 2023). CRR therefore treats its translations as provisional and open to philosophical critique, stakeholder interpretation, and comparison with other ethical traditions.

Relational analysis also requires careful integration with rights-based and risk-based governance. Individual rights, due process, equality, privacy, safety, and routes for challenge

remain essential protections. The European Union AI Act illustrates the increasing institutionalisation of risk management, documentation, oversight, and organisational duties in AI governance (European Union, 2024). CRR can complement such requirements by tracing how responsibilities travel across provider and deployment relationships. Its current role is analytical and interpretive, with legal compatibility remaining a question for future research.

7. Limitations and Future Research

CRR remains an exploratory conceptual framework. The concept selection is purposive, the philosophical translations are interpretive, and the worked applications remain conceptual illustrations. Empirical validation has yet to be undertaken. The current framework concentrates on five structural concepts and therefore leaves wider Confucian resources outside its analytical boundary. The application context is concentrated in information and knowledge environments, which may make some mechanisms more visible than they would be in other sectors. Stakeholder studies, institutional implementation, comparative legal analysis, and cross-cultural testing remain future tasks.

Future research can examine CRR through expert interviews, Delphi studies, focus groups with AI governance practitioners, institutional case studies, and procurement or deployment reviews. A first empirical question concerns whether users of the framework can identify responsibility gaps more consistently than they can with principle-based guidance alone. A second concerns whether the five categories remain conceptually distinct when applied by practitioners. A third concerns the interaction between CRR and established legal or organisational accountability models. Later research may compare CRR with other relational traditions, develop assessment instruments, and test the framework across sectors after its conceptual categories have received stakeholder evaluation.

8. Conclusion

This study addressed a focused problem in AI governance: the gap between relational ethical concepts and institutional accountability in multi-actor AI systems. A literature synthesis showed that relational and Confucian approaches have already contributed substantial work on roles, rites, trust, consent, privacy, and situated interaction, while accountability research has developed institutional mechanisms for answerability, audit, documentation, oversight, and review. CRR connects these strands through a transparent conceptual construction process.

The framework translates role, relationship, ritual propriety, trustworthiness, and coordination into role responsibility, relational responsibility, boundary-setting responsibility, trust responsibility, and coordinated accountability. It then organises these categories through actors, relationships, scenarios, responsibilities, and mechanisms. The worked example illustrates how this structure can trace duties across providers, institutions, professionals, and users and connect them to reviewable governance practices. CRR therefore offers an exploratory analytical layer between relational ethics and institutional accountability. Its broader value depends on future empirical, legal, organisational, and cross-cultural evaluation.

References

Ames, R. T. (2011). *Confucian role ethics: A vocabulary*. University of Hawaii Press.

- Ames, R. T., & Rosemont, H., Jr. (1998). *The Analects of Confucius: A philosophical translation*. Ballantine Books.
- Chan, B. S. B. (2025). *A preliminary investigation of Confucianism's impact on privacy debates in a.I. And technology ethics*. In L. Checketts & B. S. B. Chan (Eds.), *Social and ethical considerations of AI in East Asia and beyond* (pp. 15–31). Springer.
https://doi.org/10.1007/978-3-031-77857-5_2
- Coeckelbergh, M. (2010). *Robot rights? Towards a social-relational justification of moral consideration*. *Ethics and Information Technology* , 12(3), 209–221.
<https://doi.org/10.1007/s10676-010-9235-5>
- Coeckelbergh, M. (2020). *AI ethics*. Mit Press. <https://doi.org/10.7551/mitpress/12549.001.0001>
- European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*. *Official Journal of the European Union*.
- Floridi, L., & Cowls, J. (2019). *A unified framework of five principles for AI in society*. *Harvard Data Science Review* , 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Fung, P., & Etienne, H. (2023). *Confucius, cyberpunk and Mr. Science: Comparing AI ethics principles between China and the EU*. *AI and Ethics* , 3(2), 505–511.
<https://doi.org/10.1007/s43681-022-00180-6>
- Hagendorff, T. (2020). *The ethics of AI ethics: An evaluation of guidelines*. *Minds and Machines* , 30(1), 99–120. <https://doi.org/10.1007/s11023-020-09517-8>
- Jaakkola, E. (2020). *Designing conceptual articles: Four approaches*. *Ams Review* , 10, 18–26.
<https://doi.org/10.1007/s13162-020-00161-0>
- Jabareen, Y. (2009). *Building a conceptual framework: Philosophy, definitions, and procedure*. *International Journal of Qualitative Methods* , 8(4), 49–62.
<https://doi.org/10.1177/160940690900800406>
- Jobin, A., Ienca, M., & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. *Nature Machine Intelligence* , 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kim, T. W., & Strudler, A. (2023). *Should robots have rights or rites? Communications of the ACM* , 66(6), 78–85. <https://doi.org/10.1145/3571721>
- Lechterman, T. M. (2022). *The concept of accountability in AI ethics and governance*. In J. B. Bullock, Y. C. Chen, J. Himmelreich, V. M. Hudson, A. Korinek, M. M. Young, & B. Zhang (Eds.), *The Oxford handbook of AI governance* (pp. 164–182). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780197579329.013.10>
- Mökander, J. (2023). *Auditing of AI: Legal, ethical and technical approaches*. *Digital Society*, 2, Article 49. <https://doi.org/10.1007/s44206-023-00074-y>
- Novelli, C., Taddeo, M., & Floridi, L. (2024). *Accountability in artificial intelligence: What it is and how it works*. *AI & Society* , 39, 1871–1882. <https://doi.org/10.1007/s00146-023-01635-y>
- OECD. (2024). *Revised recommendation of the Council on Artificial Intelligence*. OECD.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). ACM.
<https://doi.org/10.1145/3351095.3372873>
- Rosemont, H., Jr. (2015). *Against individualism: A Confucian rethinking of the foundations of morality, politics, family, and religion*. Lexington Books.

- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). ACM. <https://doi.org/10.1145/3287560.3287598>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.
- Wang, T. X. (2021). *Confucian ritual technicity and philosophy of technology*. In P. H. Wong & T. X. Wang (Eds.), *Harmonious technology: A Confucian ethics of technology* (pp. 10–28). Routledge. <https://doi.org/10.4324/9780429292842-1>
- Whetten, D. A. (1989). *What constitutes a theoretical contribution?* *Academy of Management Review*. , 14(4), 490–495. <https://doi.org/10.5465/amr.1989.4308371>
- Wong, P. H. (2012). *Dao, harmony and personhood: Towards a Confucian ethics of technology*. *Philosophy & Technology*. , 25(1), 67–86. <https://doi.org/10.1007/s13347-011-0021-z>
- Wong, P. H. (2019). *Rituals and machines: A Confucian response to technology-driven moral deskilling*. *Philosophies*. , 4(4), Article 59. <https://doi.org/10.3390/philosophies4040059>
- Wong, P. H. (2021). *Artificial intelligence, personal decisions, consent, and the Confucian idea of oneness*. In P. H. Wong & T. X. Wang (Eds.), *Harmonious technology: A Confucian ethics of technology* (pp. 79–94). Routledge. <https://doi.org/10.4324/9780429292842-5>
- Wong, P. H. (2025). *Confucian 'trustworthy AI': Diversifying a keyword in the ethics of AI and governance*. In L. Checketts & B. S. B. Chan (Eds.), *Social and ethical considerations of AI in East Asia and beyond* (pp. 3–14). Springer. https://doi.org/10.1007/978-3-031-77857-5_1
- Wong, P. H., & Wang, T. X. (Eds.). (2021). *Harmonious technology: A Confucian ethics of technology*. Routledge. <https://doi.org/10.4324/9780429292842>
- Zhu, Q., Williams, T., & Wen, R. (2019). *Confucian robot ethics*. In D. Wittkower (Ed.), 2019 Computer Ethics Philosophical Enquiry Proceedings (11 pp.). <https://doi.org/10.25884/5qbh-m581>
- Zhu, Q., Williams, T., Jackson, B., & Wen, R. (2020). *Blame-laden moral rebukes and the morally competent robot: A Confucian ethical perspective*. *Science and Engineering Ethics*. , 26(5), 2511–2526. <https://doi.org/10.1007/s11948-020-00246-w>