Governing artificial intelligence ethically: A fuzzy DEMATEL analysis of Australia's AI principles in the Indonesian policy context

Feliks Prasepta Sejahtera Surbakti[1*], Hotma Antoni Hutahaean[2], Liquita Florencia Putri[3], Henry Edison[4]

[1,2,3]Department of Industrial Engineering, Atma Jaya Catholic University of Indonesia, Tangerang, Indonesia.
[4]Department of Software Engineering, Blekinge Institute of Technology, Karlskrona, Sweden.

Corresponding author: Feliks Prasepta Sejahtera Surbakti (*Email: feliks.prasepta@atmajaya.ac.id*)

Abstract

The rapid diffusion of artificial intelligence (AI) in Indonesia has outpaced the development of robust ethical governance mechanisms. Although Indonesia has issued ethical guidelines for AI, these remain largely declarative and lack operational enforcement. In contrast, Australia's AI Ethics Principles provide a more structured and implementable framework. However, the interactions among these ethical principles and their relative importance within the Indonesian policy context remain insufficiently understood. This study examines the causal relationships among AI ethical principles adapted from Australia's framework to identify key governance drivers for strengthening ethical AI implementation in Indonesia. A quantitative perception-based study was conducted using survey data from 109 Indonesian AI stakeholders, including developers, practitioners, and users. The Fuzzy DEMATEL method was employed to analyze causal relationships among eight AI ethical principles. The results indicate that Accountability, Transparency and Explainability, Contestability, and Privacy Protection and Security act as causal drivers within the ethical governance system, while Fairness, Human-Centered Values, Reliability and Safety, and Human, Societal, and Environmental Wellbeing function as outcome-oriented principles. These findings suggest that ethical outcomes depend primarily on strong upstream governance mechanisms. The study provides empirical guidance for localizing international AI ethics frameworks and advancing actionable AI governance in Indonesia.

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Atma Jaya Catholic University of Indonesia (protocol code: Ref. No. UAJ-ETIK/AI-2025/03, date of approval: 7 April 2025).

1. Introduction

The global expansion of Artificial Intelligence (AI) has transformed the way societies interact, make decisions, and distribute information. AI systems now influence social media algorithms, employment decisions, healthcare diagnostics, and public administration processes, producing both efficiency and ethical tension. While AI offers opportunities to enhance productivity and innovation, it simultaneously raises profound concerns about fairness, accountability, privacy, and the erosion of human autonomy [1, 2]. The discourse around AI ethics has thus shifted from technical compliance to broader questions of governance, responsibility, and trust—especially in societies with developing regulatory infrastructures.

In Indonesia, AI has rapidly entered various domains such as e-commerce, financial technology, agriculture, education, and government services. However, ethical governance has not evolved at the same pace as technological adoption. Cases of algorithmic bias in recruitment, unauthorized use of personal data, and AI-generated misinformation during elections have exposed weaknesses in Indonesia's current policy framework [3]. The absence of strong accountability mechanisms has also led to disputes about liability and trust, particularly when AI systems produce harmful or discriminatory outcomes. These challenges underline the need for a comprehensive approach that links AI ethics to national governance and socio-technical realities.

The Ministry of Communication and Informatics introduced Circular Letter No. 9/2023 as a preliminary step toward establishing an ethical framework for AI. While this guideline articulates nine ethical principles—including humanity, inclusivity, and data protection—it remains largely declarative and lacks operational mechanisms for implementation or oversight [4]. The absence of enforcement tools, assurance frameworks, and appeal mechanisms makes it difficult for practitioners and institutions to translate principles into accountable practices. In contrast, countries such as Australia have implemented more structured frameworks that include measurable standards, assurance checklists, and public transparency statements [5]. This contrast provides a valuable benchmark for understanding how ethical principles might interact and be prioritized in a policy-oriented context.

Scholarly discussions on AI ethics have generally emphasized the identification of universal principles—fairness, accountability, transparency, privacy, and beneficence—yet few studies examine how these principles influence one another within specific socio-political environments [6, 7]. Ethics, as implemented in practice, operates not as a static list of values but as an interdependent system where principles may reinforce, conflict, or mediate one another. For example, transparency can strengthen accountability, while excessive focus on privacy may obscure fairness if data collection becomes overly restrictive. Recognizing these relationships is essential for designing context-sensitive governance frameworks that move beyond ethical formalism to operational realism.

This study addresses this conceptual and practical gap by using the Fuzzy Decision-Making Trial and Evaluation Laboratory (Fuzzy DEMATEL) method to map causal and effect relationships among eight AI ethical principles adapted from Australia's framework. Unlike conventional descriptive analyses, Fuzzy DEMATEL enables the identification of cause–effect linkages between principles, revealing which act as foundational drivers and which are more outcome-oriented [8]. By applying this method to the Indonesian context, this research not only contributes empirically to the understanding of ethical interdependencies but also provides evidence-based guidance for policymakers seeking to localize international ethical standards.

Accordingly, this paper seeks to address two central research questions: which AI ethical principles are perceived as the most influential drivers in promoting ethical AI implementation in Indonesia, and how these principles interact causally to shape ethical AI governance. Through these questions, the study contributes to both theoretical and practical discussions on responsible AI. Theoretically, it advances the sociotechnical understanding of AI ethics as a dynamic system of interrelated principles rather than isolated values. Practically, it offers policymakers a prioritization framework to strengthen Indonesia's AI ethics through evidence-based governance strategies that emphasize accountability, transparency,

and contestability. By situating the analysis within the Indonesian policy landscape while benchmarking against Australia's established framework, this research aims to bridge the gap between ethical ideals and regulatory implementation.

The remainder of this paper is organized as follows: Section 2 reviews the theoretical background, Section 3 describes the research methodology, Section 4 presents the key findings, Section 5 discusses implications for policy and governance, and Section 6 concludes with recommendations for future research.

2. Literature Review

Artificial Intelligence (AI) ethics has evolved as a multidisciplinary field at the intersection of technology, philosophy, and governance. It examines how moral principles should guide the design, deployment, and impact of intelligent systems in society [9]. The rapid adoption of AI has amplified existing ethical tensions surrounding fairness, accountability, and transparency while introducing new dilemmas such as algorithmic opacity, bias, and autonomy in machine decision-making [10, 11]. As AI systems increasingly mediate communication, employment, and governance, ethical frameworks have become essential instruments for aligning innovation with social values and public trust. However, the interpretation and implementation of these frameworks remain contested across cultural and political contexts, particularly in developing countries where regulatory capacity is still emerging.

AI ethics frameworks globally share common aspirations—to ensure that AI benefits humanity, protects fundamental rights, and prevents harm—but differ in emphasis and operational maturity. Early frameworks developed by the European Union, OECD, and UNESCO primarily articulated high-level principles, while more recent models, such as those from Australia and Singapore, include operational guidance and accountability tools [12, 13]. Australia's AI Ethics Principles, in particular, offer eight integrated values: (1) human, societal, and environmental wellbeing; (2) human-centered values; (3) fairness; (4) privacy protection and security; (5) reliability and safety; (6) transparency and explainability; (7) contestability; and (8) accountability [14]. These principles form an interconnected system that balances human rights, technological integrity, and institutional oversight, and have been recognized internationally for their practical applicability [15].

In contrast, Indonesia's ethical landscape for AI remains in a formative stage. The government's Circular Letter Kominfo No. 9/2023 outlines nine ethical principles—humanity, inclusivity, safety, credibility, accountability, transparency, personal data protection, and intellectual property protection—but these are expressed primarily as normative statements without corresponding enforcement mechanisms. Several principles mirror those of Australia, but key operational components such as contestability and explainability are missing. Moreover, Indonesia's AI ethics discourse often focuses on compliance or moral appeal rather than systemic governance [16]. As a result, ethical decision-making in AI projects is highly dependent on organizational discretion, leaving gaps in accountability and oversight that hinder the translation of ethical values into tangible practices.

To advance from ethical formalism to actionable governance, scholars have emphasized the sociotechnical nature of AI ethics. Sociotechnical theory posits that technology and society are co-constitutive: technological design shapes social practices, while cultural and institutional contexts influence how technologies are used and governed [17, 18]. Applying this lens, ethical AI governance must not only codify abstract principles but also consider the institutional, economic, and cultural dynamics that condition their implementation. For instance, transparency in AI is meaningful only if it is coupled with public literacy and institutional mechanisms that enable contestation. Similarly, accountability must extend beyond individual developers to include systemic responsibility shared among organizations, regulators, and civil society.

While there is a growing body of research exploring the content of AI ethical principles, relatively few studies have examined how these principles interact with each other in practice. Jobin, et al. [1] identified over 80 global AI ethics frameworks, yet most treat ethical values as discrete categories rather than dynamic elements of a larger system. This reductionist view risks oversimplifying complex ethical relationships. For example, ensuring privacy may sometimes constrain explainability, and promoting fairness may require trade-offs with efficiency. Understanding these interdependencies is therefore crucial to designing effective policy interventions that address not only individual principles but the systemic balance among them. Recognizing the interrelated nature of ethical principles is especially relevant in contexts like Indonesia, where institutional coordination and data governance structures are still evolving.

The Fuzzy Decision-Making Trial and Evaluation Laboratory (Fuzzy DEMATEL) method provides a methodological lens for addressing these complexities. Originally developed by Fontela and Gabus [19] DEMATEL helps identify causal and effect relationships among multiple criteria in systems characterized by interdependence and uncertainty. The fuzzy extension introduced by Kacprzyk [20] enables researchers to incorporate linguistic and subjective assessments into quantitative modeling. In the context of AI ethics, Fuzzy DEMATEL has been increasingly applied to capture expert or stakeholder perceptions about how principles influence one another [21]. By revealing which principles function as drivers (causes) and which operate as dependent effects, the method offers policymakers a structured approach to prioritize interventions.

This study employs Fuzzy DEMATEL to empirically examine the interrelations among Australia's eight AI ethical principles as perceived by Indonesian AI stakeholders. Integrating the sociotechnical perspective with quantitative causal modeling allows this research to bridge normative and operational dimensions of ethics. It recognizes that ethical principles cannot be implemented effectively in isolation; rather, they form a network of influences that shape the trajectory of AI governance. By identifying the causal hierarchies among these principles, this study contributes to the emerging literature on responsible AI governance in the Global South and provides practical insights for policymakers seeking to build context-sensitive frameworks that promote transparency, accountability, and public trust.

3. Methods

This study adopts a quantitative, perception-based design to explore how Indonesian AI stakeholders understand and prioritize ethical principles of artificial intelligence. The research is grounded in the sociotechnical perspective, which recognizes that ethical governance is both a technical and social construct. Accordingly, the methodology aims to capture not only the structural relationships among ethical principles but also how human perceptions and contextual factors shape these relationships. The study integrates two complementary approaches: (1) Fuzzy Decision-Making Trial and Evaluation Laboratory (Fuzzy DEMATEL) to model interdependencies between principles, and (2) Borda Count to determine their perceived priority rankings. Together, these methods provide a robust framework to analyze how ethics functions as an interconnected system rather than as a set of isolated rules.

3.1. Research Design and Data Collection

The research employed a structured online survey distributed between January and March 2025. The questionnaire was developed in Bahasa Indonesia to ensure clarity and accessibility for local participants. It was divided into three sections: (a) demographic and professional background, (b) pairwise influence evaluation between ethical principles, and (c) ranking of principles by importance. The survey instrument was designed using Google Forms and pretested among 12 respondents (academics and AI practitioners) to refine the clarity of linguistic terms and ensure conceptual consistency with the Australian AI Ethics Principles framework [14].

The final sample comprised 109 respondents drawn from multiple sectors, including academia, information technology, government, financial services, and creative industries. Of these, approximately 70% were AI practitioners—developers, data scientists, or researchers—while the remainder were users or policymakers engaged in AI-related activities. The respondents represented a diverse age range (17–55 years, mean = 24.3), educational backgrounds (undergraduate to postgraduate), and regional distribution, though the majority (87%) were based in Java. This demographic profile reflects Indonesia's concentration of digital innovation within urban centers, where access to AI tools and literacy is higher compared to rural areas. Participation was voluntary and anonymous to reduce social desirability bias and encourage honest responses.

3.2. Survey Instrument and Measurement

The instrument measured perceived influence relationships among eight AI ethical principles: (P1) Human, Societal, and Environmental Wellbeing; (P2) Human-Centered Values; (P3) Fairness; (P4) Privacy Protection and Security; (P5) Reliability and Safety; (P6) Transparency and Explainability; (P7) Contestability; and (P8) Accountability. Respondents were asked to evaluate how strongly each principle influenced the others using a five-point fuzzy linguistic scale, ranging from No Influence (0) to Very High Influence (4). These qualitative judgments were subsequently converted into Triangular Fuzzy Numbers (TFN)—defined as triplets (l, m, r), representing the lower, middle, and upper bounds of perceived influence—to preserve uncertainty and subjectivity inherent in ethical evaluation [22]. The linguistic variables were defined as follows:

- No Influence (NO) = (0.00, 0.00, 0.25)
- Very Low Influence (VL) = (0.00, 0.25, 0.50)
- Low Influence (L) = (0.25, 0.50, 0.75)
- High Influence (H) = (0.50, 0.75, 1.00)
- Very High Influence (VH) = (0.75, 1.00, 1.00)

This scale has been widely adopted in previous Fuzzy DEMATEL studies [22, 23] for its balance between interpretability and computational efficiency. Each respondent provided 56 pairwise evaluations, covering all possible directional relationships among the eight principles (8×7 matrix).

3.3. Fuzzy DEMATEL Procedure

The Fuzzy DEMATEL method was employed to model the causal interrelations among principles. This technique is particularly suited for analyzing systems characterized by feedback loops, ambiguity, and subjective judgment [24]. The analytical process followed six stages:

1. Constructing the Fuzzy Direct-Relation Matrix (Z):
Aggregating all individual TFN evaluations to represent the average influence between each pair of principles.
2. Defuzzification (CFCS Method):
Applying the Converting Fuzzy Data into Crisp Scores (CFCS) technique Opricovic and Tzeng [25] to translate fuzzy data into crisp values while preserving their relative magnitude and shape. This step ensures comparability across principles.
3. Normalization of the Matrix (N):
Adjusting the matrix values to a consistent scale to eliminate distortions caused by extreme judgments. The normalized matrix expresses proportional influence strengths.
4. Deriving the Total-Relation Matrix (T):
Computing both direct and indirect effects using matrix operations to reflect the overall influence network among principles.
5. Calculating R and C Values:

Summing rows (R) to indicate the total influence given and columns (C) to represent the total influence received by each principle. The difference (R–C) classifies principles as causes (positive values) or effects (negative values), while the sum (R+C) indicates their overall prominence.

6. Causal Diagram Visualization:

Plotting each principle on a two-dimensional map—X-axis for importance (R+C) and Y-axis for causality (R–C)—to visually depict systemic relationships. The visualization was generated using the R programming language [26].

3.4. Validation and Reliability

Although Fuzzy DEMATEL relies primarily on expert judgment rather than psychometric constructs, reliability was enhanced through multiple validation strategies. First, the linguistic scale was tested for semantic consistency during the pilot phase to ensure respondents interpreted influence levels uniformly. Second, data were screened for outliers and inconsistent responses using the interquartile range method. Third, sensitivity analysis was performed by recalculating the total-relation matrix with and without extreme values to verify stability in causal direction. The results remained robust, confirming that the identified causal principles were not artifacts of specific respondent biases.

The study also employed the Borda Count method as a secondary validation tool. While Fuzzy DEMATEL captures interdependencies, Borda Count aggregates participants' direct ranking preferences. Comparing the two sets of results provided an additional layer of reliability and interpretive triangulation—confirming that principles identified as causal drivers in the DEMATEL model (Accountability, Transparency, Privacy, and Contestability) also emerged among the top-ranked in perceived importance.

3.5. Ethical Considerations and Limitations

All respondents were informed about the purpose of the research and gave consent prior to participation. No personal identifying information was collected, in accordance with the Indonesian Personal Data Protection (PDP) Law 2022. Ethical clearance was obtained from the Research Ethics Committee of Universitas Katolik Indonesia Atma Jaya (Ref. No. UAJ-ETIK/AI-2025/03).

Despite its methodological rigor, the study has several limitations. The sample was obtained through purposive and snowball sampling, which may limit representativeness across Indonesia's broader AI ecosystem. Furthermore, perceptions were self-reported and may be influenced by individual familiarity with ethical principles. Finally, the cross-sectional nature of the study captures perceptions at a single point in time, preventing longitudinal assessment of evolving attitudes toward AI ethics. Future studies should adopt mixed methods, including qualitative interviews or Delphi approaches, to deepen understanding and validate the causal relationships identified here.

4. Results

This section presents both the numerical and interpretive findings derived from the Fuzzy DEMATEL analysis and supporting demographic data. The results illustrate how the eight AI ethical principles, adapted from Australia's AI Ethics Principles, interact in a complex causal network within Indonesia's emerging AI governance context. Tables 1 through 4 summarize the quantitative outputs and respondent characteristics, supported by interpretative discussion of their implications.

4.1. Influence Scores and Causal Relationships

Table 1 presents the calculated influence scores (R, C, R+C, and R–C) for each ethical principle. The value of R (row sum) represents how much a principle influences others, while C (column sum) indicates how much it is influenced by others. The combination of R+C measures the overall prominence or importance of a principle in the system, and R–C identifies whether the principle functions as a cause (positive) or an effect (negative) within the network.

Table 1.
Influence Scores of Ethical Principles.

Principle	R+C	R–C	Category
P1 – Human, Societal and Environmental Wellbeing	68.852	-0.445	Effect
P2 – Human-Centered Values	69.580	-0.269	Effect
P3 – Fairness	70.346	-0.009	Effect (Near-Neutral)
P4 – Privacy Protection and Security	68.429	+0.014	Cause (Near-Neutral)
P5 – Reliability and Safety	67.820	-0.405	Effect
P6 – Transparency and Explainability	68.181	+0.433	Cause
P7 – Contestability	65.997	+0.164	Cause
P8 – Accountability	69.276	+0.517	Cause

The results show that Accountability (P8) has the highest causal influence (R–C = +0.517), confirming its central role in driving other ethical outcomes. Transparency and Explainability (P6) follows closely as a strong causal driver (R–C = +0.433), reinforcing its importance in ensuring both fairness and accountability. Contestability (P7) and Privacy Protection and Security (P4) also show positive R–C values, functioning as enabling structures for ethical governance. In contrast, Fairness (P3), Human-Centered Values (P2), and Wellbeing (P1) emerge as effects, meaning their achievement depends on

the effectiveness of causal principles. These results empirically demonstrate that Indonesia's ethical AI maturity depends largely on strengthening systemic drivers rather than focusing solely on outcome-based ethics.

4.2. Perceived Importance: Borda Count Results

To validate and complement the causal mapping, respondents were also asked to rank the eight principles by their perceived importance. The aggregated rankings were analyzed using the Borda Count method Table 2 where each rank position was assigned a corresponding score (7 points for first place, 0 for last).

Table 2.
Borda Count Ranking Results.

Principle	Borda Count	Rank
P4 – Privacy Protection and Security	483	1
P2 – Human-Centered Values	463	2
P3 – Fairness	427	3
P1 – Wellbeing	421	4
P5 – Reliability and Safety	372	5
P6 – Transparency and Explainability	339	6
P8 – Accountability	287	7
P7 – Contestability	260	8

The Borda Count results reflect perceptual priorities, indicating that respondents view privacy and fairness as the most pressing ethical concerns. Interestingly, these perceptions differ from the systemic influence pattern derived from Fuzzy DEMATEL. Table 3 integrates both perspectives, comparing perceived importance (Borda rank) and systemic importance (R+C).

Table 3.
Comparison Between Fuzzy Dematel and Borda Count.

Principle	R+C (System Importance)	Rank	Borda Count	Rank	Relative System Importance
P1 – Wellbeing	68.852	4	421	4	Equal
P2 – Human-Centered Values	69.580	2	463	2	Equal
P3 – Fairness	70.346	1	427	3	Higher in System
P4 – Privacy and Security	68.429	5	483	1	Lower in System
P5 – Reliability and Safety	67.820	7	372	5	Lower in System
P6 – Transparency	68.181	6	339	6	Equal
P7 – Contestability	65.997	8	260	8	Equal
P8 – Accountability	69.276	3	287	7	Higher in System

This comparison highlights an important distinction: public perception prioritizes immediate ethical concerns (privacy, fairness), while systemic causality emphasizes structural governance mechanisms (accountability, transparency). The divergence reinforces the need for balanced policymaking that addresses both citizens' concerns and institutional requirements for sustainable AI governance.

4.3. Demographic Profile of Respondents

The validity and interpretation of the causal relationships are influenced by the demographic characteristics of the 109 respondents. Tables 4 summarize their backgrounds.

Table 4.
Demographic Summary.

Characteristic	Distribution
Age	Mean = 24.28 years (Range 17–55)
Gender	Male 73.4%, Female 26.6%
Education	Mostly Bachelor's, followed by High School and Master's Degree
Field of Study	Predominantly STEM (Information Technology, AI, Computer Science)
Job Sector	AI/ML/Data, IT, Education, Finance, Consulting, Logistics, Arts, Government, Entrepreneurship
Work Experience	Majority <5 years; many students or early-career professionals
Role in AI	69.7% practitioners (developing, researching, or applying AI)
Type of AI Used	Generative AI dominant; few use reinforcement learning or automation tools
Geographic Distribution	87% based in Java Island; 13% from other regions

The demographic data indicate that respondents are mostly young, tech-oriented, and concentrated in urban centers. This profile explains why perceptions of ethical AI are heavily influenced by exposure to generative and user-facing

technologies. The dominance of respondents from Java also reflects Indonesia's digital divide, where AI literacy and infrastructure are significantly higher compared to rural areas. Moreover, the gender gap (only 26.6% female) highlights persistent disparities in AI participation—a finding consistent with global statistics, where women comprise less than 20% of AI researchers [27]. These contextual insights provide sociological depth to the quantitative findings, underscoring the need for inclusive policy design that addresses both technical ethics and social equity.

4.4. Visual Representation: Fuzzy DEMATEL Causal Diagram

The relationships presented in Tables 1–3 are illustrated in Figure 1, which plots each principle along two axes: the X-axis (R+C) representing importance and the Y-axis (R–C) indicating causal direction. Principles above the horizontal line (positive R–C) function as drivers, while those below are outcomes.
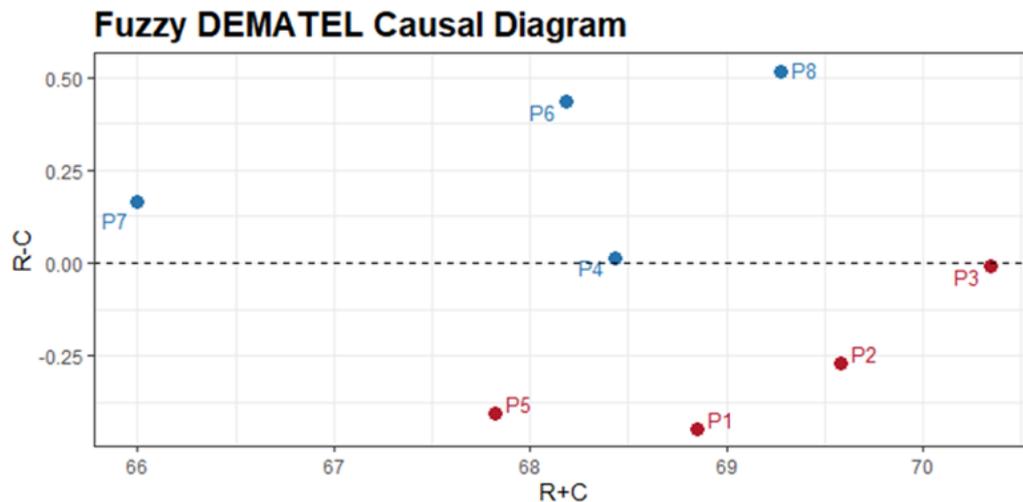

Figure 1.
Fuzzy DEMATEL Causal Diagram.

The diagram visually confirms the hierarchical configuration of ethical governance in Indonesia's AI ecosystem: Accountability and Transparency occupy the upper-right quadrant (high importance, strong causality), while Fairness, Reliability, and Wellbeing cluster as dependent outcomes. This configuration aligns with global research that conceptualizes ethical AI as a layered construct, where governance enablers precede social outcomes [28].

4.5. Synthesis of Quantitative Insights

The integration of both Fuzzy DEMATEL and Borda Count analyses provides a comprehensive picture of how ethical principles interconnect and how stakeholders perceive their importance within Indonesia's AI ecosystem. The results indicate that ethical governance in Indonesia is fundamentally structurally driven. Principles such as accountability and transparency are not merely supporting values but serve as the foundational enablers of fairness, reliability, and wellbeing. Their strong causal influence demonstrates that the ethical quality of AI systems cannot be achieved through moral awareness or compliance statements alone but must be built on institutional integrity, procedural transparency, and enforceable accountability mechanisms.

Furthermore, the findings highlight the intermediary role of privacy protection and contestability. These two principles act as bridges between governance structures and societal outcomes. Privacy protection ensures that data integrity and consent mechanisms are respected, which in turn strengthens public trust and reliability. Contestability, meanwhile, provides the procedural avenue for redress and feedback, allowing individuals or organisations to question AI-driven decisions. Together, they operationalise the ethical commitment of AI governance, transforming abstract principles into actionable procedures that enhance accountability and fairness.

The analysis also reveals that fairness, human-centred values, and wellbeing are not autonomous drivers of ethical practice but emergent results of a system where structural governance mechanisms function effectively. This insight aligns with sociotechnical theories of ethics, which argue that desirable social outcomes arise from systemic integrity rather than isolated interventions. In Indonesia's context, promoting fairness and wellbeing will therefore depend on the strength of upstream policies that institutionalise accountability and transparency across AI development, deployment, and oversight.

Another critical insight emerging from the comparative analysis is the gap between public perception and systemic reality. While stakeholders place high perceptual importance on privacy and fairness, principles that directly affect users, they tend to underestimate the enabling role of accountability and transparency. This divergence suggests that public understanding of AI ethics remains largely outcome-oriented, focused on visible harms or benefits, whereas systemic ethics requires a more sophisticated grasp of institutional governance. Bridging this gap will require deliberate public education initiatives and regulatory communication strategies to increase awareness of how structural ethics underpin visible fairness outcomes.

Finally, the demographic data provide important context for interpreting these findings. The dominance of young, urban, and technically oriented respondents explains the strong emphasis on issues such as data privacy and fairness, which

are salient in their digital experiences. However, the same demographic skew also highlights the need to expand ethical literacy and participation beyond technologically privileged groups. A more inclusive understanding of AI ethics, one that engages rural communities, women, and policy practitioners, will ensure that governance frameworks represent a broader spectrum of societal interests.

Taken together, these insights suggest a dual-layered dynamic between visible ethics, representing public concerns about fairness and privacy, and structural ethics, representing institutional mechanisms that enable sustainable governance. Strengthening the latter is essential for ensuring that ethical AI in Indonesia evolves beyond aspirational statements toward an operational model anchored in accountability, transparency, and contestability. This synthesis provides a solid empirical foundation for the subsequent policy discussion and reinforces the argument that ethical AI governance must be both systemic and participatory to achieve long-term legitimacy.

5. Discussion

The findings of this study provide a nuanced understanding of how ethical principles of artificial intelligence (AI) operate within Indonesia's sociotechnical context. Rather than functioning independently, the principles interact through a complex network of causal relationships that collectively determine the ethical maturity of AI governance. This interdependence demonstrates that responsible AI development cannot rely on a single principle—such as fairness or privacy—but instead requires an integrated governance architecture that reinforces accountability, transparency, and contestability as structural enablers of ethical practice. These insights extend the current discourse on AI ethics from a purely normative orientation toward a systems-level understanding of ethical interrelations, a shift that has been advocated by scholars such as Mäntymäki, et al. [29] and Stahl [30].

5.1. Interpreting the Causal Hierarchy of Ethical Principles

The dominance of accountability as the strongest causal driver underscores the foundational role of institutional responsibility in achieving ethical outcomes. In the Indonesian context, accountability remains largely conceptual rather than procedural. Kominfo's AI Ethics Guidelines emphasize moral obligations but provide limited enforcement mechanisms for ensuring compliance. This gap contrasts with the Australian model, where accountability is embedded through concrete measures—such as documentation of decision-making, auditability of AI systems, and assignment of lifecycle responsibility [14]. The causal prominence of accountability in this study suggests that Indonesia's policymakers must move beyond declarative ethics to establish formal oversight bodies capable of monitoring and sanctioning AI practices. Without such mechanisms, other ethical aspirations—fairness, reliability, and wellbeing—remain dependent on voluntary adherence rather than institutionalized governance.

Transparency and explainability, identified as the second strongest causal driver, are critical for enabling both accountability and fairness. The Indonesian public's increasing exposure to algorithmic decision-making—ranging from digital credit scoring to facial recognition—has heightened concerns over opacity and discrimination. Transparency, however, extends beyond the mere disclosure of information; it involves ensuring that AI systems are interpretable, explainable, and understandable by non-technical audiences [31]. This study's findings affirm that stakeholders perceive transparency not only as an ethical virtue but as a governance mechanism that empowers users and regulators to evaluate AI decisions. Institutionalizing transparency, therefore, requires mandatory disclosure standards, explainability testing, and public access to algorithmic impact assessments—practices that have proven effective in other governance contexts such as the EU's AI Act.

Contestability, while less prominent in direct perception rankings, emerged as a structurally significant enabler. Its identification as a causal principle suggests that the ability to challenge AI-driven outcomes is fundamental to maintaining human dignity and institutional trust. In Indonesia, where administrative redress mechanisms are underdeveloped, the absence of contestability has led to public skepticism toward automated decisions in recruitment, lending, and content moderation. Embedding contestability into policy frameworks would allow affected individuals or organizations to seek review and correction, aligning with principles of procedural justice [18]. Such mechanisms could take the form of independent AI ethics tribunals, ombudsman functions, or internal audit units within organizations that deploy AI. From a sociotechnical perspective, contestability not only safeguards fairness but also contributes to the continuous learning and improvement of AI systems through feedback and accountability loops.

5.2. From Declarative Ethics to Institutional Governance

One of the key contributions of this study lies in highlighting the gap between declarative ethics—the formal articulation of ethical principles—and institutional governance, which determines how those principles are enacted and enforced. In Indonesia, AI ethics currently resides in the declarative stage. The Kominfo guidelines articulate moral commitments but lack procedural guidelines for implementation, monitoring, and sanctions. This situation mirrors what Borenstein and Howard [32] describe as the "ethics-washing" problem, where organizations or governments adopt ethical language without embedding corresponding governance structures. The causal network analysis presented here empirically supports this critique by showing that unless structural drivers such as accountability, transparency, and contestability are institutionalized, ethical outcomes such as fairness and wellbeing cannot be sustained.

The findings also reinforce the sociotechnical perspective that technology ethics is inseparable from governance capacity. Indonesia's limited institutional coordination among government agencies, private sector actors, and academia hinders the translation of AI ethics into practice. The causal hierarchy identified in this study suggests a roadmap for sequencing policy interventions: first, build institutional accountability; second, establish transparency requirements; and

third, integrate contestability mechanisms. Only after these foundational elements are in place can outcome-oriented principles like fairness, reliability, and human wellbeing emerge effectively. This sequencing approach provides an evidence-based alternative to ad-hoc policymaking, aligning with international calls for context-sensitive governance models in the Global South [33].

5.3. Cultural and Contextual Dimensions of AI Ethics

While AI ethics frameworks are often presented as universal, their interpretation is deeply shaped by cultural and institutional contexts. Indonesia's collectivist culture and emphasis on community harmony influence how ethical principles are prioritized and operationalized. For instance, respondents' relatively high ranking of human-centered values and wellbeing reflects cultural tendencies to value social welfare over individual autonomy. However, this cultural orientation can also obscure accountability by diluting individual responsibility within collective structures. Hence, integrating cultural values into AI governance must be done carefully, balancing societal harmony with clear lines of responsibility and transparency. This contextualization enriches global discourse by demonstrating how ethical principles evolve when transposed from Western liberal frameworks into non-Western sociopolitical environments.

Furthermore, religion and moral philosophy play a significant role in shaping Indonesian attitudes toward technology ethics. Many respondents framed their understanding of accountability and fairness through moral and spiritual lenses rather than regulatory logic. This reflects what Meyer, et al. [34] terms "moral hybridity," where traditional ethics coexist with modern governance principles. Acknowledging these plural moral foundations is essential for designing inclusive AI governance that resonates with Indonesia's diverse moral landscape. Policymakers and AI developers should therefore engage not only with technical experts but also with ethicists, theologians, and social scientists to ensure a culturally sensitive yet globally aligned approach.

5.4. Toward a Policy-Oriented Ethical Governance Framework

The implications of these findings for AI governance in Indonesia are profound. The study demonstrates that ethical AI cannot emerge spontaneously from technological innovation but must be intentionally designed into institutions, policies, and infrastructures. Strengthening accountability requires the establishment of independent oversight mechanisms, such as a national AI Ethics Council empowered to conduct audits and issue compliance certifications. Enhancing transparency calls for mandatory AI disclosure statements, especially in high-risk sectors like finance, healthcare, and education. Operationalizing contestability involves creating accessible redress systems for users to report grievances or request explanations of AI decisions.

At the organizational level, companies and public institutions deploying AI should be encouraged or required to adopt AI assurance frameworks similar to those in Australia and the UK. These include ethical impact assessments, bias audits, and lifecycle documentation, ensuring that ethical evaluation occurs before and after deployment. Furthermore, public education initiatives are necessary to improve AI literacy, enabling citizens to understand, question, and trust algorithmic systems. Collectively, these policy directions align with the systemic relationships identified through Fuzzy DEMATEL, reinforcing the need to strengthen causal drivers to achieve sustainable ethical outcomes.

5.5. Theoretical and Practical Contributions

From a theoretical standpoint, this study advances the understanding of AI ethics as a complex adaptive system, where ethical principles interact dynamically rather than hierarchically. By integrating sociotechnical theory with the Fuzzy DEMATEL method, the research demonstrates how quantitative causal mapping can complement philosophical analysis. This hybrid approach provides a model for future studies seeking to operationalize normative ethics within empirical frameworks. Practically, the study offers an actionable roadmap for policymakers: rather than treating ethics as abstract ideals, they can prioritize interventions targeting causal drivers that exert measurable influence over the ethical system as a whole.

Overall, the discussion emphasizes that ethical AI governance is not merely about technological regulation but about fostering an ecosystem of trust, accountability, and transparency. Indonesia's challenge is to translate ethical principles into enforceable structures that reflect its social realities and policy capacities. Doing so would not only align national governance with international standards but also position Indonesia as a regional leader in responsible AI development within Southeast Asia.

6. Conclusions

This study contributes to the growing discourse on responsible and ethical artificial intelligence (AI) by empirically examining how ethical principles interact within Indonesia's emerging governance landscape. Using the Fuzzy Decision-Making Trial and Evaluation Laboratory (Fuzzy DEMATEL) method, the research mapped the causal and effect relationships among eight AI ethical principles adapted from Australia's national framework. The results demonstrate that four principles—Accountability, Transparency and Explainability, Contestability, and Privacy Protection and Security—function as foundational causal drivers, while Fairness, Human-Centered Values, Reliability and Safety, and Human, Societal, and Environmental Wellbeing serve as outcome-oriented effects. This causal hierarchy highlights that achieving fairness and wellbeing depends on strengthening the structural conditions of governance, particularly accountability and transparency. In this sense, ethical AI in Indonesia cannot be achieved through declarative codes or voluntary compliance alone but requires systematic and institutionalized governance mechanisms.

Conceptually, the findings advance the understanding of AI ethics as an interdependent and adaptive system rather than a linear set of moral guidelines. The Fuzzy DEMATEL analysis reveals that ethical principles operate dynamically—where improvements in one area, such as accountability, can trigger positive ripple effects across fairness, privacy, and wellbeing. This systemic insight moves beyond traditional ethical checklists and provides a more holistic model for understanding how sociotechnical and institutional factors shape ethical outcomes. It also bridges a critical gap in AI ethics research by integrating quantitative causal modeling with normative ethical theory, demonstrating that ethical values can be operationalized without losing philosophical depth. This hybrid methodological contribution strengthens the empirical grounding of AI ethics and opens pathways for evidence-based policymaking.

From a policy perspective, the study offers practical recommendations for designing a more coherent and enforceable AI governance framework in Indonesia. First, establishing institutional accountability mechanisms—such as a national AI Ethics Council or cross-ministerial oversight body—would ensure that ethical compliance is monitored throughout the AI lifecycle. Second, implementing transparency and explainability standards would help build public trust by enabling citizens, regulators, and developers to understand how AI systems make decisions. Third, embedding contestability mechanisms would operationalize procedural justice, granting individuals and organizations the right to question or appeal AI-driven outcomes. These structural interventions would translate ethical ideals into measurable governance practices, ensuring that principles such as fairness, human-centeredness, and wellbeing emerge as tangible results rather than rhetorical aspirations.

The study also underscores the importance of contextualizing global ethical frameworks to fit local realities. While international principles—such as those from Australia, the OECD, and UNESCO—provide valuable benchmarks, their successful implementation in Indonesia depends on alignment with the nation's regulatory capacity, cultural values, and socio-political conditions. For instance, Indonesia's collectivist ethos and emphasis on moral integrity can strengthen public support for ethical governance but must be complemented with institutional structures that ensure transparency and accountability. Thus, effective AI governance in Indonesia must harmonize universal ethical principles with local institutional and cultural dynamics. This approach not only reinforces ethical legitimacy but also positions Indonesia as a regional model for context-sensitive AI governance in the Global South.

Despite its contributions, this study is not without limitations. The use of a perception-based survey introduces subjectivity and may not fully capture the diversity of experiences within Indonesia's AI ecosystem. Moreover, the sample of 109 respondents, while sufficient for exploratory analysis, may not represent all stakeholder groups, particularly policymakers, rural communities, and marginalized users. Future research should therefore adopt mixed-method or longitudinal designs to validate these causal relationships across sectors and time. Qualitative approaches, such as interviews or focus groups, could enrich understanding of how stakeholders interpret and negotiate ethical principles in practice. Comparative studies between Indonesia and other Southeast Asian nations could also provide valuable insights into regional variations in AI governance maturity.

In conclusion, this research provides both theoretical and practical foundations for advancing ethical AI governance in Indonesia. It demonstrates that accountability, transparency, contestability, and privacy protection are not only moral imperatives but also functional preconditions for sustaining fairness, reliability, and societal wellbeing. Strengthening these causal drivers offers a strategic pathway for policymakers to transition from aspirational ethics to operational governance. More broadly, the study contributes to the international dialogue on AI ethics by showing that empirical methods such as Fuzzy DEMATEL can complement ethical reasoning, providing actionable insights for policymakers and scholars alike. Ultimately, ethical AI in Indonesia—and globally—will depend not only on the articulation of principles but on the creation of robust, transparent, and participatory governance systems that ensure technology serves humanity with integrity and justice.

References

[1]   A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence,* vol. 1, no. 9, pp. 389-399, 2019. https://doi.org/10.1038/s42256-019-0088-2

[2]   L. Floridi and J. Cowls, "A unified framework of five principles for AI in society," *Machine Learning and the City: Applications in Architecture and Urban Design,* pp. 535-545, 2022. https://doi.org/10.1002/9781119815075.ch45

[3]   W. K. Parinsi, A. D. L. Musa, and K. S. P. Musa, "Ethical use of AI in continuous recruitment: An analysis of algorithm bias towards candidates from marginalized backgrounds," *Journal Management & Economics Review,* vol. 2, no. 7, pp. 215-220, 2025.

[4]   F. Nugroho, "Artificial intelligence regulation and political ethics: An analysis of Indonesia's Position in AI Governance," *Journal of Political Innovation and Analysis,* vol. 2, no. 1, pp. 42 51-42 51, 2025. https://doi.org/10.59261/jpia.v2i1.10

[5]   C. Sanderson *et al.*, "AI ethics principles in practice: Perspectives of designers and developers," *IEEE Transactions on Technology and Society,* vol. 4, no. 2, pp. 171-187, 2023. https://doi.org/10.1109/TTS.2023.3257303

[6]   M. Ryan and B. C. Stahl, "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications," *Journal of Information, Communication and Ethics in Society,* vol. 19, no. 1, pp. 61-86, 2021.

[7]   J. Zhou, F. Chen, A. Berry, M. Reed, S. Zhang, and S. Savage, "A survey on ethical principles of AI and implementations," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020: IEEE, pp. 3010-3017.

[8]   N. Kadoić, B. Divjak, and N. Begičević Ređep, "Integrating the DEMATEL with the analytic network process for effective decision-making," *Central European Journal of Operations Research,* vol. 27, no. 3, pp. 653-678, 2019.

[9]   X. Chen, "Ethical governance of AI: An integrated approach via human-in-the-loop machine learning," *Computer Sciences & Mathematics Forum,* vol. 8, no. 1, pp. 29-32, 2023.

[10]   P. Radanliev, "AI ethics: Integrating transparency, fairness, and privacy in AI development," *Applied Artificial Intelligence,* vol. 39, no. 1, pp. 246-259, 2025. https://doi.org/10.1080/08839514.2025.2463722

[11] B. Vaassen, "AI, opacity, and personal autonomy," *Philosophy & Technology,* vol. 35, no. 4, pp. 88-100, 2022. https://doi.org/10.1007/s13347-022-00577-5

[12] P. Akbarighatar, "Operationalizing responsible AI principles through responsible AI capabilities," *AI and Ethics,* vol. 5, no. 2, pp. 1787-1801, 2025.

[13] A. Batool, D. Zowghi, and M. Bano, "AI governance: A systematic literature review," *AI and Ethics,* vol. 5, no. 3, pp. 1-15, 2025.

[14] D. Lindsay and J. Hogan, "An Australian perspective on AI, ethics and its regulatory challenges," *Journal of Law and Economic Regulation,* vol. 12, no. 2, pp. 116-127, 2019.

[15] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi, "Operationalising AI ethics: Barriers, enablers and next steps," *AI & SOCIETY,* pp. 1-13, 2023. https://doi.org/10.1007/s00146-021-01308-8

[16] R. P. Wadipalapa *et al.*, "An ambitious artificial intelligence policy in a decentralised governance system: Evidence from Indonesia," *Journal of Current Southeast Asian Affairs,* vol. 43, no. 1, pp. 65-93, 2024. https://doi.org/10.1177/18681034231226393

[17] P. Akbarighatar, I. Pappas, and P. Vassilakopoulou, "A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review," *International Journal of Information Management Data Insights,* vol. 3, no. 2, pp. 100-113, 2023.

[18] O. Kudina and I. van de Poel, "A sociotechnical system perspective on AI," *Minds and Machines,* vol. 34, no. 3, pp. 21-34, 2024. https://doi.org/10.1007/s11023-024-09680-2

[19] E. Fontela and A. Gabus, *Current perceptions of the world problematique. World modeling: A dialogue*. Amsterdam/Oxford: North-Holland Publishing Company, 1976.

[20] J. Kacprzyk, "Web intelligence, business intelligence and decision support systems: A challenge for fuzzy logic and soft computing," presented at the International IEEE Conference Intelligent Systems, IS'06, September 4, 2006 - September 6, 2006, 2006. [Online]. Available: http://dx.doi.org/10.1109/IS.2006.348382.

[21] R. Priyanka, K. Ravindran, B. Sankaranarayanan, and S. M. Ali, "A fuzzy DEMATEL decision modeling framework for identifying key human resources challenges in start-up companies: Implications for sustainable development," *Decision Analytics Journal,* vol. 6, p. 100192, 2023. https://doi.org/10.1016/j.dajour.2023.100192

[22] B. Chang, C.-W. Chang, and C.-H. Wu, "Fuzzy DEMATEL method for developing supplier selection criteria," *Expert systems with Applications,* vol. 38, no. 3, pp. 1850-1858, 2011. https://doi.org/10.1016/j.eswa.2010.07.114

[23] M.-B. Fakhrzad, M.-R. Firozpour, and H. H. Nasab, "A fuzzy DEMATEL approach for evaluating the risk factors-(Case study: The instruction center of Iran's national oil company)," *RAIRO-Operations Research,* vol. 55, no. 4, pp. 2525-2543, 2021.

[24] M. S. Ahmad, W. Fei, M. Shoaib, and H. Ali, "Identification of key drivers for performance measurement in sustainable humanitarian relief logistics: an Integrated fuzzy Delphi-DEMATEL approach," *Sustainability,* vol. 16, no. 11, pp. 44-59, 2024. https://doi.org/10.3390/su16114459

[25] S. Opricovic and G.-H. Tzeng, "Defuzzification within a multicriteria decision model," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 11, no. 05, pp. 635-652, 2003.

[26] RStudio, "RStudio: Integrated development environment for R (Version 2024) [Software]. RStudio, PBC," 2024. https://www.rstudio.com/

[27] S. M. West, M. Whittaker, and K. Crawford, "Discriminating systems," *AI Now,* vol. 2019, pp. 1-33, 2019.

[28] E. Papagiannidis, P. Mikalef, and K. Conboy, "Responsible artificial intelligence governance: A review and research framework," *The Journal of Strategic Information Systems,* vol. 34, no. 2, pp. 100-115, 2025. https://doi.org/10.1016/j.jsis.2024.101885

[29] M. Mäntymäki, M. Minkkinen, T. Birkstedt, and M. Viljanen, "Defining organizational AI governance," *AI and Ethics,* vol. 2, no. 4, pp. 603-609, 2022. https://doi.org/10.1007/s43681-022-00143-x

[30] B. C. Stahl, "Embedding responsibility in intelligent systems: From AI ethics to responsible AI ecosystems," *Scientific Reports,* vol. 13, no. 1, pp. 75-86, 2023. https://doi.org/10.1038/s41598-023-34622-w

[31] B. Mittelstadt, C. Russell, and S. Wachter, "Explaining explanations in AI," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 279-288.

[32] J. Borenstein and A. Howard, "Emerging challenges in AI and the need for AI ethics education," *AI and Ethics,* vol. 1, no. 1, pp. 61-65, 2021. https://doi.org/10.1007/s43681-020-00002-7

[33] S. Farhad, "Passengers in flight: AI governance capacity in the global South," *Digital Society,* vol. 4, no. 2, pp. 39-53, 2025.

[34] S. Meyer, S. Mandl, D. Gesmann-Nuissl, and A. Strobel, "Responsibility in hybrid societies: Concepts and terms," *AI and Ethics,* vol. 3, no. 1, pp. 25-48, 2023.