

# ARTIFICIAL INTELLIGENCE IN EDUCATIONAL GOVERNANCE: A RISK-BASED ETHICAL FRAMEWORK FOR STRENGTHENING RELIGIOUS MODERATION IN INDONESIA

<sup>1,2</sup>Ahmad Imam Khairi, <sup>3</sup>Sugiantoro, <sup>4</sup>Kacung Wahyudi, <sup>5</sup>Itaanis Tianah, and <sup>6</sup>Salman Al Farisi

<sup>1,3</sup>Universitas Negeri Surabaya, <sup>2,4,5,6</sup>Universitas Islam Negeri Madura  
Email: <sup>1</sup>ahmad.23001@mhs.unesa.ac.id <sup>2</sup>khairi.ahmadimam@gmail.com, <sup>3</sup>sugiantoro@unesa.ac.id, <sup>4</sup>kacungwahyudi@iainmadura.ac.id <sup>5</sup>ita@iainamdura.ac.id, <sup>6</sup>alfarisisalman552@gmail.com

## Abstrak

*Transformasi digital pendidikan di Indonesia telah bergeser dari sekadar digitalisasi layanan menuju tata kelola berbasis data yang semakin bergantung pada kecerdasan buatan. Integrasi kecerdasan buatan dalam tata kelola pendidikan menjanjikan efisiensi administratif dan optimalisasi pengambilan keputusan, namun sekaligus menimbulkan persoalan etis terkait bias algoritmik, akuntabilitas, transparansi, serta potensi reduksi nilai-nilai humanistik. Penelitian ini bertujuan menganalisis secara kritis implikasi etis penggunaan kecerdasan buatan dalam tata kelola pendidikan di Indonesia serta relevansinya terhadap penguatan moderasi beragama sebagai agenda strategis pendidikan nasional. Penelitian menggunakan pendekatan kualitatif dengan desain kajian literatur kritis dan analisis kebijakan. Data diperoleh dari artikel ilmiah bereputasi, dokumen kebijakan internasional, serta regulasi nasional terkait transformasi digital dan moderasi beragama. Analisis dilakukan melalui tahapan identifikasi tema, evaluasi normatif berbasis prinsip keadilan algoritmik dan tata kelola berpusat pada manusia, serta sintesis kontekstual dengan lanskap sosial-keagamaan Indonesia. Hasil penelitian menunjukkan bahwa tanpa kerangka etis yang kuat, tata kelola pendidikan berbasis kecerdasan buatan berpotensi mereproduksi ketimpangan struktural dan mempersempit ruang dialog plural. Studi ini merumuskan model konseptual tata kelola etis berbasis moderasi beragama yang menekankan pemetaan risiko, desain etis, pengawasan manusiawi, akuntabilitas publik, dan evaluasi berkelanjutan. Model ini diharapkan menjadi rujukan normatif bagi pembuat kebijakan dan institusi pendidikan dalam memastikan transformasi digital tetap selaras dengan prinsip keadilan, inklusivitas, dan kohesi sosial.*

**Kata Kunci:** Kecerdasan buatan, Tata kelola pendidikan, Etika algoritmik, Moderasi beragama, Kebijakan publik

## Abstract

*The digital transformation of education in Indonesia has evolved from basic service digitization toward data-driven governance increasingly reliant on artificial intelligence. While the integration of artificial intelligence in educational governance promises administrative efficiency and optimized decision-making, it simultaneously raises ethical concerns related to algorithmic bias, accountability, transparency, and the potential erosion of humanistic educational values. This study aims to critically analyze the ethical implications of artificial intelligence deployment in Indonesian educational governance and examine its relevance to strengthening religious moderation as a strategic national education agenda. The research employs a qualitative approach using a critical literature review design combined with policy analysis. Data were drawn from reputable scholarly publications, international policy frameworks, and national regulations concerning digital transformation and religious moderation. The analysis proceeded through thematic identification, normative evaluation grounded in principles of algorithmic justice and human-centered governance, and contextual synthesis within Indonesia's plural socio-religious landscape. The findings indicate that without a robust ethical framework, artificial intelligence-driven governance risks reproducing structural inequalities and narrowing pluralistic dialogue. This study proposes a conceptual ethical governance model aligned with religious moderation, emphasizing risk mapping, ethical design protocols, meaningful human oversight, public accountability mechanisms, and continuous evaluation. The model offers a normative and operational reference for policymakers and educational institutions to ensure that digital transformation remains aligned with justice, inclusivity, and social cohesion.*

**Key words:** Artificial intelligence, Educational governance, Algorithmic ethics, Religious moderation, Public policy

## INTRODUCTION

The digital transformation of education in Indonesia has evolved from the mere digitization of services toward data driven governance increasingly dependent on computational systems, including artificial intelligence (AI). In practice, AI is being deployed or at least articulated as a policy aspiration for learning analytics, automated assessment, early warning systems for dropout risks, intervention recommendation engines, and strategic resource planning. While these developments promise greater efficiency and institutional responsiveness, they simultaneously shift the locus of decision-making from pedagogical-humanistic judgment toward algorithmic optimization logics. At the global level, UNESCO has underscored that the integration of AI in education must be guided by a humanistic vision that safeguards human dignity, equity, inclusion, and cultural-linguistic diversity, warning that without robust ethical guardrails, technological systems risk exacerbating inequalities and eroding human agency.<sup>1</sup> A comparable normative stance is advanced by the OECD through its framework on trustworthy AI, which calls for adherence to human rights, transparency, accountability, and systemic robustness in AI governance.<sup>2</sup>

Within the framework of educational governance, this shift raises a fundamental question: when educational decisions are mediated by predictive models, who bears responsibility for their consequences the institution, the developers, or the “model” itself? The literature on algorithmic governance suggests that software systems, databases, and analytic infrastructures can function as modalities of power, governing through processes of classification, prediction, and the automation of administrative action, often with limited space for public deliberation.<sup>3</sup> Scholarship on algorithmic accountability further underscores the necessity of transparency and robust auditing mechanisms to ensure that high-stakes public-sector decisions do not devolve into opaque “black boxes” insulated from scrutiny and correction.<sup>4</sup> When historical patterns of inequality are embedded within training data, AI systems risk reproducing and amplifying systemic discrimination an issue widely documented in ethical critiques of algorithms that identify bias, opacity, and the broader social consequences of data-driven decision-making.<sup>5</sup> As well as in more

---

<sup>1</sup> UNESCO, “Recommendation on the Ethics of Artificial Intelligence” (Paris: UNESCO Publishing, 2022), <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>; UNESCO, “Guidance for Generative AI in Education and Research” (Paris: UNESCO Publishing, 2023), <https://doi.org/10.54675/ewzm9535>.

<sup>2</sup> OECD, “What Are the OECD Principles on AI?” (Paris: OECD Publishing, 2024), <https://legalinstruments.oecd.org/en/>.

<sup>3</sup> Ben Williamson, “Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education,” *Learning, Media and Technology* 40, no. 1 (2015): 83–105, <https://doi.org/10.1080/17439884.2014.924527>.

<sup>4</sup> Harshad Shah, “Algorithmic Accountability,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (2018), <https://doi.org/10.1098/rsta.2017.0362>; Michael Veale, Max Van Kleek, and Reuben Binns, “Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making,” *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, <https://doi.org/10.1145/3173574.3174014>.

<sup>5</sup> Brent D Mittelstadt et al., “The Ethics of Algorithms: Mapping the Debate,” *Big Data & Society* 3, no. 2 (2016), <https://doi.org/10.1177/2053951716679679>.

publicly oriented critiques warning of opaque and potentially harmful models that disproportionately disadvantage vulnerable populations.<sup>6</sup>

Within the field of education, AI-related discourse has coalesced around two dominant currents: functional optimism and ethical critique. Educational scholarship highlights the potential of AI to enable personalized learning pathways and enhanced pedagogical support; however, it simultaneously calls for caution, as these transformations implicate the professional role of teachers, the social dynamics of the classroom, and the normative foundations upon which educational practice is built.<sup>7</sup> Critical editorials on AI in education further caution that the “hyperbolic” rhetoric of innovation often obscures the longer historical trajectory of datafication and the expanding automation of governance. Such analyses advocate for a more reflexive interrogation of the assumptions, actors, and institutional interests embedded within educational platforms and infrastructures.<sup>8</sup> At the same time, global mappings of AI ethics guidelines reveal a convergence around core principles transparency, fairness, non-maleficence, responsibility, and privacy. Yet their operationalization frequently remains indeterminate and non-binding, raising the concern that these principles risk functioning as normative slogans rather than as enforceable mechanisms of governance.<sup>9</sup>

Indonesia’s distinctiveness lies in the fact that education functions not merely as an instrument for enhancing economic capacity, but also as a formative arena for cultivating national character and fostering socio-religious harmony. Within this framework, religious moderation has been positioned as a strategic agenda aimed at strengthening tolerance, non-violence, and acceptance of diversity. In the digital sphere, mainstreaming religious moderation is considered particularly urgent, given the susceptibility of online information ecosystems to polarization. Islamic higher education institutions, in particular, are expected to serve as counter-narrative agents through the reinforcement of digital content production and digital literacy initiatives.<sup>10</sup> Other studies examining the internalization of moderation values within Islamic universities emphasize the need for institutional strategies to ensure that moderation transcends rhetorical commitment and becomes embedded in concrete educational practice.<sup>11</sup> However, as educational

---

<sup>6</sup> Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

<sup>7</sup> Neil Selwyn, *Should Robots Replace Teachers? AI and the Future of Education* (Cambridge: Polity Press, 2019); Wayne Holmes, Maya Bialik, and Charles Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning* (Boston: Center for Curriculum Redesign, 2019).

<sup>8</sup> Ben Williamson and Rebecca Eynon, “Historical Threads, Missing Links, and Future Directions in AI in Education,” *Learning, Media and Technology* 45, no. 3 (2020): 223–35, <https://doi.org/10.1080/17439884.2020.1798995>.

<sup>9</sup> Luciano Floridi et al., “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations,” *Minds and Machines* 28, no. 4 (2018): 689–707, <https://doi.org/10.1007/s11023-018-9482-5>; Anna Jobin, Marcello Ienca, and Effy Vayena, “The Global Landscape of AI Ethics Guidelines,” *Nature Machine Intelligence* 1, no. 9 (2019): 389–99, <https://doi.org/10.1038/s42256-019-0088-2>.

<sup>10</sup> Wildani Hefni, “Moderasi Beragama Dalam Ruang Digital: Studi Pengarusutamaan Moderasi Beragama Di Perguruan Tinggi Keagamaan Islam Negeri,” *Jurnal Bimas Islam* 13, no. 1 (2020): 1–22, <https://api.semanticscholar.org/CorpusID:225547190>.

<sup>11</sup> Muhammad Nasir and Muhammad Khoirul Rijal, “Keeping the Middle Path: Mainstreaming Religious Moderation through Islamic Higher Education Institutions in Indonesia,” *Indonesian Journal of Islam and Muslim Societies* 11, no. 2 (2021): 213–41, <https://doi.org/10.18326/ijims.v11i2>.

governance becomes increasingly AI driven through mechanisms such as institutional ranking systems, content recommendation algorithms, or the detection of “risk” indicators the normative substance of religious moderation risks being reduced to a mere administrative variable. In such contexts, complex ethical and theological commitments may be oversimplified, rendered into binary classifications, or even distorted by latent biases embedded in training data and algorithmic design.

A review of the state of the art over the past decade indicates substantial advances in scholarship on algorithmic ethics and accountability,<sup>12</sup> software-mediated educational governance and datafication,<sup>13</sup> the implications of AI for educational practice,<sup>14</sup> global normative frameworks,<sup>15</sup> and comparative mappings of AI ethics guidelines.<sup>16</sup> In parallel, research on religious moderation has examined strategies for value internalization and the strengthening of moderation within digital environments.<sup>17</sup> Despite these developments, the respective bodies of literature have largely evolved in isolation. Studies of AI governance seldom interrogate the implications of algorithmic systems for the agenda of religious moderation, while scholarship on religious moderation has yet to systematically articulate ethical-policy instruments capable of addressing AI-mediated educational governance particularly at the levels of system design, accountability architecture, and bias auditing.

The central research gap addressed in this study lies in the absence of an ethical-analytical framework that explicitly connects three critical nodes: (1) AI as an instrument of educational governance rather than merely a pedagogical tool; (2) the ethical risks associated with bias, opacity, surveillance, and the delegation of decision-making authority within data-driven education policy; and (3) the normative implications of these dynamics for strengthening religious moderation as a foundational objective of national education. This gap is consequential. In the absence of a contextually grounded ethical framework, AI systems risk reinforcing entrenched patterns of inequality, constraining plural forms of expression, or generating policy recommendations that appear procedurally “neutral” while in fact embedding unexamined value preferences. Such risks underscore the need for a more integrative analytical approach capable of situating technological governance within the broader moral and civic aims of education.

The novelty of this study resides in the development of a critical-ethical reading that integrates theories of algorithmic governance and global AI ethics frameworks with

---

<sup>12</sup>Mittelstadt et al., “The Ethics of Algorithms: Mapping the Debate”; Shah, “Algorithmic Accountability”; Veale, Van Kleek, and Binns, “Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making.”

<sup>13</sup>Williamson, “Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education”; Williamson and Eynon, “Historical Threads, Missing Links, and Future Directions in AI in Education.”

<sup>14</sup>Holmes, Bialik, and Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*; Selwyn, *Should Robots Replace Teachers? AI and the Future of Education*.

<sup>15</sup>OECD, “What Are the OECD Principles on AI?”; UNESCO, “Recommendation on the Ethics of Artificial Intelligence”; UNESCO, “Guidance for Generative AI in Education and Research.”

<sup>16</sup>Floridi et al., “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”; Jobin, Ienca, and Vayena, “The Global Landscape of AI Ethics Guidelines.”

<sup>17</sup>Hefni, “Moderasi Beragama Dalam Ruang Digital: Studi Pengarusutamaan Moderasi Beragama Di Perguruan Tinggi Keagamaan Islam Negeri”; Nasir and Rijal, “Keeping the Middle Path: Mainstreaming Religious Moderation through Islamic Higher Education Institutions in Indonesia.”

the Indonesian context, in which religious moderation is positioned as a strategic objective of national education. Rather than confining the discussion to “AI for learning,” this research conceptualizes AI as a policy infrastructure an architecture of governance that shapes who is evaluated, who gains access to services, which forms of knowledge are prioritized, and which narratives are amplified or marginalized within the educational ecosystem.

This study aims to critically analyze the ethical implications of AI deployment in educational governance in Indonesia, to explicate the principal risks that may affect the strengthening of religious moderation, and to formulate operational ethical principles that can serve as normative guideposts for policymakers and educational institutions. These principles are intended to ensure that AI implementation remains aligned with justice, inclusivity, transparency, and public accountability within the broader mandate of national education.

## METHOD

This qualitative study employs critical literature review and policy analysis to examine normative constructions and ethical frameworks surrounding artificial intelligence in educational governance. Rather than testing causal relationships, the research interpretively uncovers meanings, value commitments, and underlying assumptions informing policy orientations.<sup>18</sup> The critical review interrogates epistemological assumptions, normative positions, and power implications embedded within literature on AI ethics, algorithmic governance, and religious moderation in Indonesian and global contexts.<sup>19</sup> Concurrently, policy analysis assesses alignment between AI ethics principles and Indonesia's educational regulatory framework, foregrounding values, actors, and policy consequences to evaluate how problem representations shape educational outcomes.<sup>20</sup>

Data sources comprise three categories: peer-reviewed articles from international and nationally accredited journals addressing AI ethics and algorithmic governance; international policy documents including ethical recommendations and AI-in-education guidelines; and national policies on educational digital transformation and religious moderation. Systematic database searches employed keywords: artificial intelligence ethics, algorithmic governance in education, educational data governance, religious moderation, and digital transformation. Literature selection followed thematic relevance and publication credibility criteria, with subsequent thematic classification. Selected articles underwent analysis of publication context, methodology, findings, and ethical implications, guided by transparency and replicability principles to ensure methodological rigor.<sup>21</sup>

Data analysis proceeded through three sequential stages. First, thematic analysis identified central issues: algorithmic bias, accountability, transparency, digital

---

<sup>18</sup>John W. Creswell, *Research Design: Pendekatan Metode Kualitatif, Kuantitatif, Dan Campuran*, 4 Terj. ol (Yogyakarta: Pustaka Pelajar, 2016); Lexy J Moleong and D M A Anggoro, *Metodologi Penelitian Kualitatif*, ed. Edisi Revisi (Bandung: PT Remaja Rosdakarya, 2019).

<sup>19</sup>Hannah Snyder, “Literature Review as a Research Methodology: An Overview and Guidelines,” *Journal of Business Research* 104, no. March (2019): 333–39, <https://doi.org/10.1016/j.jbusres.2019.07.039>.

<sup>20</sup>Carol Bacchi, *Analysing Policy: What's the Problem Represented to Be?* (Frenchs Forest, NSW: Pearson, 2009).

<sup>21</sup>Snyder, “Literature Review as a Research Methodology: An Overview and Guidelines.”

surveillance, and implications for educational values. Second, normative analysis applied technology ethics frameworks emphasizing justice, non-discrimination, responsibility, and human dignity protection.<sup>22</sup> Third, contextual synthesis linked global findings to Indonesia's socio-religious landscape, particularly religious moderation as a national education policy agenda. Conceptual validity employed source triangulation, comparing academic literature, policy documents, and international organization reports. A reflexive reading strategy avoided technological reductionism and determinism, situating AI-driven educational governance within broader social, political, and cultural dimensions. This design generates an ethically robust, contextually grounded framework capable of informing policymaking that upholds justice, inclusivity, and religious moderation strengthening.

## RESULT AND DISCUSSION

### Result

This study employs a critical literature review and policy analysis to examine the ethical implications of AI in educational governance. The findings, synthesized from peer-reviewed scholarship and policy documents, are presented thematically in tabular form to enhance clarity and analytical precision. Each table summarizes core insights regarding algorithmic rationality, ethical risks, the MEKAI-MB policy model, and managerial implications.

*Artificial Intelligence in Educational Governance: Transformation and Algorithmic Rationality*

**Table 1.** Classification of Data Sources Used in the Study

Category of Data Source	Type of Document	Main Authors / Institutions Referenced in the Manuscript	Thematic Focus	Relevance to Study
<b>Peer-Reviewed Scholarly Articles (Last 10 Years)</b>	International journal articles	Bruno Williamson; Neil Selwyn; Bernd Mittelstadt; Luciano Floridi; Hanna Shah; Michael Veale; Arianna Rouvroy; Rob Kitchin; Wayne Holmes	AI ethics, algorithmic governance, datafication, accountability, bias, AI in education	Provides theoretical and analytical foundation for ethical risks, algorithmic bias, transparency, and governance transformation
<b>National Scholarly Articles (Indonesia)</b>	Nationally accredited journal articles	Wahyudin Hefni; Muhammad Nasir	Religious moderation, digital religious discourse, Islamic higher education	Contextualizes religious moderation within Indonesian educational institutions
<b>International Policy Documents</b>	Ethical frameworks	UNESCO (Recommendation on the Ethics of AI;	Human rights, AI ethics principles, risk-based	Serves as global normative benchmark for

<sup>22</sup>Floridi et al., “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations”; UNESCO, “Recommendation on the Ethics of Artificial Intelligence.”

	and regulatory guidelines	Guidance for Generative AI in Education); OECD (OECD Principles on AI); European Commission (AI Act Proposal)	regulation, public sector accountability	ethical AI governance
<b>National Policy Documents (Indonesia)</b>	National education and governance policy documents	Ministry-level regulations and national strategies on digital transformation of education and strengthening religious moderation	Digital transformation, governance reform, moderation mainstreaming	Aligns global AI ethics principles with Indonesian socio-cultural and regulatory context
<b>Database &amp; Search Strategy</b>	Academic databases and systematic search	Scopus, Web of Science, Google Scholar, nationally accredited journal portals	Keywords: artificial intelligence ethics; algorithmic governance in education; educational data governance; religious moderation in education; digital transformation	Ensures transparency, thematic relevance, and credibility of source selection

*Source: data processed by the author*

**Table 2.** Mapping Ethical and Governance Issues of Artificial Intelligence in Education

<b>Main Theme</b>	<b>Issue Description</b>	<b>Manifestation in Educational Governance</b>	<b>Impact on Educational Values</b>	<b>Policy Implications</b>
Algorithmic Bias	The non-neutrality of AI systems resulting from biased training data, model design, or developers' epistemic assumptions	Student risk prediction systems, academic performance classification, and ideology-related detection based on historical data	Indirect discrimination, reduction of identity complexity, and reproduction of social inequalities	Independent algorithmic audits, data provenance documentation, pre-implementation fairness testing
Accountability	Ambiguity regarding the locus of responsibility for AI-based decisions	Administrative or evaluative decisions relying on system recommendations without correction mechanisms	Weakening of institutional moral responsibility and erosion of professional human authority	Formal appeal mechanisms, establishment of technology ethics committees, mandatory human oversight
Transparency	Algorithmic opacity (black box problem)	Absence of clear explanations regarding system decision logic for stakeholders	Decline in public trust and reduced policy legitimacy	Mandatory explainability standards, publication of AI governance policies in accessible language

Digital Surveillance	Expansion of behavioral monitoring through learning analytics and risk detection systems	Monitoring students' online activities and behavioral profiling	Risk of privacy violations and normalization of control culture	Data protection regulations, restrictions on sensitive data use, informed consent requirements
Humanistic Reductionism	Dominance of technocratic rationality over pedagogical judgment	Decision-making based solely on quantitative scores and performance indicators	Weakening of dialogue, empathy, and pedagogical relationships	Reinforcement of human-centered governance principles, value-based evaluation frameworks
Polarization and Echo Chambers	Data-driven content personalization that narrows exposure to diverse perspectives	Recommendation systems reinforcing specific ideological preferences	Reduced exposure to plural viewpoints and risk of social fragmentation	Religious moderation alignment review, diversification of learning content
Digital Access Inequality	Disparities in technological infrastructure and digital literacy across regions and institutions	Resource-limited institutions lagging in AI adoption	Widening educational quality gaps	Affirmative policies and equitable resource redistribution

*Source: Thematic synthesis conducted by the author based on literature on AI ethics and educational governance (2020–2024)*

### ***Ethical Risks and Their Implications for Religious Moderation***

**Table 3.** Risk Stratification Matrix for AI Implementation in Educational Governance

<b>Risk Level</b>	<b>AI System Characteristics</b>	<b>Examples of Implementation in Education</b>	<b>Potential Social Impact</b>	<b>Governance Requirements</b>
<b>Low Risk</b>	Administrative support systems without direct impact on individual identity or evaluation	Archival automation, schedule management, aggregate statistical processing	Minimal impact on individual rights	Basic system documentation, internal supervision, periodic technical evaluation
<b>Moderate Risk</b>	Systems that affect academic processes but do not determine social status or identity	Learning material recommendation systems, general performance analytics	Potential bias in recommendations and material access	Algorithmic bias testing, model transparency, limited public reporting
<b>High Risk</b>	Systems that affect individual classification, character evaluation, or ideological assessment	Digital-behavior-based dropout risk prediction, ideological tendency detection	Indirect discrimination, social labeling, polarization	Independent audit, mandatory human oversight, appeal mechanisms, evaluation of alignment with religious moderation
<b>Very High Risk</b>	Systems impacting fundamental rights, freedom of thought,	Online-data-based radicalism mapping systems without	Civil rights violations, stigmatization,	Temporary ban or moratorium until comprehensive ethical evaluation,

or individual social reputation	contextual verification	erosion of public trust	national regulatory oversight
------------------------------------	----------------------------	----------------------------	----------------------------------

Source: Author's conceptual synthesis based on literature on public sector AI ethics and risk-based regulation frameworks (2020–2024).

## Discussion

### *Artificial Intelligence in Educational Governance: Transformation and Algorithmic Rationality*

The digital transformation of education does not merely introduce new tools; it reconfigures the ways in which knowledge is produced, distributed, and validated. Within the framework of critical data studies, data are not neutral entities but socially constructed artifacts embedded with epistemological assumptions and power relations.<sup>23</sup> When artificial intelligence is deployed in educational governance, it contributes to the formation of new “regimes of truth” grounded in metrics, quantification, and predictive analytics.

Williamson<sup>24</sup> demonstrates that software systems and databases operate as policy infrastructures that implicitly define standards of educational success. In this configuration, algorithms do not merely analyze data; they actively shape institutional practice through performance indicators and risk classifications. As a consequence, qualitatively grounded educational values such as tolerance, dialogical engagement, and moderation resist reduction to numerical metrics without sacrificing the complexity and depth of their meaning. This phenomenon is reinforced by the growing trajectory of data-driven governance, which positions prediction as a primary source of policy legitimacy.<sup>25</sup> In educational settings, predictive models may shape the prioritization of interventions, the allocation of financial assistance, and even the monitoring of students’ online behavior.

From the standpoint of religious moderation, this question becomes particularly salient. If analytic systems are deployed to detect ideological tendencies or assess risks of radicalization without due consideration of social and pedagogical context, AI-driven governance may inadvertently evolve into a mechanism of control one that undermines the very values of dialogue, inclusivity, and mutual recognition that moderation seeks to cultivate.

The findings indicate that the integration of artificial intelligence into educational governance has extended beyond pedagogical functions into the domains of administrative management and strategic decision-making. Predictive analytics systems, automated performance evaluations, and data-driven risk mapping have become embedded within contemporary governance practices. Conceptually, this shift marks a transition from procedurally grounded governance toward datafied governance, in which decisions are no longer entirely deliberative but are increasingly mediated by computational models.<sup>26</sup>

<sup>23</sup>Rob Kitchin, “Thinking Critically About and Researching Algorithms,” *Information, Communication & Society* 20, no. 1 (2017): 14–29, <https://doi.org/10.1080/1369118X.2016.1154087>.

<sup>24</sup>Williamson, “Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education.”

<sup>25</sup>Antoinette Rouvroy and Thomas Berns, “Algorithmic Governmentality and Prospects of Emancipation,” *Réseaux* 177, no. 1 (2013): 163–96, <https://doi.org/10.3917/res.177.0163>.

<sup>26</sup>Ben Williamson, *Big Data in Education: The Digital Future of Learning, Policy and Practice* (Los Angeles, United States of America: Sage Publication, Inc., 2017),

From the standpoint of effectiveness, the literature suggests that artificial intelligence holds significant potential to enhance administrative efficiency and improve the accuracy of decision-making through large-scale data analytics.<sup>27</sup> However, the algorithmic rationality underpinning such systems centered on optimization and prediction tends to reduce social complexity into quantifiable variables. Mittelstadt et al. argue that algorithmic systems may embed latent biases arising from data structures, model design, or the assumptions of developers.<sup>28</sup> Within educational governance, such biases may shape the allocation of resources, the classification of students, and the formulation of performance-based policies, thereby influencing institutional outcomes in ways that are not always transparent or equitable.

Ethically, this transformation generates acute accountability concerns. When administrative decisions are shaped by predictive models, the locus of responsibility becomes increasingly diffuse. The principle of algorithmic accountability requires transparency, decision traceability, and the establishment of independent auditing mechanisms to ensure that automated or semi-automated determinations remain open to scrutiny and redress.<sup>29</sup>

Thematic analysis indicates that algorithmic bias and accountability constitute the two most dominant issues in the public-sector AI ethics literature. Bias is not merely technical in nature; it is structural, insofar as it reflects historically embedded patterns of inequality within datasets. In Indonesia's plural educational context, such bias carries the risk of reductively shaping socio-religious identities through classificatory logics that fail to capture their lived complexity.

The themes of transparency and digital surveillance highlight an inherent tension between managerial efficiency and the protection of individual rights. Learning analytics systems may enhance the precision of targeted interventions; however, in the absence of robust data protection regulations, digital monitoring can foster a culture of control that stands in tension with emancipatory conceptions of education.

The theme of humanistic reduction and content polarization underscores that the impact of AI is not solely administrative but epistemological. When educational values are translated into performance metrics, dialogical and reflective dimensions risk marginalization. This dynamic bears direct implications for the agenda of religious moderation, which emphasizes balance, mutual recognition, and respect for diversity.

Taken together, the thematic analysis reinforces the argument that AI-driven educational governance requires a systematic and contextually grounded ethical framework to prevent outcomes that may contradict the normative objectives of national education.

### ***Ethical Risks and Their Implications for Religious Moderation***

At the global level, several AI regulatory frameworks provide normative foundations that may serve as relevant points of reference emphasizes human rights,

---

<https://doi.org/10.4135/9781529714920>; Williamson and Eynon, "Historical Threads, Missing Links, and Future Directions in AI in Education."

<sup>27</sup>Holmes, Bialik, and Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*.

<sup>28</sup>Mittelstadt et al., "The Ethics of Algorithms: Mapping the Debate."

<sup>29</sup>Veale, Van Kleek, and Binns, "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making."

cultural diversity, and social justice as core ethical principles underpinning the governance of artificial intelligence.<sup>30</sup> Similarly, OECD<sup>31</sup> advances principles of accountability and transparency, particularly in the public sector, underscoring the need for responsible oversight in state-led AI deployment.<sup>32</sup>

This risk-based approach holds particular relevance for Indonesia, as it enables regulatory differentiation according to the social impact of specific technologies. A learning recommendation system, for instance, may be categorized as low risk, whereas systems designed for ideological classification or social risk mapping could be deemed high risk, given their implications for civil liberties and freedom of thought.

Nevertheless, sociocultural differences necessitate contextual adaptation. Indonesia's plural socio-religious structure, shaped by complex historical dynamics of identity formation, requires that global AI governance frameworks be interpreted in light of national philosophical and constitutional foundations. In this regard, alignment with Pancasila, policies promoting religious moderation, and constitutional guarantees of religious freedom is essential to ensure normative coherence.

Within the Indonesian context, education carries a pronounced ideological and cultural dimension, particularly in relation to strengthening religious moderation. This agenda emphasizes balance, tolerance, non-violence, and respect for diversity. The challenge arises when algorithmic systems mediate the distribution of information and generate content recommendations within educational environments, thereby shaping exposure, discourse, and interpretive frameworks in ways that may either support or inadvertently undermine these normative commitments.

Research in algorithmic ethics indicates that recommendation systems tend to reinforce pre-existing preference patterns through personalization mechanisms grounded in historical user data.<sup>33</sup> Within digital education ecosystems, such dynamics carry the risk of generating echo chambers that constrain exposure to diverse perspectives and limit dialogical engagement. Selwyn<sup>34</sup> cautions that the integration of artificial intelligence in education is not value-neutral; rather, it embodies particular assumptions concerning efficiency, performativity, and control. Consequently, AI-mediated environments may subtly recalibrate educational priorities, privileging measurable engagement and optimization over pluralism, critical reflection, and deliberative exchange.

From the perspective of religious moderation, the risk of value reduction emerges when theological and social complexities are rendered into binary data classifications. If algorithms are deployed to detect “radicalism risk” or to categorize ideological tendencies without a stringent ethical framework, they may generate problematic labeling effects that undermine inclusivity and stigmatize particular groups. Such classificatory practices risk transforming nuanced socio-religious identities into administratively convenient but normatively distorted categories. UNESCO<sup>35</sup> underscores the importance of meaningful human oversight in the use of generative artificial intelligence in education, particularly

---

<sup>30</sup>UNESCO, “Recommendation on the Ethics of Artificial Intelligence.”

<sup>31</sup>OECD, “What Are the OECD Principles on AI?”

<sup>32</sup>European Commission, “Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)” (Brussels, 2021).

<sup>33</sup>Floridi et al., “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.”

<sup>34</sup>Selwyn, *Should Robots Replace Teachers? AI and the Future of Education*.

<sup>35</sup>UNESCO, “Guidance for Generative AI in Education and Research.”

to prevent misinformation and cultural bias. This emphasis on human supervision is crucial in contexts where automated systems intersect with sensitive domains of belief, identity, and civic belonging.

Accordingly, the relationship between artificial intelligence and religious moderation is inherently ambivalent. On the one hand, technological systems can serve as instruments for strengthening critical literacy and fostering cross-perspectival dialogue. On the other hand, they may constrict diversity if system design fails to incorporate principles of justice and epistemic pluralism. The normative outcome, therefore, depends less on the technology per se than on the ethical architectures, governance safeguards, and value commitments embedded within its development and deployment.

The Risk Stratification Matrix is designed as an operational instrument to assist policymakers and educational leaders in determining the appropriate level of oversight and control for AI implementation. A risk-based approach rests on the premise that not all AI systems warrant identical degrees of regulation. Such differentiation is essential to maintaining equilibrium between technological innovation and the safeguarding of core educational values.

Within the low-risk category, AI systems tend to perform administrative functions that do not directly affect individual rights. Governance at this level may therefore concentrate on technical reliability, cybersecurity, and system integrity. As systems begin to influence access to learning opportunities or academic evaluation, however, they move into a medium-risk category, necessitating more rigorous bias testing, transparency requirements, and documentation standards.

High- and very high-risk categories become particularly salient in the context of religious moderation. Systems that engage in ideological classification or map religious tendencies carry the potential for significant social consequences, including stigmatization and polarization. Under such conditions, human oversight cannot be supplanted by automated processes. Independent audits, formal grievance mechanisms, and the involvement of multidisciplinary review panels become indispensable safeguards.

Normatively, the matrix reinforces the principle of proportionality: the greater the potential social impact of an AI system, the more stringent the oversight mechanisms it requires. In this way, the MEKAI-MB framework advances beyond a purely conceptual ethical model to offer an actionable managerial instrument for educational institutions and national regulators.

### ***Hypothetical Design of an Evidence-Synthesis-Based Policy Model***

To enhance managerial contribution and policy relevance, this study formulates the Ethical Governance Model of Artificial Intelligence in Education Based on Religious Moderation (MEKAI-MB) as a hypothetical policy design developed through an evidence synthesis approach. The model is not constructed speculatively; rather, it emerges from the systematic integration of empirical findings and conceptual insights drawn from the literature on AI ethics, algorithmic governance in the public sector, and studies of religious moderation within both Indonesian and global educational contexts. Accordingly, MEKAI-MB is positioned as a normative-operational conceptual framework that bridges global ethical principles with nationally grounded socio-cultural imperatives.

The first component of the model is the risk mapping layer. Any implementation of artificial intelligence within the education sector must begin with a systematic identification of potential social, ethical, and ideological risks. This mapping process classifies systems according to the magnitude of their impact on individual rights and broader social dynamics, distinguishing among low-, medium-, and high-risk categories. Systems that function solely as administrative support tools such as automated archiving or statistical data processing may reasonably be categorized as low risk. By contrast, systems that engage in identity classification, ideological profiling, character assessment, or the detection of “radicalism risk” should be designated as high risk. Such applications require independent auditing, bias testing, and social impact assessment prior to full-scale deployment. This risk-based approach is essential to preventing uniform or indiscriminate technological adoption, ensuring instead that regulatory intensity corresponds proportionately to the socio-religious sensitivities embedded within Indonesia’s educational context.

The second component is the ethical design protocol. The model stipulates that principles of algorithmic justice, model transparency, and meaningful human oversight must be embedded at the system design stage rather than retrofitted after implementation. System developers are required to document data sources, model training procedures, design assumptions, and potential biases that may arise in processes of classification or prediction. Such transparency enables both academic scrutiny and public evaluation of the system’s legitimacy. Moreover, the presence of human oversight ensures that final decisions remain grounded in ethical and pedagogical judgment rather than being determined solely by statistical optimization logics. Within educational contexts, this approach is essential to preserving the humanistic and relational dimensions that constitute the foundation of meaningful learning processes.

The third component is the public accountability mechanism. AI-driven educational governance must not erode the individual’s right to procedural clarity and fairness. Accordingly, educational institutions deploying AI systems are required to establish formal appeal mechanisms for students, educators, and administrative staff affected by algorithm-based decisions. These mechanisms should be designed to ensure transparency, accessibility, and the use of clear, publicly intelligible language. In addition, institutional policies governing AI use must be publicly disclosed, enabling stakeholders to understand the scope, objectives, and limitations of the technology in question. Through such measures, policy legitimacy rests not solely on claims of efficiency, but also on social acceptance and adherence to principles of democratic accountability.

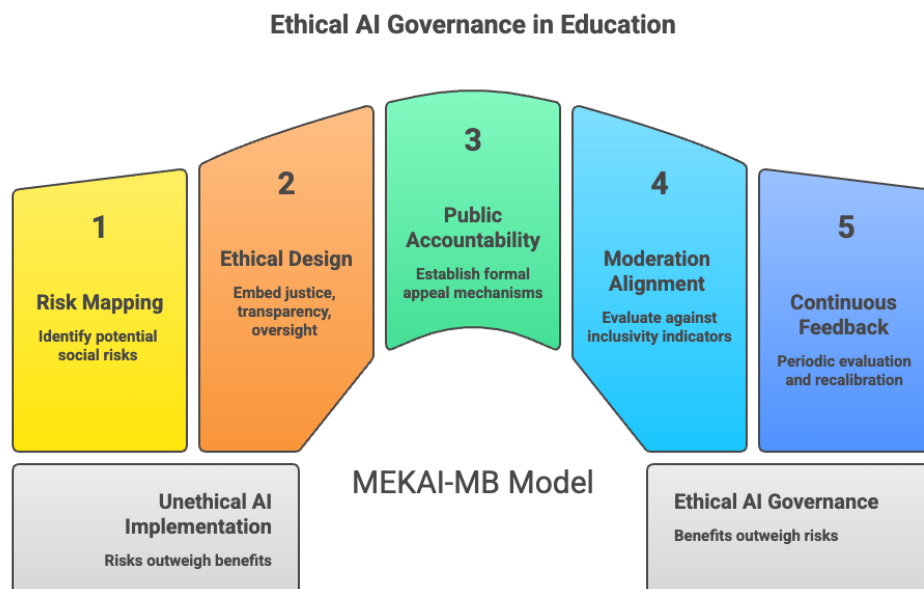
The fourth component is the moderation alignment review. The MEKAI-MB model positions religious moderation as a normative parameter that must be systematically integrated into the evaluation of educational AI systems. Each system should be assessed against indicators of inclusivity, non-violence, respect for diversity, and the prevention of discrimination. This evaluation is not conceived as a purely technocratic exercise; rather, it is conducted through a multidisciplinary panel comprising education experts, information technology specialists, scholars of religious studies, and representatives of civil society. Such a multidisciplinary approach is essential, as questions of religious moderation encompass theological, social, and political dimensions that cannot be reduced to technical considerations alone. Accordingly, AI systems implemented within

educational settings are evaluated not only in terms of technical performance but also with respect to their contribution to social harmony and national cohesion.

The fifth component is continuous evidence-based feedback. The model emphasizes that AI-based policies are inherently dynamic and must be subjected to periodic evaluation through social impact research and empirical data analysis. Longitudinal assessment is necessary to detect potential indirect effects, such as opinion polarization, unintended discrimination, or the erosion of intergroup dialogue. This feedback process enables adaptive, evidence-informed policy recalibration, preventing governance frameworks from becoming anchored to initial assumptions that may later prove flawed. Through sustained evaluative mechanisms, AI systems can be refined iteratively in light of emerging research findings and evolving social dynamics, thereby ensuring their continued alignment with educational and societal objectives.

Overall, the MEKAI-MB model is not intended as a rigid or definitive regulatory instrument; rather, it is conceived as a flexible and adaptive conceptual framework for policymakers. Its principal strength lies in the simultaneous integration of technical and normative dimensions. On the one hand, the model acknowledges the imperatives of efficiency and technological innovation in educational governance. On the other hand, it affirms that such innovation must remain subordinate to principles of justice, transparency, accountability, and alignment with the values of religious moderation.

By adopting an evidence synthesis approach, the model seeks to ensure that AI policy in the education sector is not only responsive to global technological developments but also contextually grounded and socially responsible within Indonesia's plural socio-cultural landscape.



**Figure 1.** Ethical AI Governance in Education MEKAI-MB Model

***Ethical Synthesis: Toward Value-Based Educational Governance***

Based on the normative analysis, this study identifies three core ethical principles that should be systematically integrated into AI-driven educational governance. The first is the principle of algorithmic justice, which requires that systems neither reproduce

structural bias nor operate beyond the reach of independent auditing.<sup>36</sup> This principle demands rigorous bias testing, impact assessment, and corrective mechanisms to ensure equitable outcomes across diverse social groups. The second is the principle of transparency and public accountability, whereby data-driven decision-making processes must be explicable in rational and accessible terms to relevant stakeholders.<sup>37</sup> Transparency extends beyond technical disclosure; it encompasses the communicability of underlying assumptions, data sources, and evaluative criteria that shape institutional outcomes. The third is the principle of human-centered governance, which affirms that human agents remain the ultimate decision-makers and custodians of educational values.<sup>38</sup> Under this principle, AI functions as a decision-support instrument rather than an autonomous authority, ensuring that pedagogical, ethical, and civic considerations retain primacy within governance structures.

In contrast to prior studies that concentrate primarily on technical or pedagogical dimensions, this inquiry demonstrates that the governance dimension carries direct implications for national agendas, including the strengthening of religious moderation. What further emerges from the analysis is the imperative to integrate global ethical frameworks with Indonesia's local socio-political and cultural context. Absent such contextualization, global AI ethics guidelines risk remaining normatively abstract and insufficiently operational within national policy practice. Their principles however well articulated may lack regulatory traction if not translated into institutional procedures, oversight mechanisms, and culturally resonant standards that reflect Indonesia's plural social fabric and constitutional commitments.

Conceptually, these findings affirm that AI-based educational governance must be designed not merely as a mechanism for administrative efficiency, but as a value-oriented system that safeguards equilibrium between technological innovation and social responsibility. The integration of principles of justice, inclusivity, and respect for diversity constitutes a necessary precondition to ensure that the digital transformation of education does not undermine the broader objective of strengthening religious moderation. In this formulation, technological advancement is normatively legitimate only insofar as it remains aligned with the ethical commitments and civic purposes that underpin national education.

### ***Managerial and Educational Governance Implications***

From an educational management perspective, the findings indicate that institutional leaders cannot fully delegate decision-making authority to technology vendors. The development of algorithmic literacy at the managerial level is essential to ensure that AI-based decision processes remain under institutional oversight and aligned with organizational values and regulatory obligations. Without such capacity, institutions risk becoming passive adopters of externally designed systems whose underlying assumptions may not reflect local priorities. Furthermore, the establishment of technology ethics committees within universities and ministries of education is warranted. These committees would function as internal oversight bodies tasked with evaluating the

---

<sup>36</sup>Mittelstadt et al., "The Ethics of Algorithms: Mapping the Debate"; Veale, Van Kleek, and Binns, "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making."

<sup>37</sup>OECD, "What Are the OECD Principles on AI?"

<sup>38</sup>UNESCO, "Recommendation on the Ethics of Artificial Intelligence."

feasibility, ethical integrity, and contextual suitability of AI systems prior to large-scale implementation. Through multidisciplinary review and structured assessment procedures, such bodies can help institutionalize responsible innovation and mitigate potential ethical and social risks.

Strengthening human resource capacity is equally critical. Teachers and educational staff must develop a clear understanding of the limitations and potential biases inherent in AI systems to prevent uncritical or absolute reliance on automated recommendations. Such capacity-building efforts should encompass not only technical literacy but also ethical awareness, enabling educators to interpret algorithmic outputs with professional judgment and contextual sensitivity. In this way, AI remains a decision-support tool rather than a substitute for pedagogical expertise and moral discernment.

## **CONCLUSION**

This study set out to critically examine the ethical implications of artificial intelligence in educational governance in Indonesia and its relevance to strengthening religious moderation. Drawing upon a systematic analysis of scholarly literature and policy frameworks, it concludes that the integration of AI into educational governance has generated significant transformations in administrative efficiency, data-driven accountability, and the optimization of decision-making processes. However, these transformations are not value-neutral. Algorithmic systems may reproduce structural biases, narrow the diversity of perspectives, and obscure lines of public responsibility if implemented without a robust ethical framework. The governance of AI in education, therefore, cannot be reduced to technical performance metrics alone; it must be situated within broader normative commitments that safeguard justice, inclusivity, and democratic accountability.

Within the framework of religious moderation, the deployment of artificial intelligence in the education sector must be situated within a value-based paradigm that foregrounds justice, inclusivity, and respect for diversity. Absent meaningful human oversight, system transparency, and rigorous bias auditing mechanisms, AI-driven governance risks constructing reductive socio-religious categorizations that may undermine pluralism and social cohesion. Accordingly, this study underscores the urgency of developing AI-based educational governance grounded in a human-centered paradigm. Such an approach requires that global ethical principles serve as normative foundations while being carefully contextualized to Indonesia's socio-cultural realities. In this configuration, technological advancement is aligned not only with administrative efficiency but also with the broader civic mission of sustaining diversity, mutual recognition, and responsible public governance.

Conceptually, the principal contribution of this study lies in its synthesis of AI ethics scholarship with the agenda of religious moderation within the framework of educational governance. By bridging these domains, the study advances a value-oriented perspective that situates technological governance within the broader civic and cultural objectives of national education. It recommends that policymakers institutionalize principles of algorithmic transparency, public accountability, and stakeholder participation at every stage of AI implementation in the education sector. Embedding these principles into regulatory design, institutional procedures, and oversight mechanisms is essential to ensuring that AI deployment remains ethically grounded and socially legitimate. Future research should undertake empirical investigations within

educational institutions to assess the effectiveness of the proposed ethical governance model and to evaluate its concrete impact on the practice of religious moderation. Such studies would provide critical evidence to refine the model and strengthen its applicability in real-world policy contexts.

## REFERENCE

Bacchi, Carol. *Analysing Policy: What's the Problem Represented to Be?* Frenchs Forest, NSW: Pearson, 2009.

Commission, European. "Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)." Brussels, 2021.

Creswell, John W. *Research Design: Pendekatan Metode Kualitatif, Kuantitatif, Dan Campuran*. 4 Terj. ol. Yogyakarta: Pustaka Pelajar, 2016.

Floridi, Luciano, Josh Cows, Monica Beltrametti, Rachid Chatila, Patrick Chazerand, Virginie Dignum, Christoph Luetge, et al. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28, no. 4 (2018): 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.

Hefni, Wildani. "Moderasi Beragama Dalam Ruang Digital: Studi Pengarusutamaan Moderasi Beragama Di Perguruan Tinggi Keagamaan Islam Negeri." *Jurnal Bimas Islam* 13, no. 1 (2020): 1–22. <https://api.semanticscholar.org/CorpusID:225547190>.

Holmes, Wayne, Maya Bialik, and Charles Fadel. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Boston: Center for Curriculum Redesign, 2019.

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389–99. <https://doi.org/10.1038/s42256-019-0088-2>.

Kitchin, Rob. "Thinking Critically About and Researching Algorithms." *Information, Communication & Society* 20, no. 1 (2017): 14–29. <https://doi.org/10.1080/1369118X.2016.1154087>.

Mittelstadt, Brent D, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3, no. 2 (2016). <https://doi.org/10.1177/2053951716679679>.

Moleong, Lexy J, and D M A Anggoro. *Metodologi Penelitian Kualitatif*. Edited by Edisi Revisi. Bandung: PT Remaja Rosdakarya, 2019.

Nasir, Muhammad, and Muhammad Khoirul Rijal. "Keeping the Middle Path: Mainstreaming Religious Moderation through Islamic Higher Education Institutions in Indonesia." *Indonesian Journal of Islam and Muslim Societies* 11, no. 2 (2021): 213–41. <https://doi.org/10.18326/ijims.v11i2>.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

OECD. "What Are the OECD Principles on AI?" Paris: OECD Publishing, 2024. <https://legalinstruments.oecd.org/en/>.

Rouvroy, Antoinette, and Thomas Berns. "Algorithmic Governmentality and Prospects

- of Emancipation.” *Réseaux* 177, no. 1 (2013): 163–96. <https://doi.org/10.3917/res.177.0163>.
- Selwyn, Neil. *Should Robots Replace Teachers? AI and the Future of Education*. Cambridge: Polity Press, 2019.
- Shah, Harshad. “Algorithmic Accountability.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (2018). <https://doi.org/10.1098/rsta.2017.0362>.
- Snyder, Hannah. “Literature Review as a Research Methodology: An Overview and Guidelines.” *Journal of Business Research* 104, no. March (2019): 333–39. <https://doi.org/10.1016/j.jbusres.2019.07.039>.
- UNESCO. “Guidance for Generative AI in Education and Research.” Paris: UNESCO Publishing, 2023. <https://doi.org/10.54675/ewzm9535>.
- . “Recommendation on the Ethics of Artificial Intelligence.” Paris: UNESCO Publishing, 2022. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>.
- Veale, Michael, Max Van Kleek, and Reuben Binns. “Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making.” *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018. <https://doi.org/10.1145/3173574.3174014>.
- Williamson, Ben. *Big Data in Education: The Digital Future of Learning, Policy and Practice*. Los Angeles, United States of America: Sage Publication, Inc., 2017. <https://doi.org/10.4135/9781529714920>.
- . “Governing Software: Networks, Databases and Algorithmic Power in the Digital Governance of Public Education.” *Learning, Media and Technology* 40, no. 1 (2015): 83–105. <https://doi.org/10.1080/17439884.2014.924527>.
- Williamson, Ben, and Rebecca Eynon. “Historical Threads, Missing Links, and Future Directions in AI in Education.” *Learning, Media and Technology* 45, no. 3 (2020): 223–35. <https://doi.org/10.1080/17439884.2020.1798995>.