El-Rusyd
https://ejournal.stitahlussunnah.ac.id/index.php/el-rusyd

Scientific Neutrality in the Age of Artificial Intelligence: A Critical Analysis of the Value-Free Ideal

Budi Harto[1], Ridha Ahida[2]
[1]Sekolah Tinggi Ilmu Ekonomi El Hakim, Indonesia
[2]Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia
budiharto212@gmail.com *

Article Information:
Received October 21, 2025
Revised November 28, 2025
Accepted December 5, 2025

Keywords: *Scientific neutrality, value-free ideal, artificial intelligence (AI), Thomas Kuhn, critical analysis*

Abstract

The debate on the neutrality of scientific knowledge whether science is value-free or value-laden has been a central discussion in the philosophy of science from the era of logical positivism to the present. The positivist tradition of Carnap and Reichenbach, along with Popper's falsificationism, argues that the process of scientific justification must be separated from non-epistemic values in order to secure objectivity. However, the use of artificial intelligence (AI) in contemporary scientific research presents empirical evidence that challenges this ideal of value-free science. This study critically examines how the use of AI in science supports the value-laden position advocated by Thomas Kuhn, Helen Longino, and feminist epistemology. Employing a qualitative method with a content analysis approach, the study is analyzed through a philosophical analytical framework. The findings identify three major positions: neopositivism, which defends the value-free ideal; the Kuhnian position, which acknowledges the role of epistemic values; and the radical value-laden position. The discussion demonstrates that artificial intelligence substantiates the value-laden view through four dimensions: algorithmic bias as a manifestation of social values, value-laden design choices in artificial intelligence systems, the incommensurability of artificial intelligence paradigms, and situated objectivity, which requires explicit recognition of embedded values. The study concludes that artificial intelligence not only confirms but reinforces the argument that the value-free ideal is a philosophical illusion, and that responsible science requires critical reflexivity toward the values embedded within scientific practice.

INTRODUCTION

The debate surrounding the neutrality of scientific knowledge whether science can or should be free from non-epistemic values has long been one of the most fundamental issues in the philosophy of science. The logical positivist tradition asserts that scientific objectivity requires a strict separation between facts and values, as emphasized in the works of Carnap, Reichenbach, and proponents of the analysis of scientific language (Politi, 2024). In contrast, critiques from Thomas Kuhn

How to cite:	Harto, B., Ahida, R. (2025). Scientific Neutrality in the Age of Artificial Intelligence: A Critical Analysis of the Value-Free Ideal. *El-Rusyd*, *10*(2), 179-190.
E-ISSN:	2580-0256
Published by:	The Institute for Research and Community Service

to feminist epistemologists demonstrate that social, cultural, and even political values cannot be entirely removed from scientific practice (Brown, 2024). This debate carries significant implications for our understanding of the epistemic authority of science, the social responsibility of scientists, and the relationship between science and society.

In the contemporary context, these classical discussions gain renewed relevance with the integration of artificial intelligence into scientific research activities. Artificial intelligence is no longer understood as a neutral technical tool but as a complex system that embeds design choices, epistemological assumptions, and social values within its algorithms. When artificial intelligence is used to generate knowledge from data analysis to hypothesis formulation the question of whether science is value-free or value-laden becomes more than a theoretical discourse; it materializes in technical configurations that can be empirically examined (Resnik & Hosseini, 2025).

The philosophical roots of this debate can be traced back to logical positivism, which developed within the Vienna Circle in the 1920s and 1930s. This movement sought to construct a unified science grounded in verifiable observations and the logical analysis of scientific language (Vagelli, 2024). For positivists, meaningful statements are those that can be empirically verified or are logical tautologies, whereas value statements are considered to lack epistemic status. Reichenbach later differentiated between the context of discovery and the context of justification, locating values only within the context of discovery and asserting that scientific justification must remain entirely value-free (Ward, 2021).

Karl Popper, though critical of the verification principle, maintained the idea that scientific objectivity requires freedom from values and subjective biases. In falsificationism, objective science is achieved through open criticism and intersubjective testing (Mitra, 2020). However, several challenges arise to this value-free ideal. First, the problem of underdetermination shows that more than one theory can be consistent with the same evidence, meaning that theory choice involves non-empirical considerations. Second, the theory-ladenness of observation reveals that observation depends on particular conceptual frameworks. Third, inductive risk requires scientists to make decisions based on value judgments concerning the consequences of error.

A major shift came with Thomas Kuhn's The Structure of Scientific Revolutions, which introduced the concept of paradigms as constellations of beliefs, values, and techniques that structure scientific practice (Haryanto & Paryanto, 2023). Kuhn argued that paradigm shifts are incommensurable because each paradigm carries distinct definitions of problems, methods, and criteria of truth. He also highlighted the role of epistemic values such as accuracy, consistency, and simplicity which may be weighed differently by scientists without undermining rationality.

Feminist epistemology then advanced a more radical critique. Helen Longino argued that background assumptions linking evidence and hypotheses always contain value elements (Bueter, 2022). Sandra Harding, through the concept of "strong objectivity," suggested that scientific objectivity is strengthened through critical engagement with social values. Donna Haraway (Faqih et al., 2025) added that all knowledge is situated, meaning that genuine objectivity requires acknowledging social positions and the limits of perspective.

In more recent developments, the use of artificial intelligence in scientific research provides an opportunity to empirically examine the value-free ideal. Artificial intelligence systems particularly machine learning models encode values through data selection, feature construction, model architectures, and optimization objectives (Resnik et al., 2025). When these algorithms generate scientific knowledge, the value entanglement that once seemed abstract becomes concrete and measurable

(Schintler et al., 2023). Therefore, artificial intelligence emerges as a critical case study illustrating that scientists cannot fully avoid values in the scientific process (Ferrario et al., 2024).

This study offers a novel contribution by providing the first systematic philosophical analysis of how artificial intelligence empirically challenges the long-standing ideal of value-free science. Unlike previous literature that discusses value-laden science only at a conceptual level, this research demonstrates through a structured examination of artificial intelligence systems, algorithmic design choices, and paradigm-level constraints that artificial intelligence operationalizes social, epistemic, and ethical values within scientific practice. The study further advances the debate by integrating insights from logical positivism, Kuhnian theory, and feminist epistemology into a unified analytical framework, which has not been explicitly synthesized in prior research. Through this integration, the study reveals that artificial intelligence not only reflects value entanglement but also amplifies it, thereby reframing objectivity as a situated and reflexive construct. This offers a new epistemological foundation for responsible scientific inquiry in the era of intelligent systems.

METHODS

This study employs a qualitative method with a philosophical content-analysis approach to examine the debate between value-free and value-laden science in the context of artificial intelligence (Khan et al., 2022). This method is chosen because it aligns with the characteristics of philosophical research, which require in-depth analysis of arguments, assumptions, and theoretical positions found in academic literature. The data sources for this research consist of scholarly journal articles, books, and philosophical encyclopedias that discuss issues of values in scientific inquiry and artificial intelligence. The databases used include PhilPapers for philosophical literature, Scopus and Web of Science for interdisciplinary coverage, and Google Scholar to ensure that no relevant literature is overlooked. Additional searches were conducted in the Stanford Encyclopedia of Philosophy and specialized journals such as Philosophy of Science, the British Journal for the Philosophy of Science, and Ethics and Information Technology (Hagendorff, 2024).

The literature search employed English keywords: "value-free" OR "value-laden" OR "objectivity," combined with "artificial intelligence" OR "machine learning" AND "science" OR "scientific research." Additional searches were performed for specific philosophers such as Carnap, Reichenbach, Popper, and Kuhn, combined with the keywords "values" and "AI" (Douglas & Branch, 2024). The period covered by the literature review spans January 2018 to October 2024. The criteria for selecting literature were as follows: (1) the article discusses issues of values, objectivity, or neutrality in the context of artificial intelligence or science; (2) it provides substantial philosophical analysis; (3) it is published in a reputable peer-reviewed journal; (4) it is accessible in full-text form; and (5) it contains a minimum analytical depth of 3,000 words for journal articles (Hasanzadeh et al., 2025). The selection process was conducted in four stages. The first stage yielded 428 potentially relevant items. The second stage, based on title screening, narrowed the list to 186 items. The third stage, based on abstracts, resulted in 103 articles. The fourth stage, after full-text reading, identified 72 articles that met all criteria (Delgado-Chaves et al., 2025).

Data analysis was carried out using a philosophical reconstruction method combined with thematic analysis. Each article was reviewed to identify its central thesis, supporting arguments, philosophical frameworks, underlying assumptions, and stated implications (Monaro et al., 2022). Each article was then categorized based on its position regarding value-free versus value-laden science, its philosophical tradition, and its stance on objectivity. Thematic analysis was subsequently conducted to identify patterns emerging across the literature (Engkizar et al., 2025). The initial coding process identified 12 preliminary themes, which were later consolidated into three main positions through an iterative process. For each position, representative articles were selected, and their arguments were reconstructed in detail (Bar-Gil, 2025).

RESULT AND DISCUSSION

Neopositivism and the Defense of the Value-Free Ideal

The first position, identified in 18 articles (25%), can be categorized as neopositivism or a contemporary defense of the value-free ideal. This position maintains a fundamental commitment to the separation between facts and values in the context of scientific justification. Its central argument is that although artificial intelligence technologies involve design choices that encode values, the core scientific content produced through artificial intelligence can and should be evaluated independently of those values (Brown, 2024). Literature in this category distinguishes between "AI as a tool" and "AI-generated knowledge." As a tool, artificial intelligence indeed involves engineering trade-offs such as between accuracy and interpretability but these are treated as analogous to traditional scientific instruments such as microscopes or telescopes, which also involve design choices without rendering the resulting observations problematically value-laden. What matters, according to this view, is that the data or findings generated by artificial intelligence can be independently validated using standard scientific methods (Alvarado, 2023).

Several articles draw on Reichenbach's distinction between the context of discovery and the context of justification to argue that artificial intelligence is primarily relevant to the discovery phase for example, in generating hypotheses or identifying patterns in large datasets while the testing and validation of those hypotheses must still adhere to value-free epistemic standards (Nazer et al., 2023). The choice to use a particular artificial intelligence tool may be influenced by values, but proponents argue that this is not fundamentally different from methodological or instrumental choices scientists always make (Baroud et al., 2025; Engkizar et al., 2024, 2025). This position also emphasizes that concerns about bias in artificial intelligence are primarily technical problems that can be resolved through improved engineering. Algorithmic bias is framed as a defect or imperfection that can and should be eliminated through better data collection, improved algorithms (Deckker & Sumanasekara, 2025), and rigorous validation. The ultimate ideal is the development of fully objective artificial intelligence systems that do not reflect social biases or particular interests (Cross et al., 2024).

Moderate Kuhnian Perspectives and Epistemic Values

The second position, found in 28 articles (38.9%), can be characterized as a moderate Kuhnian stance that acknowledges the role of values but seeks to maintain a distinction between legitimate epistemic values and illegitimate non-epistemic values. This position is inspired by Kuhn's analysis of theory choice and paradigm shifts. Literature in this category argues that the use of artificial intelligence in science clearly demonstrates the essential role of values in scientific reasoning, but not all values are equally acceptable. Epistemic values such as empirical adequacy, consistency, explanatory power, simplicity, and fruitfulness are considered intrinsic to scientific inquiry and serve as standards for evaluating theories and methods. These are contrasted with non-epistemic values such as political preferences, economic interests, or cultural biases, which should not influence scientific judgment (Kassymova et al., 2025; Engkizar et al., 2023, 2025).

In the context of artificial intelligence, this position acknowledges that choices in designing or deploying artificial intelligence systems inevitably involve value judgments. For example, deciding acceptable trade-offs between Type I and Type II errors in classification tasks requires weighing values. However, such value judgements are permissible when they are transparent, well-reasoned, and subject to critical scrutiny within the scientific community (Zanetti et al., 2023). What distinguishes this position is its emphasis on pluralism and contextual sensitivity in value assessment. Moderate Kuhnians accept that epistemic values can be weighted differently depending on the context without rendering science arbitrary or subjective. Scientists may legitimately disagree about whether simplicity or scope should carry more weight in a given case, and such disagreements are resolved through rational debate and community consensus.

Several articles use Kuhn's paradigm concept to analyze artificial intelligence in scientific practice. Artificial intelligence is viewed as introducing new paradigms in certain domains for instance, in data-intensive science or pattern-recognition approaches that carry different epistemic values and standards compared with traditional hypothesis-testing paradigms (Dhar, 2024). Incommensurability between paradigms means that there is no neutral standpoint from which to determine which is "better," but this does not imply that paradigm choice is purely subjective. Some literature also explores whether certain values traditionally classified as "non-epistemic" may in fact have epistemic dimensions. For example, values related to inclusivity or diversity may yield epistemic benefits by helping reveal biases or blind spots in research. Accordingly, some articles support expanding the conception of epistemic values to include certain social values conducive to more robust knowledge production.

Radically Value-Laden Perspectives and Situated Knowledge

The third position, identified in 26 articles (36.1%), takes the most radical stance, asserting that science is fundamentally and irreducibly value-laden, and that attempts to separate epistemic from non-epistemic values are misguided (Stamenkovic, 2024). This position draws heavily on feminist epistemology, Science and Technology Studies, and critical theory, with explicit references to the works of Longino, Harding, and Haraway. The central argument is that distinctions relied on by positivists and moderate Kuhnians between discovery and justification, between epistemic and non-epistemic values, between facts and values are untenable because they obscure the ways in which values permeate every aspect of scientific practice. Background assumptions required to link evidence with hypotheses, the categories used to classify phenomena, and the standards used to evaluate claims all involve substantive commitments that cannot be justified purely epistemically (Parker, 2024).

In the context of artificial intelligence, this position argues that these systems offer a uniquely vivid demonstration of value entanglement because values are encoded in mathematically formalized structures that can be inspected directly. Literature extensively documents how design choices in artificial intelligence from data collection strategies, feature selection, model architectures, and training objectives, to contexts of deployment reflect and reinforce particular social arrangements, power relations, and epistemic assumptions (Widder, 2024). These are not peripheral features that can be engineered away, but constitutive elements of how artificial intelligence systems function. Several articles employ Haraway's concept of "situated knowledge" to analyze artificial intelligenceartificial intelligence. Artificial intelligence systems, like all knowledge-producing systems, are situated: they are developed by specific actors, within particular institutional settings, using datasets that reflect specific perspectives (Trächtler, 2024). The assumption that artificial intelligence offers a "view from nowhere" or purely objective analysis is precisely what Haraway critiques as the "god trick." Genuine objectivity what Harding calls "strong objectivity" requires acknowledging situatedness and conducting critical examination of how it shapes what can be seen and known (Ye, 2024).

This literature also emphasizes the material and infrastructural dimensions of value in artificial intelligence. Data centers powering artificial intelligence consume vast amounts of energy with environmental costs; low-wage workers perform essential data labeling; surveillance infrastructures collect data that train models. These material relations are not external to the epistemic content of artificial intelligence; they shape the questions that can be asked, the problems that are prioritized, and the solutions that become imaginable (Bueter, 2022). This position rejects the value-free ideal as either desirable or attainable. Instead, it argues for "value-conscious" or "value-explicit" science that openly acknowledges the values informing it, subjects them to democratic deliberation, and holds scientific practice accountable for its impacts on diverse communities. Objectivity is redefined not as the absence of values or perspective, but as critical engagement with a plurality of perspectives and systematic examination of how power relations shape knowledge production.

Algorithmic Bias as a Manifestation of Values in Technical Structures

The phenomenon of algorithmic bias provides the most concrete evidence that technical systems traditionally regarded as neutral, in fact encode social values within their mathematical structures (Zajko, 2021). Thomas Kuhn argued that scientific paradigms encompass not only theories but also the values, methods, and assumptions that shape what scientists see and how they interpret observations (Greene et al., 2022). In the context of artificial intelligence, this paradigm materializes in algorithmic architectures that can be empirically examined. The case of facial recognition systems that exhibit differential accuracy across racial and gender groups offers a striking illustration (Leslie, 2020). Research by Joy Buolamwini and Timnit Gebru demonstrates that leading commercial systems show error rates of up to 34.7% for dark-skinned women compared with 0.8% for light-skinned men (Birhane, 2022). This is not simply a "technical flaw" solvable through better engineering, but rather a manifestation of values embedded throughout the development process.

Helen Longino argues that background assumptions linking evidence to hypotheses always involve contextual values. In machine learning, such assumptions materialize in decisions about which training data are considered "representative," which features count as "relevant," and which metrics sufficiently measure "success" (Huang et al., 2022). When training datasets are dominated by light-skinned male faces not by accident but due to power structures that determine who develops technology, for whom, and using whose data these biases encode values about who matters, who is visible, and who can be ignored (Dotan & Milli, 2019). The neopositivist claim that this is a technical problem solvable with "better data" fails to acknowledge that no neutral standard exists for determining what makes data "better." Every decision concerning data collection, labeling, and preprocessing involves normative judgments about which categories matter, which distinctions are significant, and what constitutes adequate representation (Balagopalan et al., 2023). The fact that these decisions can and must be made does not render them value-free; rather, it confirms that values are constitutive of scientific practice (G. M. Johnson, 2023).

Donna Haraway reminds us that all knowledge is produced from a particular location, and claims to a "view from nowhere" obscure the situatedness of knowledge producers. artificial intelligence systems developed primarily by Silicon Valley companies with demographically homogeneous teams reflect specific situated perspectives but these perspectives are masked as "objective" or "universal" through mathematical formalization (Burgess, 2022). Acknowledging that these systems are value-laden is not relativism but a prerequisite for genuine objectivity, which requires accountability for how situatedness shapes what can be seen (Rutting et al., 2023).

Artificial Intelligence Design Choices as Ineliminable Value Judgments

Every stage of artificial intelligence system development involves design decisions that cannot be determined solely by epistemic or technical considerations, but instead require value judgments about acceptable trade-offs and legitimate priorities. This challenges Reichenbach's distinction between the context of discovery and the context of justification, showing that "justification" itself relies on value-laden standards. Consider the choice between predictive accuracy and model interpretability. Complex deep-learning models often achieve higher accuracy but operate as opaque "black boxes." Simpler models are more interpretable but may be less accurate. No purely epistemic criterion can resolve these trade-offs; the choice depends on the application context and the values prioritized.

In medical diagnosis, interpretability may be prioritized because clinicians must understand the basis of decisions and patients have a right to explanations. In content recommendation systems, predictive accuracy may be favored because the consequences of error are perceived as less severe. But judgments about the "severity of consequences" are value judgments, not empirical facts. This confirms Longino's claim that scientific reasoning is never purely deductive from evidence to conclusion, but always relies on additional assumptions imbued with values.

Choices regarding objective functions in machine learning are also value-laden. Should a system minimize mean error, maximum error, or disparities in error across groups? Should it optimize for precision, recall, or some balance between the two? No "correct" answer can be derived from logic or evidence alone. Each choice encodes implicit values about fair distributions of risk, who may be harmed, and how much inequality is acceptable (Andrus et al., 2021).

Karl Popper argued that objectivity is achieved through intersubjective testing and critical traditions. Yet in artificial intelligence, the "critical tradition" itself is situated within specific communities with particular values. When artificial intelligence development communities are demographically or ideologically homogeneous, what counts as "legitimate criticism" or an "adequate evaluation metric" reflects community-specific values. This does not invalidate criticism; it merely confirms that no neutral standpoint exists from which to evaluate systems. Sandra Harding notes that marginalized standpoints may provide epistemic advantages by enabling critical scrutiny of dominant assumptions. In artificial intelligence, this is evident when researchers from underrepresented groups identify biases invisible to mainstream developers. The fact that different perspectives reveal different aspects of a system confirms that there is no "value-free objectivity," but only objectivity achieved through critical engagement with a plurality of perspectives (Hagendorff & Danks, 2022).

Paradigm Incommensurability in Artificial Intelligence Methodology

Kuhn argued that different scientific paradigms cannot be fully compared because they determine different meanings of terms, standards of evidence, and criteria of success. The use of artificial intelligence in science introduces a new methodological paradigm that demonstrates this form of incommensurability, thereby challenging the positivist image of science as rational progress guided by neutral epistemic standards (Norton, 2021). Data-driven machine-learning approaches differ fundamentally from traditional hypothesis-testing methodologies in their epistemology. Traditional science begins with theory or hypotheses derived from causal understanding that are then tested against data. Machine learning, by contrast, identifies patterns in data without assuming or producing explicit causal models.

Questions about which paradigm is "better" cannot be answered neutrally, because each paradigm determines what counts as an adequate explanation (Greene et al., 2022). For traditional paradigms, correlation without causation is epistemically unsatisfying; the goal is to understand underlying mechanisms. For machine-learning paradigms, predictive accuracy may suffice even without causal understanding. Their criteria for success differ: one emphasizes theoretical understanding, the other empirical performance. No neutral standpoint exists from which to judge which is "objectively" better, because such judgments presuppose standards drawn from one paradigm or the other (D. G. Johnson & Verdicchio, 2025).

Kuhn clarified that epistemic values such as accuracy, consistency, scope, simplicity, and fruitfulness play roles in theory choice, but these values can be weighted differently. In artificial intelligence, this appears in debates over whether a highly accurate but opaque model is preferable to a more interpretable but less accurate one (G. M. Johnson, 2023). Scientists may legitimately disagree about how to weight these values without being irrational, which confirms that choice involves values, not just facts or logic (Huang et al., 2022).

Incommensurability also arises in how paradigms define legitimate problems. Machine learning shifts causal questions such as "What causes this disease?" into predictive questions such as "Who is likely to develop this disease?" This transformation is not neutral but reflects values regarding which forms of knowledge are important. Predictive knowledge is useful for interventions but may overlook structural understandings necessary for systemic change (Andrus et al., 2021).

Situated Objectivity and the Necessity of Value Reflexivity

Feminist epistemology particularly Haraway's concept of "situated knowledge" and Harding's notion of "strong objectivity" offers the most productive framework for understanding artificial intelligence's implications for scientific objectivity. This framework rejects both objectivism, which claims a view from nowhere, and relativism, which denies the possibility of better knowledge. Artificial intelligence demonstrates the situatedness of knowledge in empirically inspectable ways. Every artificial intelligence system is produced from a particular location: developed by teams with specific demographic compositions, within institutional contexts with particular incentives, using data collected from particular populations, for applications serving particular interests (Burgess, 2022). This situatedness is not a defect to be eliminated but a condition that shapes what can be known (Zajko, 2021).

For example, widely used criminal-risk prediction systems in the U.S. justice system are trained on historical arrest data. These data reflect not "objective crime rates" but racially biased policing patterns. Systems trained on these data predict higher risk for Black individuals not because they are more likely to commit crimes, but because they are more likely to be arrested (Birhane, 2022). This is then used to justify increased surveillance of Black communities, producing more arrests, which "confirm" the system's predictions in a self-reinforcing cycle (Leslie, 2020). The neopositivist claim that this can be fixed with "better data" presupposes the existence of an "objective" measure of criminality independent of the social practices that produce arrest data. Yet standpoint theory shows that no access to social reality exists that is not mediated by situated social practices. All data about social phenomena are produced through processes that reflect and shape power relations (Balagopalan et al., 2023).

Harding's "strong objectivity" requires that instead of hiding situatedness, responsible knowledge practices acknowledge it explicitly and subject it to critical scrutiny from diverse perspectives. In artificial intelligence, this entails identifying embedded assumptions and values, analyzing who benefits and who is harmed, involving affected communities in system design and evaluation, and holding systems accountable for their social consequences (Hagendorff & Danks, 2022). Haraway argues that genuine objectivity requires "response-ability" the ability and willingness to respond to and account for knowledge claims (Dotan & Milli, 2019). In artificial intelligence, this means acknowledging that systems are not "objective fact-finders" but sociotechnical constructions embodying particular choices. Recognizing this does not undermine artificial intelligence's epistemic authority; rather, it situates that authority within a more epistemologically honest framework.

Literature advocating "value-aware" or "value-explicit" artificial intelligence proposes that system design and deployment should involve explicit deliberation about the values they serve. This may include: explicit value statements documenting assumptions and priorities; distributive impact analyses examining who benefits and who is harmed; participatory design processes involving diverse stakeholders; and ongoing audits assessing whether systems function in accordance with their stated values (Dignum, 2022). This position does not deny the possibility of better or worse, more or less accurate, more or less reliable knowledge. Instead, it argues that criteria for "better" cannot be derived from a neutral standpoint but must be articulated explicitly, publicly justified, and accountable to diverse communities. Objectivity is thus redefined not as the absence of values or perspectives, but as critical and reflexive engagement with a plurality of perspectives in the production of knowledge (Jedličková, 2025).

CONCLUSION

This study demonstrates that the ideal of value-free science is untenable, and that artificial intelligence makes the entanglement of values in knowledge production increasingly explicit. Four key findings support this conclusion: algorithmic bias, normative trade-offs inherent in artificial intelligence design, methodological shifts that are incommensurable with traditional paradigms, and the socially situated nature of artificial intelligence systems. These findings underscore the need for a form of science that is value-reflexive, transparent about its normative assumptions, and accountable for its social impacts. Rather than pursuing value-free objectivity, this study advocates for situated objectivity as a more realistic and ethical framework. Several limitations should be noted, including reliance on philosophical literature, the dominance of Western perspectives, the rapid evolution of artificial intelligence, and the possibility of oversimplifying philosophical positions. These limitations highlight the need for empirical studies and non-Western perspectives to broaden our understanding of the relationship between artificial intelligence, values, and knowledge.

REFERENCES

Alvarado, R. (2023). AI as an Epistemic Technology. *Science and Engineering Ethics*, *29*, 32. https://doi.org/10.1007/s11948-023-00451-3

Andrus, M., Dean, S., Gilbert, T. K., Lambert, N., & Zick, T. (2021). *AI Development for the Public Interest: From Abstraction Traps to Sociotechnical Risks*. https://doi.org/10.48550/arXiv.2102.04255

Balagopalan, A., Madras, D., Yang, D. H., Hadfield-Menell, D., & Ghassemi, M. (2023). Judging facts, judging norms: Training machine learning models to judge humans requires a modified approach to labeling data. *Science Advances*, *9*(19). https://doi.org/10.1126/sciadv.abq0701

Bar-Gil, O. (2025). Examining trends in AI ethics across countries and institutions via quantitative discourse analysis. *AI & Society*. https://doi.org/10.1007/s00146-025-02673-4

Baroud, N., Ardila, Y., Akmal, F., & Sabrina, R. (2025). Opportunities and Challenges for Islamic Education Teachers in Using Artificial Intelligence in Learning. *Muaddib: Journal of Islamic Teaching and Learning*, *1*(2), 1–11. https://doi.org/https://muaddib.intischolar.id/index.php/muaddib/article/view/6

Birhane, A. (2022). The unseen Black faces of AI algorithms. *Nature*, *610*(7932), 451–

452. https://doi.org/10.1038/d41586-022-03050-7

Brown, M. J. (2024). For values in science: Assessing recent arguments for the ideal of value-free science. *Synthese*, *204*(1), 1–15. https://doi.org/10.1007/s11229-024-04762-1

Bueter, A. (2022). Bias as an epistemic notion. *Studies in History and Philosophy of Science*, *91*, 307–315. https://doi.org/10.1016/j.shpsa.2021.12.002

Burgess, J. (2022). Everyday data cultures: beyond Big Critique and the technological sublime. *AI & Society*, *38*(2). https://doi.org/10.1007/s00146-022-01503-1

Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. *PLOS Digital Health*. https://doi.org/10.1371/journal.pdig.0000651

Deckker, D., & Sumanasekara, S. (2025). Bias in AI Models: Origins, Impact, and Mitigation Strategies. *Journal of Advanced Artificial Intelligence (JAAI)*, *3*(3), 234–247. https://doi.org/10.18178/JAAI.2025.3.3.234-247

Delgado-Chaves, F. M., Jennings, M. J., Atalaia, A., & Baumbach, L. (2025). Transforming literature screening: The emerging role of large language models in systematic reviews. *Proceedings of the National Academy of Sciences (PNAS)*, *122*(2), e2411962122–e2411962122. https://doi.org/10.1073/pnas.2411962122

Dhar, V. (2024). The Paradigm Shifts in Artificial Intelligence. *Communications of the ACM*, *67*(11), 50–59. https://doi.org/10.1145/3664804

Dignum, V. (2022). Responsible Artificial Intelligence – from Principles to Practice. In *arXiv preprint: Vol. arXiv:2205.10785*. arXiv. https://doi.org/10.48550/arXiv.2205.10785

Dotan, R., & Milli, S. (2019). Value-laden Disciplinary Shifts in Machine Learning. In *arXiv preprint arXiv:1912.01172*. https://doi.org/10.48550/arXiv.1912.01172

Douglas, H., & Branch, T. Y. (2024). The social contract for science and the value-free ideal. *Synthese*, *203*, 40. https://doi.org/10.1007/s11229-023-04477-9

Engkizar, E., Jaafar, A., Masuwd, M. A., Rahman, I., Datres, D., Taufan, M., Akmal, F., Dasrizal, D., Oktavia, G., Yusrial, Y., & Febriani, A. (2025). Challenges and Steps in Living Quran and Hadith Research: An Introduction. *International Journal of Multidisciplinary Research of Higher Education (IJMURHICA*, *8*(3), 426–435. https://doi.org/https://doi.org/10.24036/ijmurhica.v8i3.396

Engkizar, E., Jaafar, A., Sarianto, D., Ayad, N., Rahman, A., Febriani, A., Oktavia, G., Guspita, R., & Rahman, I. (2024). Analysis of Quran Education Problems in Majority Muslim Countries. *International Journal of Islamic Studies Higher Education*, *3*(1), 65–80. https://doi.org/https://doi.org/10.24036/insight.v3i1.209

Engkizar, E, Jaafar, A., Alias, M., Guspita, B., & Albizar, R. (2025). Utilisation of Artificial Intelligence in Qur'anic Learning: Innovation or Threat? *Journal of Quranic Teaching and Learning*, *1*(2), 1–17. https://joqer.intischolar.id/index.php/joqer/index

Engkizar, Engkizar, Jaafar, A., Taufan, M., Rahman, I., Oktavia, G., & Guspita, R. (2023). Quran Teacher: Future Profession or Devotion to the Ummah? *International Journal of Multidisciplinary Research of Higher Education (IJMURHICA)*, *6*(4), 196–210. https://doi.org/https://doi.org/10.24036/ijmurhica.v6i4.321

Engkizar, Engkizar, Muslim, H., Mulyadi, I., & Putra, Y. A. (2025). Ten Criteria for an Ideal Teacher to Memorize the Quran. *Journal of Theory and Research Memorization Quran*, *1*(1), 26–39. https://joqer.intischolar.id/index.php/joqer

Faqih, Nadaa, & Indallah, S. M. (2025). Standpoint Epistemology in Feminist Philosophy of Science: An Analysis of Sandra Harding's Thought. *Prosiding Konferensi Integrasi Interkoneksi Islam Dan Sains (KIIIS)*, *6*(1).

Ferrario, A., Facchini, A., & Termine, A. (2024). Experts or Authorities? The Strange Case of the Presumed Epistemic Superiority of Artificial Intelligence Systems. *Minds & Machines*, *34*, 30. https://doi.org/10.1007/s11023-024-09681-1

Greene, T., Dhurandhar, A., & Shmueli, G. (2022). Atomist or Holist? A Diagnosis and Vision for More Productive Interdisciplinary AI Ethics Dialogue. In *arXiv preprint arXiv:2208.09174*. https://doi.org/10.48550/arXiv.2208.09174

Hagendorff, T. (2024). Mapping the Ethics of Generative AI: A Comprehensive Scoping Review. *Minds & Machines*, *34*, 39. https://doi.org/10.1007/s11023-024-09694-w

Hagendorff, T., & Danks, D. (2022). Ethical and methodological challenges in building morally informed AI systems. *AI and Ethics*, *3*(2), 553–566. https://doi.org/10.1007/s43681-022-00188-y

Haryanto, R., & Paryanto, P. (2023). Thomas Kuhn's Scientific Revolution Paradigm and Its Relevance to Islamic Studies. *Zawiyah: Jurnal Pemikiran Islam*, *9*(2), (not provided)-(not provided). https://doi.org/10.31332/zjpi.v9i2.6904

Hasanzadeh, F., Josephson, C. B., Waters, G., Adedinsewo, D., Azizi, Z., & White, J. A. (2025). Bias recognition and mitigation strategies in artificial intelligence healthcare applications. *Npj Digital Medicine*, *8*, 154. https://doi.org/10.1038/s41746-025-01503-7

Huang, L. T.-L., Chen, H.-Y., Lin, Y.-T., Huang, T.-R., & Hung, T.-W. (2022). Ameliorating Algorithmic Bias, or Why Explainable AI Needs Feminist Philosophy. *Feminist Philosophy Quarterly*, *8*(3/4). https://doi.org/10.5206/fpq/2022.3/4.14347

Jedličková, A. (2025). Ethical approaches in designing autonomous and intelligent systems: a comprehensive survey towards responsible development. *AI & Society*, *40*(4), 2703–2716. https://doi.org/10.1007/s00146-024-02040-9

Johnson, D. G., & Verdicchio, M. (2025). The sociotechnical entanglement of AI and values. *AI & Society*, *40*(1), 67–76. https://doi.org/10.1007/s00146-023-01852-5

Johnson, G. M. (2023). Are Algorithms Value-Free? Feminist Theoretical Virtues in Machine Learning. *Journal of Moral Philosophy*, *21*(1–2), 27–61. https://doi.org/10.1163/17455243-20234372

Kassymova, G. K., Talgatov, Y. K., Arpentieva, M. R., Abishev, A. R., & Menshikov, P. V. (2025). Artificial Intelligence in the Development of the Theory and Practices of Self-Directed Learning. *Multidisciplinary Journal of Thought and Research*, *1*(3), 66–79. https://mujoter.intischolar.id/index.php/mujoter/article/view/19

Khan, A. A., Badshah, S., Liang, P., Waseem, M., Khan, B., Ahmad, A., Fahmideh, M., Niazi, M., & Akbar, M. A. (2022). Ethics of AI: A Systematic Literature Review of Principles and Challenges. *EASE '22: Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*, 383–392. https://doi.org/10.1145/3530019.3531329

Leslie, D. (2020). Understanding bias in facial recognition technologies. In *arXiv preprint arXiv:2010.07023*. https://doi.org/10.48550/arXiv.2010.07023

Monaro, S., Gullick, J., & West, S. (2022). Qualitative Data Analysis for Health Research: A Step-by-Step Example of Phenomenological Interpretation. *The Qualitative Report*, *27*(4), 1040–1057. https://doi.org/10.46743/2160-3715/2022.5249

Mitra, S. (2020). An Analysis of the Falsification Criterion of Karl Popper: A Critical Review. *Tattva Journal of Philosophy*, *12*(1), 1-18. https://doi.org/10.12726/tjp.23.1

Nazer, L. H., Zatarah, R., Waldrip, S., Chen Ke, J. X., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*. https://doi.org/10.1371/journal.pdig.0000278

Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.

https://doi.org/10.2307/j.ctv25wxcb5

Parker, W. S. (2024). The Epistemic Projection Approach to Values in Science. *Philosophy of Science*, *91*(1), 18–36. https://doi.org/10.1017/psa.2023.107

Politi, V. (2024). The value-free ideal, the autonomy thesis, and cognitive diversity. *Synthese*, *204*(1), 1–21. https://doi.org/10.1007/s11229-024-04673-1

Resnik, D. B., & Hosseini, M. (2025). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI Ethics*, *5*((Issue date: April 2025)), 1499–1521. https://doi.org/10.1007/s43681-024-00493-8

Resnik, D. B., Hosseini, M., & Hauswald, R. (2025). *Autonomous Artificial Intelligence, Scientific Research, and Human Values* (2025th-06–13th ed.). https://doi.org/10.31219/osf.io/b4m25_v1

Rutting, L., Vervoort, J., Mees, H., Pereira, L., Veeger, M., Muiderman, K., Mangnus, A., Winkler, K., Olsson, P., Hichert, T., Lane, R., Bottega Pergher, B., Christiaens, L., Bansal, N., Hendriks, A., & Driessen, P. (2023). Disruptive seeds: a scenario approach to explore power shifts in sustainability transformations. *Sustainability Science*, *18*(3), 1117–1133. https://doi.org/10.1007/s11625-022-01251-7

Schintler, L. A., McNeely, C. L., & Witte, J. (2023). *A Critical Examination of the Ethics of AI-Mediated Peer Review*. https://doi.org/10.48550/arXiv.2309.12356

Stamenkovic, P. (2024). Straightening the 'value-laden turn': minimising the influence of extra-scientific values in science. *Synthese*, *203*, 20. https://doi.org/10.1007/s11229-023-04446-2

Trächtler, J. (2024). The world as witty agent—Donna Haraway on the object of knowledge. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1389575

Vagelli, M. (2024). Styles of Science and the Pluralist Turn: Between Inclusion and Exclusion. *Revue de Synthèse*, *145*(3–4), 1–39. https://doi.org/10.1163/19552343-14234053

Ward, Z. B. (2021). On value-laden science. Studies in History and Philosophy of Science Part A, 85, 54-62. https://doi.org/10.1016/j.shpsa.2020.09.006

Widder, D. G. (2024). Epistemic Power in AI Ethics Labor: Legitimizing Located Complaints. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 1295–1304. https://doi.org/10.1145/3630106.3658973

Ye, A. (2024). And Then the Hammer Broke: Reflections on Machine Ethics from Feminist Philosophy of Science. *ArXiv*. https://doi.org/10.48550/arXiv.2403.05805

Zajko, M. (2021). Conservative AI and social inequality: conceptualizing alternatives to bias through social theory. *AI & Society*, *36*(3), 1047–1056. https://doi.org/10.1007/s00146-021-01153-9

Zanetti, L., Chiffi, D., & Petrini, L. (2023). Epistemic and Non-epistemic Values in Earthquake Engineering. *Science and Engineering Ethics*, *29*, 18. https://doi.org/10.1007/s11948-023-00438-0