philosophies

MDPI

Article

Operationalizing Pluralist AI Governance with the Integrated Axiology–MCDA Framework

Fei Sun [1,*], Damir Isovic [1] and Gordana Dodig-Crnkovic [2]

[1] Department of Computer Science, Mälardalen University, 721 26 Västerås, Sweden; damir.isovic@mdu.se
[2] Chalmers University of Technology, 412 96 Gothenburg, Sweden; gordana.dodig-crnkovic@chalmers.se
* Correspondence: fei.sun@mdu.se

Abstract

AI systems generate ethical tensions that cannot be addressed through principle-based guidance alone. This paper brings forward an Integrated Axiology–MCDA Framework for AI ethics that distinguishes intrinsic, instrumental, and relational values and uses multi-criteria analysis to operationalize value pluralism in practice. The framework structures ethical evaluation by making value commitments explicit, enabling transparent examination of trade-offs, and supporting context-sensitive judgment. A healthcare hyper-scenario with sensitivity analysis shows how different weight configurations influence the relative acceptability of diagnostic systems and clarifies the thresholds at which accuracy considerations outweigh privacy or fairness. Cross-domain applications in education, criminal justice, and finance further illustrate how domain-specific value tensions require distinct criteria sets and weighting structures. The analysis shows that ethical challenges in AI arise from genuine value pluralism. Explicit value classification enables more accountable decision making across the AI lifecycle.

Keywords: AI ethics; axiology; value pluralism; relational values; multi-criteria decision analysis; sensitivity analysis

1. Introduction

Artificial intelligence has moved from a future possibility to an established part of contemporary life. The rapid spread of AI systems across multiple sectors has raised urgent questions about ethics, governance, and the role of human values. Although AI ethics frameworks articulate principles such as fairness, transparency, accountability, and dignity, the practical implementation often exposes tensions and conflicts that cannot be resolved at the level of principles alone [1–4]. These challenges become apparent when applied in domain-specific practice. In healthcare, expanding data access may improve predictive performance while simultaneously increasing privacy exposure and governance burdens [5]. In policing and public safety, predictive systems can be framed as efficiency-enhancing, but they may entrench historically uneven patterns of surveillance and enforcement [6]. Similar tensions appears in other contexts as well: efficiency may clash with human oversight, rapid automation can undermine procedural fairness, and requirements for openness may conflict with proprietary interests [7,8].

Established ethical traditions remain essential, but each introduces blind spots when applied independently to complex AI governance contexts. Deontological theories specify non-negotiable duties; however, such duties can conflict and prove difficult to harmonize in practice [9]. Utilitarian approaches enable assessment of aggregate outcomes, though they

Academic Editor: Soraj Hongladarom

Received: 13 April 2026
Revised: 26 May 2026
Accepted: 29 May 2026
Published: 8 June 2026

risk providing inadequate safeguards for vulnerable or minority groups [10]. Virtue ethics emphasizes moral character, judgment, and practical wisdom, but offers limited procedural guidance for institutional structures or technical system design [11]. In line with broader critiques of AI ethics guidance, these limitations indicate that lists of principles alone do not constitute a sufficient method for managing the conflicts that arise during real-world deployment [12].

To address this problem, this paper proposes an Integrated Axiology–MCDA Framework for AI ethics and examines its application through Multi-Criteria Decision Analysis (MCDA). In this paper, axiology refers to the philosophical study of values and, more specifically, to the systematic identification, classification, and justification of the values that should guide AI design, deployment, and governance. Value pluralism refers to the view that several legitimate values may matter at the same time, that these values cannot always be reduced to a single master value or common metric, and that conflicts among them may require contextual judgment rather than purely technical optimization. MCDA is used here as a structured method for comparing alternatives across multiple criteria while making the underlying value assumptions and trade-offs explicit. The framework distinguishes intrinsic values: dignity, justice, and autonomy; instrumental values: performance, efficiency, and scalability; and relational values: trust, accountability, and intelligibility. By explicitly structuring these value dimensions, the approach supports plural normative perspectives, transparent trade-off analysis, and stakeholder-oriented deliberation in context-specific governance processes [13,14].

The framework is evaluated through philosophical analysis, MCDA modelling, sensitivity testing, and applications across multiple societal domains. The analysis shows how different value configurations influence outcomes in fields including healthcare, education, criminal justice, and finance. The concluding section evaluates the methodological strengths and limitations of the approach and identifies implications for future research and practice in AI governance. The framework does not aim to resolve value conflicts through calculation. Instead, it provides a structured way to make such conflicts explicit, to support deliberation, and to clarify the consequences of different value commitments.

2. Philosophical Foundations of Axiology

The philosophical study of values has a long history. Axiology emerged as a distinct field in the early twentieth century [15,16]. This section reviews major developments in axiological theory and outlines their significance for AI ethics.

2.1. Classical Foundations: Moore and Ross

G. E. Moore's Principia Ethica (1903) introduced several foundational ideas that continue to shape modern axiology. Moore argued that the concept of good is simple, unanalyzable, and cannot be reduced to any natural properties. Its nature is grasped through rational intuition rather than empirical observation [17]. He distinguished intrinsic goods, which possess value in themselves, from instrumental goods, which derive value from the ends they promote. To assess intrinsic value, Moore proposed the isolation test, which considers whether something would remain valuable if it existed entirely on its own, independent of any consequences. On this basis, he concluded that certain forms of conscious experience, such as aesthetic appreciation and personal affection, have intrinsic value.

Moore's position was later refined by W. D. Ross, who developed a pluralist theory of moral duties. Ross maintained that human agents are guided by several prima facie duties, which include fidelity, reparation, gratitude, justice, beneficence, self-improvement, and non-maleficence [18]. These duties can conflict, and no single principle can determine their

priority in all circumstances. Judgment is therefore required to decide which duty is most important in a given situation.

Ross's pluralist view of morality helps explain key challenges in the ethics of artificial intelligence. Because he argues that several moral values matter at the same time, conflicts between them are unavoidable. AI systems show this clearly: they often create tensions between privacy and security, autonomy and welfare, or innovation and precaution. From an axiological perspective, these tensions do not result from poor system design but from the basic fact that important values sometimes collide. Ross's position therefore suggests that such conflicts cannot be fully removed. Instead, they must be handled through careful judgment in each specific case.

2.2. Contemporary Value Pluralism

Value pluralism is a major position within axiology. Later developments in value theory further expanded the foundations of pluralist axiology. Isaiah Berlin argued that many central human values are both incommensurable and sometimes incompatible [19]. Berlin also emphasized that some value conflicts are tragic in the sense that they cannot be resolved without genuine loss, even when all competing values are fully legitimate. Values are incommensurable when they cannot be measured on a shared scale, and they are incompatible when the realization of one necessarily restricts the realization of another. Berlin used the relation between liberty and equality to illustrate this point and rejected the idea that all values can be derived from a single overarching principle.

Joseph Raz extended this analysis by distinguishing between incommensurability and incomparability [20]. According to Raz, values may lack a common measure but still allow for rough or partial comparisons. He also emphasized that incommensurability does not guarantee comparability; it only means that the absence of a shared scale does not by itself rule out the possibility of rational comparison. Ruth Chang developed this line of thought through the concept of values being on a par, meaning that two options may be roughly comparable without being precisely equal, better, or worse [21]. When values stand in this relation, rational choice requires judgment that cannot be reduced to mechanical calculation.

Martha Nussbaum's capabilities approach offers another influential form of pluralistic value theory [22,23]. She identifies several central human capabilities, including bodily integrity, practical reason, affiliation, and control over one's environment. These capabilities are distinct and only partly substitutable. She also argues that each capability must reach a minimum threshold for a society to count as just. A society that develops some capabilities while neglecting others therefore cannot be said to support human flourishing in a comprehensive or equitable way.

These contemporary theories underline several points of particular relevance for AI ethics. Conflicts between values in the development and use of AI systems often reflect genuine tensions in human value frameworks rather than technical failures. Ethical evaluation therefore requires contextual reasoning that takes into account specific institutional and social settings. Moreover, different communities can reasonably prioritize values in different ways, suggesting that the ethical governance of AI must remain sensitive to cultural and political variation. This introduces a methodological tension for any operational framework: if values are incommensurable, then any formal aggregation risks oversimplification. The present approach does not eliminate this tension but makes it explicit, using formal structure to support, rather than replace, context-sensitive judgment.

2.3. Axiology and AI Ethics

Axiology offers a structured way to understand the values that shape AI systems and addresses several limitations of principle-based ethical frameworks. Existing AI ethics guidelines often group concerns under principles such as fairness, accountability, transparency, and privacy [24,25]. Although these principles are valuable, they often provide limited clarification of how they relate to one another, leave underlying value assumptions unstated, and offer only partial guidance for managing conflicts when principles pull in different directions [1,26].

An axiological approach responds by shifting attention from enumerated principles to the value commitments that support them. This perspective recognizes that values are plural and sometimes in tension, and that such tensions cannot be resolved through technical solutions alone. Ethical evaluation therefore requires processes in which stakeholders identify and justify the values at stake and consider how those values should be interpreted within specific contexts. This approach makes the basis of disagreement more transparent and supports more careful reasoning about trade-offs.

Within this framework, methods such as Multi-Criteria Decision Analysis (MCDA) can provide formal support for structured deliberation. While some forms of MCDA aggregate values into composite measures, its primary contribution lies in organizing diverse criteria, making value assumptions explicit, and clarifying their implications for decision making [27,28]. An axiological foundation therefore enables more transparent and context-sensitive governance by linking ethical evaluation to explicit and accountable value commitments.

2.4. A Tripartite Classification of Values for AI Ethics

Building on the axiological foundations outlined above, values relevant to AI ethics can be organized into three categories: intrinsic, instrumental, and relational values. The specific values listed under each category were selected through three criteria: their recurrence in established AI ethics and governance literature, their fit with the normative role assigned to each value type, and their practical relevance across the domains examined in this paper. This classification helps clarify the different roles that values play across the AI lifecycle, from design and development to deployment, monitoring, and governance. The categories are analytically distinct, but they are not always mutually exclusive. Some values, such as explainability or transparency, may operate across categories depending on how they are interpreted. The framework should therefore be understood as a heuristic for analysis rather than a rigid taxonomy. The following subsections explain how each value type functions within the framework.

2.4.1. Intrinsic Values

Intrinsic values are those that should be respected for their own sake. In AI contexts, central intrinsic values include dignity, justice, autonomy, and privacy. These values function mainly as moral constraints, setting limits on what may be pursued in the name of efficiency or other collective goals. They reflect the idea that people must be treated as ends in themselves rather than merely as means [29]. Intrinsic values, therefore, define the basic normative boundaries for how AI systems should be designed, implemented, and used. For this reason, intrinsic values are not treated as optional preferences in the MCDA process and cannot be assigned a weight of zero; alternatives that violate minimum intrinsic-value constraints should be excluded or reconsidered before ordinary aggregation is applied.

2.4.2. Instrumental Values

Instrumental values concern the effectiveness of means in achieving ethically justified aims. In AI systems, key instrumental values include accuracy, efficiency, reliability, and scalability. These values matter for system performance, but their ethical importance depends on the purposes they serve and on whether they remain consistent with intrinsic constraints. Instrumental values can support intrinsic values, for example when accurate medical diagnosis contributes to fairness in clinical decision making and improves patient outcomes. However, they may also conflict with intrinsic or relational values, such as when efficiency compromises privacy or accountability. Making these relationships explicit helps ensure that technical optimization does not override broader ethical commitments.

2.4.3. Relational Values

The concept of relational values has been developed primarily within environmental ethics. It refers to values that arise from relationships between people, communities, and the natural world rather than from intrinsic properties or instrumental use [30–33]. This perspective extends traditional value theory by highlighting the importance of social context, shared practices, and collective meanings. It also emphasizes that values may be grounded in relationships of care, responsibility, and interdependence rather than solely in individual preferences or ecological functions. Relational values are dynamic and context-dependent, shaped through ongoing engagement and changing with the relationships in which they are embedded.

In applying this idea beyond environmental contexts, relational values can also be understood within sociotechnical systems such as AI. Relational values arise from the interactions between individuals, institutions, and AI systems. Examples include trust, accountability, transparency, and meaningful participation. These values are not captured solely by intrinsic rights or by performance metrics. Instead, they describe the social and institutional conditions under which AI systems are viewed as legitimate and aligned with the interests of affected stakeholders [34,35].

This interpretation is consistent with work on relational autonomy and trust [36,37], which similarly understand values as emerging within social relationships rather than residing solely in individuals or technical artefacts. It is also aligned with approaches such as value-sensitive design, which emphasize the importance of social context and stakeholder experience in technological systems [38]. Relational values therefore highlight the role of governance structures, institutional design, and stakeholder engagement in shaping ethical outcomes. Within the present framework, relational values play a dual role. They function both as evaluative criteria and as conditions for the legitimacy of the decision process itself, since trust, accountability, and participation shape how value weights are assigned and interpreted.

2.4.4. Advantages of the Tripartite Structure

The tripartite structure extends the traditional intrinsic–instrumental distinction by including values, such as trust and accountability, that do not fall cleanly into either type. Its main advantage is analytical: it separates values that set moral boundaries, values that support effective action, and values that sustain legitimate sociotechnical relationships. This distinction clarifies how value priorities vary across different contexts and points in the AI lifecycle, while also showing how institutional design supports relational values and influences ethical outcomes. Table 1 summarizes the tripartite value classification and provides examples of how each value type can be instantiated across domains.

Table 1. Tripartite value classification with domain examples.

Value Type	Core Examples	Domain-Specific Instantiation
Intrinsic	Dignity, Justice, Autonomy, Privacy	Healthcare: Respect for patient autonomy in treatment decisions; Education: Equal access to learning opportunities; Criminal justice: Protection of due process rights; Finance: Fair access to financial services
Instrumental	Accuracy, Efficiency, Reliability, Scalability	Healthcare: Diagnostic accuracy and cost-effectiveness; Education: Improvement of learning outcomes; Criminal justice: Predictive accuracy in risk assessment; Finance: Fraud detection and transaction efficiency
Relational	Trust, Accountability *, Transparency *, Participation	Healthcare: Trust and coordination among clinicians, patients, and AI systems; Education: Interaction between teachers, students, and digital platforms; Criminal justice: Public legitimacy of algorithmic decisions; Finance: Consumer trust in automated decision systems

Note: Values marked with * may span multiple categories depending on their interpretation and use.

3. Normative Foundations: Responsible AI and Digital Humanism

3.1. Responsible AI: From Principles to Embedded Ethics

Responsible AI has become a central framework for the governance of artificial intelligence, expressed in national and supranational guidelines, industry codes of conduct, and academic proposals [13,39]. It identifies several properties considered necessary for ethically acceptable systems, including transparency, accountability, fairness, safety, and human oversight. The European Union Artificial Intelligence Act (Regulation (EU) 2024/1689) is currently the most comprehensive regulatory instrument and implements these commitments through risk-based classification, conformity assessment procedures, and documentation requirements for high-risk applications [40].

Despite these developments, Responsible AI faces a well-documented gap between principles and practice [3,4,12,26]. The high-level formulation of principles allows broad support but provides limited guidance when concrete conflicts arise. Situations in which transparency conflicts with privacy, or in which fairness across different demographic groups cannot all be satisfied simultaneously, illustrate the limits of relying on principles alone [41]. The ethics-by-design agenda [42] seeks to integrate ethical reasoning into development processes, but it requires more precise methodological tools than most existing frameworks supply.

The present framework addresses this need by grounding Responsible AI in axiology and supporting practical reasoning through multi-criteria decision analysis (MCDA). Axiology clarifies the values to which stakeholders are committed, and MCDA provides a structure for examining how these commitments should guide specific decisions. This approach does not replace principles; instead, it translates them into a process that makes value assumptions explicit and enables transparent, revisable judgment.

3.2. Digital Humanism: Technology in Service of Human Flourishing

Digital Humanism provides a complementary normative orientation, framing AI in relation to human dignity, autonomy, and democratic participation. The Vienna Manifesto on Digital Humanism (2019) calls for digital technologies to support human flourishing rather than reducing persons to data points in automated processes [43]. This perspective draws on a broad intellectual tradition, including critical accounts of technological rationality [44], theories of communicative agency [45], and philosophical accounts of human self-determination [46].

Digital Humanism contributes to the axiological framework in several ways. It maintains that the identification of values for AI governance cannot be left solely to technical

experts and requires deliberation among those affected by these systems. This supports a participatory approach to assigning value weights within MCDA. Digital Humanism also rejects the notion that AI systems can be value-neutral [44,46], reinforcing the view that value classification and weighting must be explicit and justified. Its emphasis on human dignity as an inviolable constraint aligns with the axiological treatment of intrinsic values as non-negotiable moral limits. Together, these points strengthen the normative foundation for ethical AI governance.

3.3. Ethical Pluralism as the Integrating Principle

Axiology, Responsible AI, and Digital Humanism all share a commitment to ethical pluralism. Ethical pluralism holds that several values, each important in their own right, must be considered when evaluating AI systems. This reflects the fact that these systems operate in complex moral environments.

Ethical monism, by contrast, assumes that a single value or principle can solve all moral problems. Although simple, this view does not work in practice. It risks overlooking values important to less powerful groups and ignores the reality that legitimate values often conflict [19]. Ethical pluralism recognizes these conflicts and treats them as real ethical challenges that must be addressed rather than removed.

For a framework grounded in axiology, the key task is to create structured and transparent ways to deliberate among competing values in a manner that is legitimate and open to democratic oversight. Combining axiological analysis with MCDA supports this task by making value commitments explicit and enabling reasoned comparison across them. These perspectives provide the normative content that the Axiology–MCDA Framework operationalizes: axiology clarifies what values are at stake, Responsible AI identifies governance expectations, and Digital Humanism situates these within a broader commitment to human-centered technological development.

4. The Integrated Axiology–MCDA Framework

The Integrated Axiology–MCDA Framework connects philosophical analysis of value with the formal tools of Multi-Criteria Decision Analysis (MCDA). By combining the tripartite value classification with a structured decision method, the framework operationalizes value pluralism in a transparent and systematic way. It is designed as a deliberative cycle that supports ethical evaluation across the entire AI lifecycle, from initial design to deployment, monitoring, and eventual decommissioning.

4.1. Architectural Components and Normative Foundations

The framework consists of four connected components:

1. Normative foundation: This draws on axiology for a theory of value, on Responsible AI for governance principles, and on Digital Humanism for commitments to human agency and democratic participation. It defines the ethical space in which trade-offs are evaluated.
2. Value classification: As outlined in Section 2.4, values are grouped into intrinsic, instrumental, and relational categories. This taxonomy provides a shared vocabulary for deliberation and helps prevent the flattening of values into a single dimension.
3. MCDA operational method: MCDA offers the structure needed to compare alternatives across multiple criteria. Alternatives $A = \{A_1, \ldots, A_m\}$ are evaluated across criteria $C = \{C_1, \ldots, C_n\}$ using a performance matrix.
4. Ethical outcome and review: The ranking of alternatives is documented as an ethics audit trail and is used for iterative refinement as systems are deployed and new evidence emerges.

4.2. Axiological Adaptation of MCDA

Ethical evaluation requires adapting standard MCDA practices:

- Intrinsic constraints: Intrinsic values such as dignity and justice function as normative constraints rather than ordinary criteria. They define admissible regions of the decision space and may be implemented through threshold conditions, lexicographic ordering, or exclusion of alternatives that violate these constraints prior to aggregation.
- Ordinal evaluation: Many ethical considerations do not admit precise numerical measurement. Ordinal scoring (for example, 1 to 5) is used to reflect qualitative judgments from experts and stakeholders.
- Context sensitivity: Value priorities vary across domains. For this reason, the framework does not seek a universal set of weights but evaluates each context separately.

4.3. Formal Structure

The weighted-sum model is selected not for mathematical completeness but for transparency and accessibility in participatory contexts, where interpretability is essential for stakeholder deliberation. The overall score for an alternative A_i is:

$$S(A_i) = \sum_{j=1}^{n} w_j x_{ij}, \tag{1}$$

subject to:

$$\sum_{j=1}^{n} w_j = 1, \qquad w_j > 0 \text{ for intrinsic criteria.} \tag{2}$$

The aggregation model applies within the space of ethically admissible alternatives defined by intrinsic constraints.

In this formula, $S(A_i)$ denotes the aggregate score assigned to alternative A_i; w_j denotes the weight assigned to criterion C_j; and x_{ij} denotes the score or performance of alternative A_i on criterion C_j. The condition $\sum_{j=1}^{n} w_j = 1$ normalizes all criterion weights, while $w_j > 0$ for intrinsic criteria expresses the assumption that intrinsic values cannot be ignored or weighted to zero when they define ethical admissibility.

Other methods, such as the Analytic Hierarchy Process or outranking approaches [47,48], offer more sophisticated formal tools than a simple weighted-sum model, but may be less accessible for broad stakeholder deliberation.

4.4. Methodological Procedure

The framework is implemented through a seven-stage process, shown in Figure 1. It is designed to be iterative: the final stage often leads to refinement of earlier stages as new evidence or perspectives emerge.

1. Context specification: Define the decision setting, stakeholders, and relevant AI alternatives.
2. Criteria identification: Identify the intrinsic, instrumental, and relational values relevant to the context.
3. Performance assessment: Score each alternative on each criterion using evidence or expert judgment.
4. Weight elicitation: Deliberate on the relative importance of each criterion and assign weights.
5. Scoring and ranking: Aggregate scores using the weighted-sum model.
6. Sensitivity analysis: Test how robust the ranking is to changes in value weights. Identify tipping points at which preferences shift.

7. Reasoned judgment: Interpret results, record the ethical reasoning, and consider whether the process should be repeated with revised assumptions.

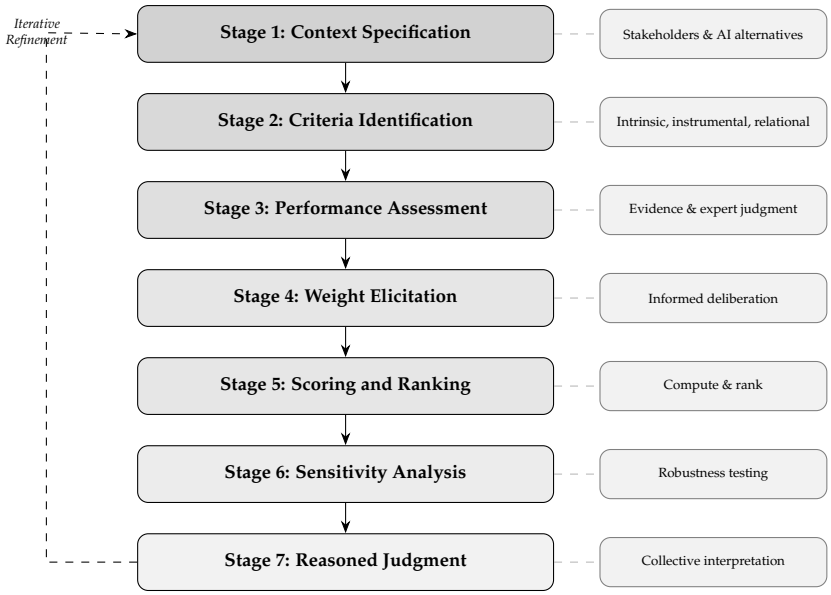

Figure 1. Seven-stage methodological procedure for integrated axiology–MCDA framework.

Sensitivity analysis is central to responsible use of the framework. By examining how rankings change with shifts in value weights, stakeholders can understand not only which alternative is preferred, but also why. This supports transparency, accountability, and defensible decision making across the AI lifecycle.

The role of MCDA in this framework requires clarification. It is not intended to compute or resolve ethical conflicts. Given the incommensurability of many values, no algorithmic procedure can determine a single correct outcome. Instead, MCDA provides a structured representation of value commitments that supports deliberation among stakeholders by making trade-offs explicit, comparable, and open to justification within specific sociotechnical contexts. Within this process, different forms of knowledge are integrated. Expert assessments are essential for evaluating system behavior, feasibility, and potential impacts, while stakeholders contribute perspectives on value priorities, social context, and acceptable trade-offs. The framework therefore mediates between technical expertise and value pluralism rather than privileging a single source of authority.

5. Case Study: AI Diagnostics in Healthcare

5.1. Scenario Description

This section presents a hypothetical case study involving two artificial intelligence diagnostic systems used in healthcare. The case combines realistic features of clinical practice with controlled variations designed to make ethical tensions analytically visible. The aim is not to reproduce empirical conditions, but to illustrate how the axiological and Multi-Criteria decision analysis (MCDA) framework supports context-sensitive ethical evaluation.

- System A: An accuracy-optimized deep learning model trained on extensive patient data, including genomics, electronic health records, and family history. Its performance depends on broad data access and is therefore associated with increased privacy risk. The training dataset is large but demographically imbalanced, leading to uneven performance across groups.

- System B: A privacy-enhancing diagnostic system employing consent-driven data minimization and differential privacy techniques. Its restricted data access reduces training richness, lowering accuracy when compared to System A, but it offers stronger privacy guarantees and more equitable demographic performance.

The evaluation covers five clinical contexts: routine outpatient care, emergency medicine, pediatrics, geriatric care, and community health. These contexts differ in urgency, vulnerability, and operational demands, which leads to different priorities among accuracy, privacy, and fairness for each stakeholder group.

It also involves three stakeholder groups. Each stakeholder group prioritizes different values depending on the specific clinical context.

- Clinicians prioritize diagnostic accuracy and clinical safety, especially in high-stakes contexts such as emergency medicine and pediatric critical care.
- Patient advocates emphasize privacy, autonomy, and fairness, particularly in community health, pediatrics, and geriatric care where vulnerability and equity concerns are central.
- Public health officials balance accuracy with community trust, especially in routine outpatient and community health contexts where acceptance of AI systems shapes effective deployment.

5.2. MCDA Analysis

Three ethical criteria are included in the analysis: Accuracy (Acc), Privacy (Priv), and Fairness (Fair). Fairness is defined as demographic performance equity, and System B scores higher because of its more balanced training data. Using a 1–4 performance scale, System A scores 4 on Accuracy, 2 on Privacy, and 2 on Fairness; System B scores 3, 4, and 4, respectively.

Illustrative stakeholder-informed weight vectors and resulting ethical scores across clinical contexts are presented in Table 2. For simplicity, relational values (such as clinician trust and perceived legitimacy) are not modelled as separate criteria here, though they could be included in more extensive applications.

Table 2. Ethical evaluation of AI diagnostic systems across five clinical contexts.

Clinical Context	$w_{\mathrm{Acc}}/w_{\mathrm{Priv}}/w_{\mathrm{Fair}}$	$S(A)$	$S(B)$	Preferred
Routine Outpatient Clinic	0.50/0.25/0.25	3.0	3.5	B
Emergency Department	0.70/0.15/0.15	3.4	3.3	A
Community Health	0.30/0.35/0.35	2.6	3.7	B
Paediatric Clinic	0.40/0.30/0.30	2.8	3.6	B
Geriatric Care	0.30/0.35/0.35	2.6	3.7	B

The results reveal a consistent pattern: System B is preferred in four of the five clinical contexts, where privacy and fairness carry substantial weight. System A is preferred only in the emergency department, where the weight on accuracy is sufficiently high to offset System B's advantages in privacy and fairness. Even in this case, System A's margin is narrow (3.4 vs. 3.3), indicating that the result depends on sustaining a large accuracy differential and could be reversed by modest improvements in System B.

5.3. Sensitivity Analysis

Sensitivity analysis evaluates how the outcome of a decision model changes when the importance assigned to different values is varied, and is a standard step in multi-criteria decision analysis for assessing the robustness of rankings [47,49]. Sensitivity analysis tests how the preferred system changes under variation in value weights. Table 3 reports

results for four weighting schemes in the outpatient context, varying the weight assigned to Fairness.

Table 3. Sensitivity analysis: ethical scores under varying weight configurations (outpatient context).

Criteria Set	w_{Acc}	w_{Priv}	w_{Fair}	$S(A)/S(B)$
Baseline (Acc/Priv only)	0.50	0.50	—	3.0/3.5 *
Add Fairness (equal)	0.34	0.33	0.33	2.7/3.7 *
Accuracy-prioritised	0.60	0.20	0.20	3.2/3.4 *
Fairness-prioritised	0.20	0.20	0.60	2.4/3.8 *

* Preferred system under each weighting configuration.

System B is preferred in all four weighting schemes, including the accuracy-prioritized case ($w_{\text{Acc}} = 0.60$). This shows that the preference for System B is robust across balanced, accuracy-oriented, and fairness-oriented configurations. The point at which System A becomes preferred lies at approximately $w_{\text{Acc}} > 0.67$, assuming equal weights on Privacy and Fairness. This threshold is ethically informative: it identifies how strongly accuracy must be prioritized before its benefits are sufficient to outweigh the privacy and fairness disadvantages associated with System A.

5.4. Insights and Limitations

This case study illustrates several key features of the Axiology–MCDA Framework:

- Context-sensitivity: No system is superior in all settings; the appropriate choice depends on clinical context and stakeholder values.
- Value transparency: Explicit weight vectors allow stakeholders to examine their own priorities and understand the ethical trade-offs involved.
- Decision robustness: Sensitivity analysis reveals the stability of conclusions across reasonable variations in value weights.

The case study also has limitations. The performance scores are hypothetical rather than empirically validated; real deployment would require detailed clinical trials and privacy audits. The weight vectors are illustrative and not derived from stakeholder deliberation; rigorous applications would require structured value elicitation. Moreover, the analysis considers only two AI systems; real procurement processes typically involve multiple alternatives. These limitations suggest a research agenda for empirical validation, which is outlined in Section 8.

6. Cross-Domain Applications

A key test of the Axiology–MCDA Framework is its ability to support ethical reasoning across the different domains in which AI systems are now used. This section applies the framework to education, criminal justice, and finance, showing how domain-specific value tensions shape criteria selection and how axiological analysis clarifies the ethical priorities appropriate to each setting. Table 4 summarizes the cross-domain axiological analysis.

Table 4. Cross-domain axiological analysis.

Domain	Key Ethical Tensions	Dominant Intrinsic Values	Primary Instrumental Values
Healthcare	Accuracy vs. Privacy	Dignity, Autonomy, Beneficence	Accuracy, Efficiency
Education	Personalization vs. Equity	Fairness, Inclusion, Autonomy	Accuracy, Scalability
Criminal Justice	Efficiency vs. Due Process	Justice, Dignity, Non-discrimination	Predictive Accuracy, Cost
Finance	Profitability vs. Fairness	Fairness, Transparency, Trust	Efficiency, Scalability

6.1. Education

AI systems in education are used for adaptive learning, automated essay scoring, early-warning systems, and admissions screening. The central ethical tension is between personalization, which relies on detailed individual data, and equity, which seeks to prevent the reproduction of existing educational inequalities [50]. Empirical studies show that educational AI tools often perform unevenly across racial, socio-economic, and disability groups, making fairness and inclusion key intrinsic values in this domain [51]. Autonomy is also important, since AI-mediated learning environments affect the agency of both teachers and students.

Applying the Axiology–MCDA Framework in educational contexts requires criteria that reflect these priorities. Suitable criteria include prediction accuracy, equity across student groups, strength of data protection, and the degree of teacher agency maintained. Weighting should reflect institutional missions and student needs: institutions serving historically marginalized groups should assign relatively high weight to equity, whereas accuracy and personalization may be given greater emphasis in resource-rich settings, subject to non-negotiable equity constraints.

6.2. Criminal Justice

Predictive policing, recidivism risk assessment, and bail algorithms illustrate some of the most ethically contested uses of AI [6]. The core tension here is between efficiency and justice. Although AI systems can process more information than human judges or officers, widely used risk tools have been shown to reproduce racial and socio-economic disparities, raising questions about violations of due process and distributive justice [52].

Within the Axiology–MCDA Framework, justice and dignity function as intrinsic values that impose strict moral limits: no gain in efficiency can justify a system that systematically disadvantages particular groups. This reflects the tripartite value structure, in which intrinsic values cannot be weighted to zero. Within this constraint, MCDA can still guide deliberation about permissible trade-offs, such as balancing different dimensions of accuracy or determining appropriate levels of human oversight.

6.3. Finance

AI systems in finance, which include credit scoring, fraud detection, algorithmic trading, and insurance pricing, operate under strong incentives to maximize efficiency and profitability [53]. These pressures create ethical tensions with fairness, transparency, and accountability. Opaque credit scoring models may prevent individuals from understanding or contesting decisions that affect their financial inclusion. Algorithmic trading systems can amplify market instability, and insurance pricing models risk indirect discrimination through proxies for protected characteristics.

In this domain, the Axiology–MCDA Framework highlights a tension between relational values, especially trust and accountability that depend on transparency, and instrumental values such as efficiency and competitive advantage. Transparency is not only an instrumental good; it is also a relational value that supports the legitimacy of financial markets. Without sufficient transparency, trust erodes and the social conditions that enable markets to function are weakened. This axiological analysis therefore supports regulatory requirements for algorithmic explainability and offers a normative justification for such requirements that goes beyond arguments based solely on economic efficiency.

7. Strengths and Limitations

7.1. Strengths of the Framework

The integrated Axiology–MCDA Framework brings several contributions to ethical AI governance. First of all, it is grounded in philosophy. Rather than relying on lists of principles assembled for policy or industry purposes, the framework draws on a substantive tradition that analyzes the nature, structure, and justification of values. This grounding enables more stable and defensible reasoning about difficult cases than principle-based approaches.

Another strength is its sensitivity to context. Ethical priorities differ across domains, culture, institutional settings, and stakeholder groups. This framework accommodates the variation without assuming that a single ranking of values applies universally. By structuring deliberation around specific contexts, it reflects the plural and contested character of real-world ethical decisions.

The framework also enhances transparency. Explicit documentation of criteria, weights, and the reasoning behind them provides an audit trail that supports accountability and enables critical review. Such traceability is especially important in high-stakes environments in which the legitimacy of AI-supported decisions depends on their intelligibility to affected parties.

In addition, the deliberative structure of the framework fits naturally with participatory approaches to technology governance. It offers a way for affected communities to express and justify their value commitments, reinforcing the democratic aims of Digital Humanism and counteracting the tendency to treat ethics as a purely technical matter.

7.2. Limitations

The framework has several limitations. Weight assignment is inherently subjective, since different stakeholders may prioritize values differently, and there is no neutral rule for deciding which weighting is correct. The framework makes these differences visible, but it cannot remove value disagreement.

There is also a risk of false precision. Numerical scores and thresholds may appear more exact than they really are, and the results depend on performance estimates and weights that involve judgment. Users must therefore avoid treating MCDA outputs as substitutes for genuine ethical reasoning [54].

Criteria selection is another challenge. Values that are easy to quantify, such as accuracy or efficiency, may unintentionally dominate values that are harder to measure, such as dignity or trust. Care is needed to ensure that important but less measurable values are not overlooked.

Applying the full MCDA process can be resource-intensive. It may only be feasible for high-stakes decisions, which raises questions about how to identify which decisions require the full procedure and which can rely on simplified approaches. The framework should also account for the entire AI lifecycle, including hardware and software considerations from design through decommissioning [55].

Finally, some values may be genuinely incommensurable. The weighted-sum model assumes that values can be compared on at least an ordinal scale. When this is not the case—for example, in assessments involving violations of human dignity—the model may mislead. In situations involving violations of intrinsic values, aggregation may not be appropriate, and constraint-based exclusion or rights-based reasoning may be required. In such contexts, alternative MCDA methods that are more robust to incommensurability, such as outranking approaches, may be more appropriate.

7.3. Comparison with Alternative Approaches

Placed alongside existing approaches, the Axiology–MCDA Framework occupies an intermediate position. Principle-based approaches, such as the EU Ethics Guidelines for Trustworthy AI, offer accessibility and broad agreement but lack the means to address concrete value conflicts. Computational ethics approaches, including formal logic systems and algorithmic fairness metrics, provide precision and auditability but struggle with the pluralism and contextual variation characteristic of real-world ethical issues [56]. Participatory design approaches strengthen democratic legitimacy but often lack a structured method for articulating and comparing value trade-offs. The Axiology–MCDA Framework brings these strands together: it retains formal structure, incorporates philosophical depth, and supports inclusive deliberation.

8. Conclusions

This paper has developed an Axiology–MCDA Framework for artificial intelligence ethics grounded in philosophy. By distinguishing intrinsic, instrumental, and relational values, and by operationalizing value pluralism through Multi-Criteria analysis, the framework offers a structured method for evaluating AI systems across diverse contexts.

The contributions include a philosophical account of axiology, a tripartite value taxonomy, an adaptation of Multi-Criteria analysis for ethical evaluation, a healthcare case study with sensitivity analysis, cross-domain applications, and a critical appraisal situating the framework within broader debates in AI governance.

The core insight is that ethical tensions in AI stem from genuine value pluralism rather than conceptual error. As pluralist theorists note, conflicts between legitimate values cannot always be resolved by a single overarching principle [19]. Making value commitments explicit through structured analysis supports more deliberative, transparent, and context-sensitive governance. Since value conflicts cannot be removed through technical optimization, the framework offers guidance for navigating them in a principled and accountable way.

Future work can develop the Integrated Axiology–MCDA Framework in several related directions. Empirical studies in institutional settings—such as healthcare procurement, educational technology assessment, and AI-assisted judicial processes—would test its practical utility and perceived legitimacy. Methodological refinement should explore advanced MCDA techniques, including outranking methods such as ELECTRE and PROMETHEE and multi-attribute utility approaches suitable for non-linear and incommensurable value relations. Closer integration with participatory design would help ensure that marginalized and underrepresented groups are included in value deliberation. Governance procedures also require refinement to specify how criteria and weights should be revised as systems evolve. Finally, testing the framework against difficult cases involving incommensurable values or moral dilemmas would help clarify the limits of axiological analysis and indicate where supplementary philosophical resources may be needed.

The framework should therefore be understood not as a decision-making algorithm that produces definitive answers, but as a deliberative support system that enables structured negotiation under conditions of value pluralism, allowing disagreements to be articulated, examined, and revised.

Author Contributions: Conceptualisation: G.D.-C., F.S. and D.I. Methodology: F.S. Formal analysis: F.S. Investigation: F.S. Writing—original draft preparation: F.S. Writing—review and editing: D.I. and G.D.-C. Supervision: D.I. and G.D.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [CrossRef]
2. Floridi, L.; Cowls, J. A unified framework of five principles for AI in society. *Harv. Data Sci. Rev.* **2019**, *1*. [CrossRef]
3. Floridi, L. Translating principles into practices of digital ethics: Five risks of being unethical. *Philos. Technol.* **2019**, *32*, 185–193. [CrossRef]
4. Floridi, L. From principles to practices: The risks of being unethical. In *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*; Oxford University Press: Oxford, UK, 2023. [CrossRef]
5. Williamson, S.M.; Prybutok, V. Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Appl. Sci.* **2024**, *14*, 675. [CrossRef]
6. Lau, T. *Predictive Policing Explained*; Brennan Center for Justice: New York, NY, USA, 2020. Available online: https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained (accessed on 26 March 2026).
7. Kakarala, M.R.K.; Rongali, S.K. Existing challenges in ethical AI: Addressing algorithmic bias, transparency, accountability and regulatory compliance. *World J. Adv. Res. Rev.* **2025**, *25*, 2511–2516. [CrossRef]
8. AlJadaan, O.T.; Zaidi, H.; Al Faress, M.Y.; Jabas, A.O. Ethics in AI and computation in automated decision-making. In *Enhancing Automated Decision-Making Through AI*; Hai-Jew, S., Ed.; IGI Global: Hershey, PA, USA, 2024; pp. 397–424. [CrossRef]
9. Alexander, L.; Michael, M. Deontological Ethics. In *The Stanford Encyclopedia of Philosophy*, Winter 2021 ed.; Zalta, E.N., Ed.; The Metaphysics Research Lab: Stanford, CA, USA, 2021. Available online: https://plato.stanford.edu/archives/win2021/entries/ethics-deontological/ (accessed on 26 May 2026).
10. Driver, J. The History of Utilitarianism. In *The Stanford Encyclopedia of Philosophy*; Zalta, E.N., Ed.; The Metaphysics Research Lab: Stanford, CA, USA, 2014. Available online: https://plato.stanford.edu/entries/utilitarianism-history/ (accessed on 26 May 2026).
11. Hursthouse, R.; Pettigrove, G. Virtue Ethics. In *The Stanford Encyclopedia of Philosophy*, Spring 2022 ed.; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2022. Available online: https://plato.stanford.edu/archives/spr2022/entries/ethics-virtue/ (accessed on 26 May 2026).
12. Hagendorff, T. The ethics of AI ethics: An evaluation of guidelines. *Minds Mach.* **2020**, *30*, 99–120. [CrossRef]
13. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef] [PubMed]
14. Dubljević, V.; List, G.F.; Milojevich, J.; Ajmeri, N.; Bauer, W.; Singh, M.P.; Samandar, M.S. Toward a rational and ethical sociotechnical system of autonomous vehicles: A novel application of multi-criteria decision analysis. *PLoS ONE* **2021**, *16*, e0256224. [CrossRef] [PubMed]
15. Lapie, P. *Logique de la Volonté*; Félix Alcan: Paris, France, 1902.
16. von Hartmann, E. *Grundriss der Axiologie*; B. G. Teubner: Wiesbaden, Germany, 1908.
17. Moore, G.E. *Principia Ethica*; Cambridge University Press: Cambridge, UK, 1903. Available online: https://www.gutenberg.org/files/53430/53430-h/53430-h.htm (accessed on 26 May 2026).
18. Ross, W.D. *The Right and the Good*; Oxford University Press: Oxford, UK, 1930.
19. Berlin, I. *Four Essays on Liberty*; Oxford University Press: Oxford, UK, 1969.
20. Raz, J. *The Morality of Freedom*; Oxford University Press: Oxford, UK, 1986.
21. Chang, R. *Making Comparisons Count*; Routledge: London, UK, 2002.
22. Nussbaum, M.C. *Women and Human Development: The Capabilities Approach*; Cambridge University Press: Cambridge, UK, 2000.
23. Nussbaum, M.C. *Creating Capabilities: The Human Development Approach*; Harvard University Press: Cambridge, MA, USA, 2011.
24. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]
25. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; Srikumar, M. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*; Publication No. 2020-1; Berkman Klein Center Research Publication: Cambridge, MA, USA, 2020.. [CrossRef]
26. Whittlestone, J.; Nyrup, R.; Alexandrova, A.; Cave, S. The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, HI, USA,27–28 January 2019; pp. 195–200. [CrossRef]

27. Dodig-Crnkovic, G.; Sapienza, G. Ethical aspects of technology in the multi-criteria decision analysis. In Proceedings of the IACAP Conference, Ferrara, Italy, 14–17 June 2016.
28. Sapienza, G.; Dodig-Crnkovic, G.; Crnkovic, I. Inclusion of ethical aspects in multi-criteria decision analysis. In Proceedings of the WICSA and CompArch Conference, 1st International Workshop on Decision Making in Software ARCHitecture (MARCH), Venice, Italy, 5–8 April 2016. [CrossRef]
29. Kant, I. *Groundwork of the Metaphysics of Morals*, Revised ed.; Original work published 1785; Gregor, M., Timmermann, J., Eds. and Trans.; Cambridge University Press: Cambridge, UK, 2012. [CrossRef]
30. Chan, K.M.; Balvanera, P.; Benessaiah, K.; Chapman, M.; Díaz, S.; Gómez-Baggethun, E.; Turner, N.J. Why protect nature? Rethinking values and the environment. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1462–1465.. [CrossRef] [PubMed]
31. Klain, S.C.; Olmsted, P.; Chan, K.M.A.; Satterfield, T. Relational values resonate broadly and differently than intrinsic or instrumental values, or the New Ecological Paradigm. *PLoS ONE* **2017**, *12*, e0183962. [CrossRef] [PubMed]
32. Chan, K.M.; Gould, R.K.; Pascual, U. Relational values: What are they, and what's the fuss about? *Curr. Opin. Environ. Sustain.* **2018**, *35*, A1–A7. [CrossRef]
33. Himes, A.; Muraca, B. Relational values: The key to pluralistic valuation of ecosystem services. *Curr. Opin. Environ. Sustain.* **2018**, *35*, 1–7. [CrossRef]
34. Gunkel, D.J.; Coeckelbergh, M. A relational approach to moral standing: Reframing ethical boundaries in the age of artificial intelligence. In *New Directions in Relational Sociology, Volume Two*; Vandenberghe, F., Papilloud, C., Eds.; Palgrave Macmillan: New York, NY, USA, 2025. [CrossRef]
35. Coeckelbergh, M. Three challenges for a global AI ethics: Towards a more relational normative vision. *AI Ethics* **2025**, *5*, 5527–5533. [CrossRef]
36. Mackenzie, C.; Stoljar, N. Autonomy refigured. In *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*; Mackenzie, C., Stoljar, N., Eds.; Oxford University Press: Oxford, UK, 2000; pp. 3–31. [CrossRef]
37. Hardin, R. The street-level epistemology of trust. In *Trust and Trustworthiness*; Russell Sage Foundation: New York, NY, USA, 2002. [CrossRef]
38. Friedman, B.; Kahn, P.H., Jr.; Borning, A. Value sensitive design and information systems. In *Human–Computer Interaction and Management Information Systems: Foundations*; Zhang, P., Galletta, D., Eds.; Routledge: London, UK, 2013; pp. 55–95. Available online: https://nissenbaum.tech.cornell.edu/papers/Value_Sensitive_Design.pdf (accessed on 26 May 2026).
39. UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021. Available online: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (accessed on 26 May 2026).
40. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**. Available online: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed on 26 May 2026).
41. Chouldechova, A. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* **2017**, *5*, 153–163. [CrossRef] [PubMed]
42. Dignum, V. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*; Springer: New York, NY, USA, 2019. [CrossRef]
43. Vienna Manifesto on Digital Humanism. 2019. Available online: https://dighum.ec.tuwien.ac.at/wp-content/uploads/2019/07/Vienna_Manifesto_on_Digital_Humanism_EN.pdf (accessed on 26 May 2026).
44. Morley, J.; Floridi, L.; Kinsey, L.; Elhalal, A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **2020**, *26*, 2141–2168. [CrossRef] [PubMed]
45. Habermas, J. The Theory of Communicative Action. In *Reason and the Rationalization of Society*; McCarthy, T., Trans.; Beacon Press: Boston, MA, USA, 1984; Volume 1.
46. Floridi, L. *The Logic of Information: A Theory of Philosophy as Conceptual Design*; Oxford University Press: Oxford, UK, 2019. [CrossRef]
47. Belton, V.; Stewart, T.J. *Multiple Criteria Decision Analysis: An Integrated Approach*; Kluwer Academic Publishers (Springer): Dordrecht, The Netherlands, 2002. [CrossRef]
48. Ishizaka, A.; Nemery, P. *Multi-Criteria Decision Analysis: Methods and Software*; John Wiley & Sons: New York, NY, USA, 2013. [CrossRef]
49. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis: The Primer*; John Wiley & Sons: New York, NY, USA, 2008. [CrossRef]
50. Holmes, W.; Bialik, M.; Fadel, C. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*; Center for Curriculum Redesign: Boston, MA, USA, 2019. Available online: https://curriculumredesign.org/our-work/artificial-intelligence-in-education/ (accessed on 26 May 2026).
51. Baker, R.S.; Hawn, A. Algorithmic bias in education. *Int. J. Artif. Intell. Educ.* **2022**, *32*, 1052–1092. [CrossRef]

52. Angwin, J.; Larson, J.; Mattu, S.; Kirchner, L. *Machine Bias*; ProPublica: New York, NY, USA, 2016. Available online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed on 26 May 2026).
53. Bartlett, R.; Morse, A.; Stanton, R.; Wallace, N. Consumer-lending discrimination in the FinTech era. *J. Financ. Econ.* **2022**, *143*, 30–56. [CrossRef]
54. Green, B. The flaws of policies requiring human oversight of government algorithms. *Comput. Law. Secur. Rev.* **2022**, *45*, 105681. [CrossRef]
55. Holstein, T.; Dodig-Crnkovic, G.; Pelliccione, P. Steps towards real-world ethics for self-driving cars: Beyond the trolley problem. In *Machine Law, Ethics, and Morality in the Age of Artificial Intelligence*; Thompson, S.J., Ed.; IGI Global: Hershey, PA, USA, 2021. [CrossRef]
56. Wallach, W.; Allen, C. *Moral Machines: Teaching Robots Right from Wrong*; Oxford University Press: Oxford, UK, 2009. [CrossRef]