Human–AI Decision-Making Ethical Governance Framework: Integrating Algorithmic Fairness and Managerial Intuition

Mustafa Kaya[*1]

[1,]School of Administrative and Social Sciences, Ankara Medipol University, Ankara, Türkiye ORCID: 0000-0002-3054-3108,
* Corresponding author: Mustafa Kaya (mustafa.kaya@ankaramedipol.edu.tr)
Received: 10 November 2025, Revised: 04 December 2025, Accepted: 14 December 2025, Published: 17 December 2025

Abstract: The increasing use of artificial intelligence (AI)-supported decision-making systems in critical areas such as healthcare, finance, and justice raises important concerns regarding ethics, bias, and fairness. This conceptual article proposes a holistic framework that integrates the principles of algorithmic fairness with the unique value of managerial intuition to ensure ethical governance in human–AI decision-making contexts. The interdisciplinary literature synthesis reveals that existing ethical principles often remain abstract and that a practical roadmap is needed. The proposed framework consists of five core components: (i) human-centered design and continuous oversight, (ii) transparency and explainability, (iii) bias management and mitigation, (iv) integration of managerial intuition, and (v) ethical governance and continuous adaptation. Together, these components safeguard ethical responsibility in human–AI collaboration, address algorithmic biases, and ensure that intuition serves as a "moral compass" under uncertainty. The article highlights the efficiency and consistency advantages of algorithmic objectivity while emphasizing that the contextual understanding, empathy, and ethical judgment capacity of human intuition creates a complementary force. This synergy is the key to decision-making processes that are not only efficient but also just and humane. The study provides a practical roadmap for policymakers, developers, and managers and proposes a research agenda for future empirical validation, cultural adaptations, and ethical governance models.

Keywords: Artificial Intelligence, Algorithmic Fairness, Human–AI Collaboration, Ethical Governance

1. Introduction

The increasing application of artificial intelligence in critical decision-making across sectors such as healthcare, banking, and justice has ushered in a new era of efficiency and analytical capability (Osasona et al., 2024). However, this extensive utilization also prompts significant ethical inquiries, particularly about equity and prejudice (Osasona et al., 2024). This conceptual article introduces a comprehensive paradigm that integrates the principles of algorithmic fairness with the critical role of management intuition to ensure ethical governance in human–AI decision-making contexts. This method acknowledges that while AI excels in particular areas, human oversight and judgment are crucial for its responsible implementation (Sigfrids et al., 2023).

The historical impact of technology on decision-making has been a subject of ongoing discourse (Cremer & Narayanan, 2023). Significant milestones such as IBM's Watson and Google DeepMind's AlphaGo demonstrate the formidable capabilities of artificial intelligence; they can analyze vast datasets, uncover concealed relationships, and generate concepts that significantly enhance human cognitive abilities (Akinnagbe, 2024). This analytical ability aids individuals in making improved selections (Vincent, 2021). However, despite these advances, AI possesses inherent limitations, particularly in situations characterized by significant uncertainty or ambiguity (Vincent, 2021). In these contexts, human attributes like creativity, intuition, and social intelligence are paramount (Vincent, 2021).

Consequently, the "human + machine" or "hybrid intelligence" paradigm of human-AI collaboration has gained prominence, indicating that the amalgamation of both strengths yields superior outcomes compared to solitary efforts (Hao et al., 2024). Research in medical diagnostics has shown that error rates in cancer identification significantly decrease when AI and human pathologists work together (Reverberi et al., 2022). Therefore, AI should be viewed not as an opponent but as a significant augmentation of human intelligence (Akinnagbe, 2024). The amalgamation of human intuition, value assessments, and creativity with AI's analytical prowess is a crucial strategy for navigating complicated, unpredictable, and ambiguous situations (Geng & Varshney, 2022).

This study presents a new ethical framework that integrates algorithmic fairness with managerial intuition, aiming to rectify a significant gap in the existing literature by providing a practical guide for ethical AI governance (Rodgers et al., 2022). This concept enriches the Management Information Systems literature by amalgamating AI-driven decision support systems with ethical and management principles, thus connecting technical algorithmic design to organizational practice.

This is the central research question this article seeks to answer: How can a holistic ethical governance framework be constructed that integrates algorithmic fairness and managerial intuition? By answering this question, this study aims to fill the gap between abstract ethical principle sets and organizational practice in the current literature. The proposed framework is essential because it provides a practical roadmap that explicitly addresses the critical human element (intuition) within the often purely technical discourse of algorithmic fairness.

2. Literature Review

The discourse regarding AI's capabilities and its potential to displace human employment is not novel; it is a continuation of a longstanding narrative of technological evolution (Deming et al., 2025). John Maynard Keynes coined the term "technological unemployment" in 1930. He referred to it as a "novel disease" that civilizations would need to confront (Minssen et al., 2025). Subsequently, scholars like Shoshana Zuboff examined the organizational ramifications of emerging information technologies, characterizing them as "smart technologies" that convert actions into informational representations (Dwivedi et al., 2019). Zuboff highlighted benefits such as enhanced transparency, while cautioning against their reductionist use solely for automation and control (Dwivedi et al., 2019). This historical context underscores that automation has been a focal point of academic and managerial discussions for decades (García-Murillo & MacInnes, 2019).

Contemporary discourse regarding AI, receiving significant media coverage, frequently presents two contrasting perspectives (Dwivedi et al., 2019). Some individuals believe that AI will soon surpass human intelligence and dominate the workforce. Some individuals consider these assertions excessively robust and recall previous, frequently erroneous, forecasts regarding the future of AI (Dwivedi et al., 2019). During the 1960s and 1970s, individuals such as Herbert Simon and Marvin Minsky made audacious forecasts that artificial intelligence will attain human-level intellect in a few period. The projections proved to be inaccurate (Dwivedi et al., 2019).

Both historical and contemporary narratives frequently overlook the potential for AI and humans to operate synergistically instead of competitively (Geng & Varshney, 2022). Early pioneers recognized that "the combination of computers and humans outperforms either independently" (Hemmer et al., 2025). This principle is exemplified in various domains, including chess. in 1997, IBM's Deep Blue defeated Garry Kasparov, demonstrating that artificial intelligence excels at mathematical computations. Subsequently, Kasparov endorsed "centaur chess," in which human-machine collaborations aim to enhance performance (Hao et al., 2024).

The collaboration between human and machine intelligence extends beyond game scenarios. In medical diagnostics, AI systems possess a distinct mistake rate, whilst human experts exhibit a different one; yet, their coordinated efforts significantly reduce diagnostic errors (Reverberi et al., 2022). This aligns with J. C. R. Licklider's notion of human–machine symbiosis, in which the benefits of one mitigate the limitations of the other (Geng & Varshney, 2022).

The emergence of AI has rendered the concept of human-machine complementarity increasingly significant, particularly in organizational decision-making (Oppioli et al., 2023). To effectively resolve this issue, it is crucial to distinguish between analytical and intuitive decision-making, acknowledging the challenges presented by uncertainty, complexity, and ambiguity that affect organizational results (Trunk et al., 2020). Management strategies sometimes include both deliberate, analytical approaches and intuitive decisions (Leyer & Schneider, 2021). Artificial intelligence thrives in analytical domains, since expert systems and predictive analytics can process vast datasets and generate insights that surpass human logical reasoning abilities (Oppioli et al., 2023).

Nonetheless, human decision-making frequently relies on intuition, which involves generating

judgments without explicit rational analysis, instead depending on accumulated experiences, tacit knowledge, and creative insight (Vincent, 2021). This mode of cognition illustrates our perception and conceptualization of entities in their entirety (Müller, 2021). Artificial intelligence excels in analytical tasks; yet, it often struggles with common-sense reasoning and ambiguous situations (Vincent, 2021). Conversely, humans exhibit superior capabilities in managing uncertainty and unpredictability compared to AI (Chen et al., 2022). This is due to the fact that individuals possess common sense and discernment, which AI lacks. This illustrates the synergy between their roles: AI excels in handling complexity, while humans have an advantage in traversing uncertainty and ambiguity through intuition, imagination, and creativity (Chen et al., 2023).

3. Methodology

This conceptual article aims to explore the complex relationship between algorithmic fairness, managerial intuition, and ethical frameworks in human–AI collaborative decision-making. The primary objective is to synthesize and integrate existing theoretical knowledge to propose a new ethical framework that bridges these distinct yet interconnected domains. To achieve this objective, the following methodological approaches have been adopted.

3.1. Conceptual Analysis and Comprehensive Literature Synthesis

The primary methodology of this study is based on a comprehensive review and conceptual analysis of the relevant academic literature. For conceptual articles, literature review and synthesis are the most appropriate methods to explain the current state of knowledge and provide a foundation for future empirical testing (Trunk et al., 2020). Studies in multiple disciplines (information ethics, management sciences, philosophy, social sciences, law, artificial intelligence engineering) covering AI ethics, algorithmic fairness, bias in AI, human–AI collaboration, hybrid intelligence, and managerial intuition were examined systematically (Bani-Hani et al., 2024; Lai et al., 2023; Tahaei et al., 2023). This interdisciplinary approach has made it possible to understand how these concepts are defined, applied, and discussed in different contexts. Thus, the social, ethical, and legal dimensions of AI have been addressed from a holistic perspective (Prem, 2023).
The literature review included the following steps:

- **Keyword Identification:** A broad set of keywords such as “algorithmic fairness,” “AI ethics,” “managerial intuition,” “human–AI decision-making,” “hybrid intelligence,” “bias mitigation,” “ethical framework,” “AI governance,” “transparency,” “explainability,” “human-in-the-loop AI,” and “decision support systems,” as well as their derivatives, were used.
- **Database Searches:** Leading academic databases such as IEEE Xplore, ACM Digital Library, Web of Science, Scopus, and Google Scholar were searched. In addition, specific conferences on AI ethics (e.g., FAccT, AIES) and journal publications were also targeted.
- **Article Selection and Filtering:** Thousands of articles obtained from search results were initially filtered based on title, abstract, and keyword relevance. The articles selected for full-text review were approximately 180, and the core studies included in the final synthesis as references are exactly 90. Then, full-text review was conducted to evaluate whether the article addressed the study’s objective and scope, that is, whether it dealt with the balance of ethics, fairness, and intuition in AI decision-making. Publications after 2018 were particularly preferred, but older, seminal studies were also included for fundamental conceptual definitions. The selection criteria prioritized peer-reviewed journal articles and proceedings from major AI ethics conferences (e.g., FAccT, AIES). However, given the rapid development of the field, highly influential pre-prints and working papers from reputable repositories (such as arXiv and SSRN) were also systematically included to capture the most recent conceptual advancements. Only English-language publications were selected.

- **Data Extraction and Synthesis:** From the selected articles, critical information was extracted regarding principles of algorithmic fairness and ethical AI, theoretical foundations and practical applications of managerial intuition, human–AI collaboration models, strengths and weaknesses of existing ethical frameworks, ethical risk factors, and mitigation strategies. These data were categorized under relevant themes and subthemes and synthesized. The emergence of these themes (the 5 Components of the Framework) was achieved through an inductive, iterative classification process derived directly from the literature synthesis; that is, the most frequently recurring and intersecting ethical risks and solution strategies across different disciplines were identified as core components. This process enabled the comparison and integration of perspectives from different disciplines.

In the conceptual analysis, emphasis was placed on clarifying the terminological meanings of terms such as "fairness," "bias," and "ethics" across different disciplines and their interpretations in algorithmic applications (Mulligan et al., 2019; Spitale, 2020). This is particularly important in discussions of algorithmic fairness, as differences between engineering and social science perspectives directly affect definitions of fairness and the fairness metrics of AI systems (Kuppler et al., 2022; Nakao et al., 2022; Valença & Santos, 2025).

3.2. Framework Development Approach

The proposed ethical framework is based on the findings obtained from literature synthesis and conceptual analysis. The framework development process was inspired by the approaches of Ethics by Design and Value Sensitive Design (Bleher & Braun, 2023; d'Aquin et al., 2018). These approaches emphasize the integration of ethical principles into the architecture and operation of AI systems from the earliest stages of development. It has been accepted that these principles must be translated into concrete tools and processes to ensure that AI ethics principles do not remain abstract but transform into practical applications (Prem, 2023).

- **Interdisciplinary and Multidisciplinary Approach:** AI ethics issues are typically multidimensional and complex problems that cannot be addressed by purely technical solutions. Therefore, perspectives from different disciplines such as philosophy, law, sociology, management science, computer science, and psychology were integrated (Stahl et al., 2020). This interdisciplinary and multidisciplinary approach made it possible to more holistically understand the complex social, ethical, and human rights impacts of AI, bring together different perspectives, and generate more inclusive solutions.
- **Systemic Perspective:** AI systems were not considered merely technological tools but as integral parts of socio-technical systems (Xu & Gao, 2024). In this context, a broad systemic perspective was adopted in the design, development, deployment, and use of AI, including human–machine interactions, organizational processes, social norms, and cultural contexts. This requires considering not only the technical performance of AI systems but also their social compatibility and ethical acceptance.
- **Participatory and Stakeholder Orientation:** The active participation of different stakeholders is crucial for implementing fairness in practice and developing ethical AI systems (Curto & Comim, 2023). The proposed framework aims to consider the values, expectations, and concerns of all stakeholders affected by AI (AI developers, end users, regulators, policymakers, civil society organizations, and affected communities). This participatory approach ensures that AI systems are not only technically functional but also socially acceptable, trustworthy, and ethically sound. In addition, it emphasizes the potential of the "conscientious design" approach for hybrid communities formed through AI systems (Noriega et al., 2021; Zhang et al., 2023).

3.3. Structuring of the Conceptual Framework

The conceptual framework was designed to mitigate ethical issues arising from the integration of AI into

decision-making processes and to enhance the synergistic potential of human–AI collaboration. This framework aims to provide a holistic model that combines fundamental ethical principles (transparency, accountability, fairness, human-centeredness) with the critical role of managerial intuition. Considering the criticism that existing AI ethical frameworks in the literature may remain abstract (Liao, 2023; Prem, 2023), the framework proposed in this study will seek to provide more concrete applications and guidance and will focus on practical steps to ensure the alignment of AI systems with ethical responsibility and social justice.

This conceptual article's primary limitation is the absence of empirical testing. The framework remains theoretical and requires validation through case studies and experimental designs in organizational contexts. Methodologically, the limitation of the literature review being restricted to certain databases and predominantly English-language sources creates potential constraints in fully addressing global governance and cultural adaptations.

4. Proposed Ethical Governance Framework

In this section, a holistic framework is presented that integrates the principles of algorithmic fairness with managerial intuition and aims to ensure ethical management in human–AI decision-making processes. The framework seeks to benefit from the analytical power of AI while maintaining human control, ethical responsibility, and social justice concerns. To address the ethical, bias, and fairness issues arising from the integration of AI into decision-making systems, an approach centered on the human factor has been adopted throughout the design, development, implementation, and continuous monitoring of AI systems (Tahaei et al., 2023; Xu et al., 2023). This ethical framework, hereafter referred to as the Ethical Governance Framework, combines both structural ethical guidelines and governance mechanisms.

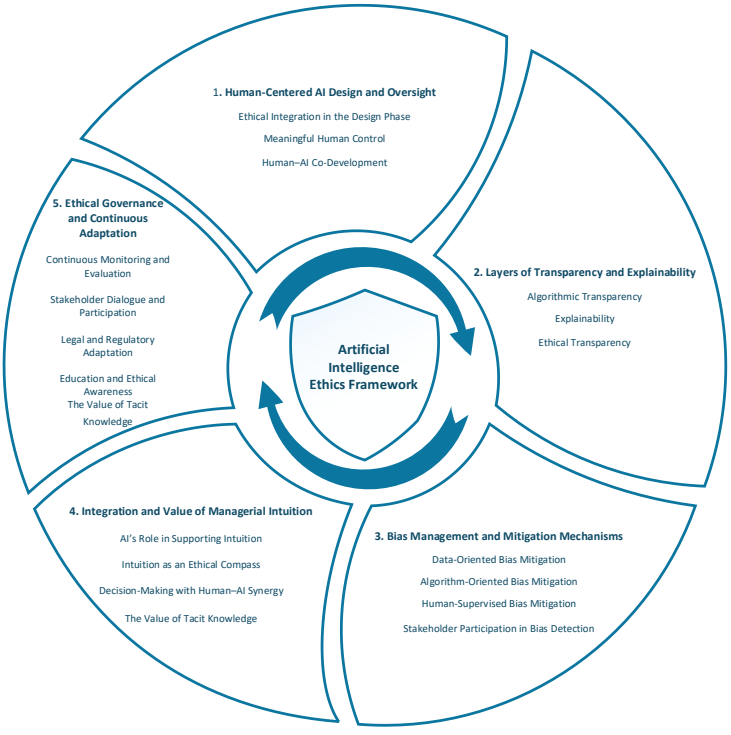

Figure 1. Ethical Governance Framework integrating algorithmic fairness, managerial intuition, and human–AI collaboration

The five components of the proposed framework are derived directly from the conceptual synthesis in Section 3 and reflect the most recurring themes across disciplines. This ensures that the framework is not an abstract construction but a translation of interdisciplinary insights into practical guidance for managers, developers, and policymakers.

4.1. Human-Centered AI Design and Oversight

Ethical AI systems must be designed from the outset with a human-centered approach. This means ensuring that AI systems are not only technically efficient but also respect human values, rights, and autonomy (Stahl et al., 2020; Xu & Gao, 2024). At the core of the framework lies the view of AI as a tool that complements human capabilities, never as a replacement (Auernhammer, 2020; Garibay et al., 2023; Schmager et al., 2025).

- **Ethical Integration in the Design Stage:** Incorporating ethical considerations and fairness principles into the system architecture at the earliest stages of AI development is critical. This approach reduces the potential of AI to produce unethical or biased outcomes at the core of the system and aims to apply the principle of "ethics by design" (d'Aquin et al., 2018). For example, at every stage, from data collection methods to model evaluation metrics, ethical and fairness-oriented decisions must be made, and values such as human dignity, privacy, and autonomy must be adopted as fundamental design principles (Yue & Shyu, 2023).
- **Meaningful Human Control:** In AI-supported decision systems, ultimate responsibility and critical decisions must remain with humans. This requires not only the possibility of overriding AI outputs but also active involvement in the entire decision loop (Jonker et al., 2024). "Human-in-the-loop" approaches mandate integrated human participation to ensure the ethical management of AI (Chen, 2023).
- **Human–AI Co-Development:** AI systems must be continuously improved through the collaboration of human–AI teams, where AI learns from human feedback and ethical judgments (Wang et al., 2020; Bao et al., 2023; Trunk et al., 2020).
-

4.2. Layers of Transparency and Explainability

Transparency and explainability of AI systems are fundamental for user trust and for detecting possible biases and errors. These concepts are multidimensional and can carry different meanings in different contexts.

- **Algorithmic Transparency:** Access to information about how AI systems operate, including architecture, datasets, algorithmic processes, and the logic behind outputs (Prem, 2023).
- **Explainability:** The ability to explain in understandable language how AI reached a decision, which factors were most influential, and why these factors were weighted (Binns et al., 2018; Binns, 2020).
- **Ethical Transparency:** Transparency should also serve as an ethical tool by proactively disclosing biases, potential harms, or uncertainties. This empowers stakeholders to provide informed feedback about AI's behavior and builds mutual trust.

4.3. Bias Management and Mitigation Mechanisms

Algorithmic bias is one of the most significant threats to the fairness of AI systems (Demartini et al., 2023; Fu et al., 2020). Bias management requires continuous and multidimensional strategies.

- **Diverse and Representative Datasets:** Ensuring equitable representation across demographic groups (Risser et al., 2022).
- **Data Auditing and Preprocessing:** Regular auditing and corrective techniques such as re-sampling and re-weighting (Demartini et al., 2023).
- **Fairness Metrics:** Applying measures such as equal opportunity, demographic parity, or individual fairness (Barsotti & Koçer, 2022; Waller et al., 2023).
- **Fairness-Enhancing Algorithms:** Integrating algorithms specifically designed to reduce bias (Orphanou et al., 2021).
- **Human-Supervised Mitigation:** Ongoing human monitoring to detect and address subtle biases (Orphanou et al., 2021).

- **Stakeholder Participation:** Involving communities in bias detection, consistent with the "conscientious design" approach (Noriega et al., 2021; Zhang et al., 2023).

4.4. Integration and Value of Managerial Intuition

The framework acknowledges managerial intuition as an indispensable complement to AI in decision-making

.

- **AI's Role in Supporting Intuition:** Providing data-driven insights that can strengthen or challenge intuitive judgments (Müller, 2021; Saghafian & Idan, 2024).
- **The Ethical Compass Role of Intuition:** Enabling consideration of ethical dimensions, social impacts, and values in ambiguous situations (Janhonen, 2023; Karakuş et al., 2025).
- **Human–AI Synergy:** Combining AI's analytical rigor with human contextual and moral reasoning for optimal decisions (Dellermann et al., 2019; Seufert & Meier, 2023).
- **Tacit Knowledge:** Leveraging human sensitivity to emotions, beliefs, cultural norms, and contextual subtleties that AI cannot digitize (Guan et al., 2022; Janhonen, 2023).

4.5. Ethical Governance and Continuous Adaptation

The Ethical Governance Framework must be understood as dynamic and continuously adaptive, rather than a static structure.

- **Continuous Monitoring and Evaluation:** Regular assessments of fairness, performance, and ethical outcomes (Zhang et al., 2023).
- **Stakeholder Dialogue and Participation:** Establishing permanent mechanisms for inclusive governance (Decker et al., 2024; Trunk et al., 2020). Creating AI ethics boards within organizations could operate this dialogue.
- **Legal and Regulatory Adaptation:** Bridging gaps between technological advances and regulation, contributing to harmonized global AI governance (Corrêa et al., 2023; Minssen et al., 2025; Wu & Liu, 2023b).
- **Education and Ethical Awareness:** Training programs to improve ethical literacy, strengthen accountability, and embed AI ethics into organizational culture.

5. Discussion

The proposed Ethical Governance Framework brings together the principles of algorithmic fairness and the critical role of managerial intuition with a holistic approach to ensure that human–AI decision-making processes are managed ethically and responsibly. However, this integration also brings various tensions, potential synergies, and implementation challenges. In this section, the multifaceted interactions of AI in the complex decision-making environment will be analyzed in depth.

5.1. Tensions and Conflicting Values Between Algorithmic Fairness and Managerial Intuition

In the interaction between AI's analytical power and the holistic understanding of human intuition, certain internal tensions exist. AI systems generally prioritize values such as efficiency, objectivity, consistency, and scalability, while human intuition is typically characterized by contextual understanding, empathy, flexibility, moral judgment, and emotional intelligence (Helberger et al., 2020; Janhonen, 2023). This situation becomes particularly evident in ethically sensitive or high-stakes decisions.

- **Conflict Between Objectivity and Individuality:** Algorithms tend to produce objective decisions based on generalized patterns and statistical relationships. However, in situations requiring individual justice, the ability to focus on the unique circumstances of each case and evaluate contextual nuances is exclusive to human judgment (Binns, 2020). In areas such as criminal justice, healthcare, or recruitment, algorithmic risk assessment tools may be contrary to

fair legal or social processes due to their inability to recognize shades of gray and incorporate individual bias (Peng & Simard-Halm, 2020). For example, in recruitment processes, algorithmic scoring systems may overlook context-specific managerial judgments, such as recognizing a candidate's unique career trajectory or organizational fit, which human intuition can better capture. People may show reluctance toward AI (algorithm aversion) especially in decision-making situations with serious consequences, as they worry that algorithms may be opaque in moral decision-making, potentially biased, or replacing human discretion (Filiz et al., 2023; Jauernig et al., 2022; Szafran & Bach, 2024).

- **Balance of Emotion and Logic:** Human intuition develops in a rich context that often includes emotional, social, and cultural factors and influences decision-making processes. The emotionless nature of AI can sometimes be perceived as impartiality, yet it may fall short in decisions requiring empathy, compassion, and the ability to make exceptions (Helberger et al., 2020). For AI to learn what is important and acceptable to us, it must incorporate all relevant social and psychological forces into its comprehensive calculations (Janhonen, 2023). In fields such as education, the difference between AI's objective, data-driven decisions and teachers' pedagogical intuition and ethical values may require different approaches to resolve ethical dilemmas. Moreover, whether human ethical values can be successfully modeled by AI remains a subject of debate (Karakuş et al., 2025).
- **Differences in Definitions of Justice:** Algorithmic fairness can be defined by different mathematical metrics (for example, equal opportunity, demographic parity, predictive equality). However, there are often trade-offs between these metrics, making it difficult to find a single fair solution (Nakao et al., 2022; Waller et al., 2023). Social justice understanding, on the other hand, is based on human values, social norms, and moral judgments. These two understandings of justice may not always overlap and may even conflict. This creates complexity in the design and evaluation of AI systems. Tensions between ethical principles lead to ethical dilemmas, creating difficulties in deciding which principle should be prioritized, and such situations require careful balancing judgment (Jedličková, 2024).

5.2. Synergies and Complementary Roles

Despite these tensions, algorithmic fairness and managerial intuition can play complementary roles and, when combined, create stronger, more ethical, and more informed decision-making processes. The hybrid intelligence approach lies at the center of these synergies.

- **AI-Supported Enhanced Intuition:** AI can help managers enrich their intuition by providing access to previously unavailable massive amounts of data, advanced analysis, and insights. AI can detect hidden patterns and correlations in large datasets, thereby providing evidence to confirm, question, or present new perspectives for managers' intuitive insights (Müller, 2021; Saghafian & Idan, 2024). This helps managers make more informed and contextually appropriate intuitive decisions while also allowing AI to search for answers to questions about when human intuition may be useful. By focusing on complex analytical tasks with its high analytical capabilities, while humans focus on tasks involving uncertainty and ambiguity, human–AI collaboration can be enhanced (Vincent, 2021).
- **Bias Mitigation Through Human Intuition:** While AI on its own may contain data-related or algorithm design-related biases, human oversight and intuition play a critical role in detecting and mitigating these biases. Humans can identify subtle social, cultural, or contextual biases overlooked by algorithms and intervene to prevent unethical outcomes. It has also been suggested that the similarity of human and machine errors in AI systems may have benefits for machine-assisted decision-making. This demonstrates the complex nature of human–AI collaboration (Grgić-Hlača et al., 2022; Vakali & Tantalaki, 2024).
- **Ethical Leadership and Strengthened Frameworks:** Ethical leadership ensures that ethical principles are embedded in institutional culture during AI integration, strengthening both

algorithmic fairness and the ethical dimension of managerial intuition (Kandasamy, 2024; Uddin, 2023). Institutional ethical frameworks guide managers in complex ethical dilemmas, providing consistency and responsibility in decision-making processes (Gkeredakis et al., 2023). It has been suggested that AI can serve as a mirror to help managers understand the psychological foundations of ethical behavior, thereby increasing ethical maturity and improving the ethical capabilities of organizations (Cremer & Narayanan, 2023).

5.3. Challenges and Barriers to Implementing the Ethical Framework

Despite the theoretical robustness of the proposed Ethical Governance Framework, its implementation in the real world may encounter significant challenges. These challenges range from technological limitations to institutional resistance and regulatory gaps.

- **Translation of Ethical Principles into Practice:** Although many AI ethics principles and guidelines exist, difficulties are encountered in transforming these abstract principles into concrete, actionable tools (Agbese et al., 2023; Prem, 2023). Organizations may face criticism of "ethics washing" in AI ethics, as the adoption of ethical principles only in appearance may not reflect real practice. For example, some financial institutions have adopted AI ethics charters but continued to deploy credit-scoring algorithms that disproportionately disadvantage minority groups, showing the gap between stated ethical commitments and actual implementation. This stems from the lack of practical tools and applications in the field of AI ethics.
- **Institutional Resistance and Management Support:** The adoption of new ethical frameworks requires profound changes in organizational culture and processes. Efforts to integrate AI ethics into business processes may conflict with business goals and the pursuit of competitive advantage, slow down the development of AI-based projects, and create difficulties in balancing different stakeholder interests (Baker-Brunnbauer, 2020; Brendel et al., 2021; Hossain et al., 2025; Maiti et al., 2025). The lack of awareness and support of management regarding AI ethics or limited guidance in establishing ethical management may hinder the widespread adoption of ethical practices.
- **Attribution of Responsibility and Accountability:** As AI systems become more complex, the question of who is responsible for faulty or unethical decisions becomes even more complicated (Cañas, 2022). Uncertainty in the distribution of responsibility in human–AI collaboration may create ethical gaps. Since AI systems cannot fully account for ethical principles or moral theories, developing ethical guidance functions may be difficult (Pflanzer et al., 2022). Instead of reducing problems in decision-making, AI may increase them, thereby raising human responsibility (Trunk et al., 2020).
- **Technological Limitations:** AI's ability to make moral judgments or render ethical principles computable is still limited. The inability of AI to fully grasp the ethical principles and moral theories informing its decisions makes it difficult to develop ethical guidance functions in human–AI teams (Pflanzer et al., 2022). This makes the direct integration of ethical principles into AI systems technically difficult and requires continuous research.
- **Balance Between Trust and Transparency:** People's trust in AI is closely linked to the system's transparency and explainability. However, how cultural biases in AI systems affect human decision-making processes and undermine trust is a complex issue. It has been shown that biased AI can improve human decision-making but may simultaneously reduce trust (Lai et al., 2025). Excessive transparency can sometimes lead to meaningless details and reduce trust, while insufficient transparency can also lead to distrust.
- **Regulatory Gaps and Adaptation:** The rapid development of AI technologies has caused existing legal and regulatory frameworks to lag. These gaps make it difficult to standardize ethical practices and support them with legal sanctions (Buyl & Bie, 2024; Minssen et al., 2025; Wu & Liu, 2023b). Due to cultural and managerial differences, there is chaos among global AI regulation proposals, making international coordination difficult. Different regulatory

approaches of various countries (for example, USA, EU, China) increase the complexity of creating a harmonized framework at the global level (Chun et al., 2025; Levi-Faur et al., 2021). AI governance requires flexible approaches in a rapidly changing regulatory environment (Zaidan & Ibrahim, 2024).

5.4. Evolution of Managerial Roles and Competencies

In the AI-supported decision-making environment, the roles and required competencies of managers are significantly transforming. Managers are no longer just decision-makers but must also act as ethical guides, supervisors of AI systems, and facilitators of human–AI teams.

- **Ethical Leadership:** Managers must understand the ethical implications of AI, establish an ethical culture in their organizations, and promote the responsible use of AI (Cremer, 2024; Kandasamy, 2024; Uddin, 2023). This includes leadership qualities and ethical awareness that maximize the potential of AI while minimizing risks.
- **Human–AI Relationship Competencies:** Middle managers managing AI-integrated service teams must possess mechanical, instrumental, collegial, and collaborative relationship competencies (Koponen et al., 2023; Lou et al., 2025). These competencies involve understanding and managing how AI can collaborate in complex social interactions. Developing human–AI collaboration requires new interaction protocols, delegation strategies, and responsibility-sharing frameworks.
- **Critical Thinking and Contextual Understanding:** Managers must critically evaluate AI's recommendations rather than blindly accepting them, considering the contextual relevance and ethical consequences of the information provided by AI. While AI may enhance decision-making abilities, human biases may lead to suboptimal choices, requiring managers to confront these biases (Leyer & Schneider, 2021). Managers' ability to understand AI and recognize potential moral dilemmas is a critical combination for responsible AI adoption (Cremer, 2024).

6. Conclusion and Future Research

6.1. Conclusion

This conceptual article presents a holistic approach to the ethical challenges that arise at the intersection of algorithmic fairness, managerial intuition, and human–AI decision-making processes. The increasingly significant role of artificial intelligence in decision-making processes requires a new way of thinking and a practical framework regarding ethics, bias, and justice. The proposed Ethical Governance Framework consists of five key components: human-centered artificial intelligence design and supervision, layers of transparency and explainability, mechanisms for managing and reducing bias, the integration of managerial intuition into ethical decision-making processes, and ethical governance and continuous adaptation. This Ethical Governance Framework is designed to benefit from the analytical advantages of artificial intelligence while preserving human control, ethical responsibility, and social justice concerns.

The article emphasizes the limitation of algorithmic systems in fully ensuring individual justice and stresses that this gap must be filled by human judgment and intuition (Binns, 2020). At the same time, it has demonstrated that the pursuit of efficiency and objectivity in artificial intelligence must be balanced with the contextual understanding and moral judgment capability of human intuition (Janhonen, 2023; Saghafian & Idan, 2024). This synergy enables artificial intelligence to act as an effective support tool in decision-making processes, while ensuring that ultimate ethical responsibility remains with humans (Cañas, 2022). Ethical leadership and institutional ethical frameworks are of critical importance for the successful realization of this integration (Gkeredakis et al., 2023; Kandasamy, 2024). Although artificial intelligence enhances decision-making capabilities, it can also mirror the biases and ethical immaturity of human decision-makers, thereby serving as a mirror for the development of ethical competencies (Cremer & Narayanan, 2023).

For practitioners, this framework offers a roadmap to integrate fairness metrics into AI-based decision

support tools, train managers to critically interpret AI outputs, and institutionalize ethical governance within MIS infrastructures. The core contribution of this study lies in the explicit integration of managerial intuition as a "moral compass" and mitigation strategy against algorithmic rigidity, bridging the traditional management discourse with modern algorithmic fairness concerns.

6.2. Future Research Directions

Although the proposed Ethical Governance Framework provides an important theoretical foundation for the responsible and ethical use of artificial intelligence, it also offers a rich agenda for future research. The adoption of human-centered artificial intelligence has identified six major challenge areas: making artificial intelligence understandable, making artificial intelligence transparent, making artificial intelligence safe, making artificial intelligence accountable, making artificial intelligence fair, and using artificial intelligence in a way that empowers humans (Garibay et al., 2023). These challenges shape the research directions summarized below.

- **Empirical Validation of the Framework:** The proposed framework needs to be empirically tested in different sectors (health, finance, public services, education) and across different artificial intelligence applications. Case studies, experimental designs, and action research can evaluate the practical effectiveness of the framework and its potential to improve ethical outcomes in artificial intelligence decision-making processes (Söllner et al., 2025). In particular, comparative studies between teachers and artificial intelligence systems in resolving ethical dilemmas can provide a more holistic perspective (Karakuş et al., 2025).
- **Cultural and Contextual Adaptation:** Perceptions of artificial intelligence ethics and justice may vary across cultures and regions (Maiti et al., 2025). Future research should examine how the proposed framework can be adapted to different cultural contexts and how global ethical standards can be developed. The global diversity of artificial intelligence regulation policies should be analyzed in depth in this context (Wu & Liu, 2023a).
- **Psychological Factors in Human–AI Collaboration:** Trust, dependence, attribution of responsibility, automation bias, and emotional intelligence in human–artificial intelligence collaboration need to be examined in greater depth (Lou et al., 2025; Söllner et al., 2025). The fact that artificial intelligence makes mistakes similar to humans is also an important research area in terms of its impact on human–machine-supported decision-making (Grgić-Hlača et al., 2022). In particular, studies on how the partiality of artificial intelligence affects human decision-making and trust can reveal these complex dynamics (Lai et al., 2025).
- **Evolution of Legal and Regulatory Frameworks:** The rapid development of artificial intelligence technologies exerts constant pressure on existing legal and regulatory frameworks. Future research should investigate how the proposed Ethical Governance Framework can contribute to the development of artificial intelligence laws and regulations and how a harmonized approach can be achieved at the international level (Buyl & Bie, 2024; Minssen et al., 2025). The inability of regulatory frameworks to keep up with technological developments carries the risk of legislative obsolescence (Kiškis, 2023; Wu & Liu, 2023b). Comparative analyses of global artificial intelligence governance approaches in this area can provide policy recommendations (Chun et al., 2025; Levi-Faur et al., 2021).
- **Artificial Intelligence-Supported Development and Training of Managerial Intuition:** Further research should be conducted on how artificial intelligence can enhance managers' intuitive capabilities and support their ethical judgments. This also includes updating management training programs to incorporate artificial intelligence ethics and human–AI collaboration competencies (Koponen et al., 2023). Developing models that account for artificial intelligence's social interaction, moral judgments, and values can ensure that humans and artificial intelligence merge more holistically in ethical decision-making processes (Awad et al., 2024; Janhonen, 2023; Seeamber & Badea, 2023; Yamamoto & Suzuki, 2025).

- **Ethical Governance Models and Organizational Culture:** Further research is needed on how ethical artificial intelligence can be effectively managed at the organizational level and how ethical principles can be embedded into organizational culture. This includes stakeholder participation models, accountability mechanisms, and continuous monitoring processes. For the effective management of artificial intelligence ethics, it is important to examine leadership roles and ethical management models in detail (Brendel et al., 2021; Kandasamy, 2024). The way leaders who promote the responsible use of artificial intelligence influence organizations in terms of ethical practices should also be examined (Cremer, 2024; Hossain et al., 2025). In addition, the practical limitations and solutions of approaches that determine decision-making authority for performance and justice optimization in human–AI collaboration should be investigated (Leitão et al., 2022).

These research directions are of critical importance to advance our efforts to bring together algorithmic fairness and managerial intuition in an ethical manner in the age of artificial intelligence. While revealing the full potential of human–AI collaboration, continuing to preserve human values and social justice should remain the primary goal of research in these fields.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

Agbese, M., Mohanani, R., Khan, A. A., & Abrahamsson, P. (2023). Implementing AI ethics: Making sense of the ethical requirements. *Proceedings of the 62nd European Conference on Artificial Intelligence*, *62*, 10–17. https://doi.org/10.1145/3593434.3593453

Akinnagbe, O. B. (2024). Human-AI collaboration: Enhancing productivity and decision-making. *International Journal of Education Management and Technology*, *2*(3), 387–396. https://doi.org/10.58578/ijemt.v2i3.4209

Auernhammer, J. (2020). Human-centered AI: The role of human-centered design research in the development of AI. *Proceedings of DRS*, *1*, 381–392. https://doi.org/10.21606/drs.2020.282

Awad, E., Levine, S., Loreggia, A., Mattei, N., Rahwan, I., Rossi, F., Talamadupula, K., Tenenbaum, J. B., & Kleiman-Weiner, M. (2024). When is it acceptable to break the rules? Knowledge representation of moral judgements based on empirical data. *Autonomous Agents and Multi-Agent Systems*, *38*(2), 1–28. https://doi.org/10.1007/s10458-024-09667-4

Baker-Brunnbauer, J. (2020). Management perspective of ethics in artificial intelligence. *AI and Ethics*, *1*(2), 173–182. https://doi.org/10.1007/s43681-020-00022-3

Bani-Hani, I., Alawadi, S., & Elmrayyan, N. (2024). AI and the decision-making process: A literature review in healthcare, financial, and technology sectors. *Journal of Decision System*, *33*(3), 1321–1348. https://doi.org/10.1080/12460125.2024.2349425

Bao, Y., Gong, W., & Yang, K. (2023). A literature review of human–AI synergy in decision making: From the perspective of affordance actualization theory. *Systems*, *11*(9), 442. https://doi.org/10.3390/systems11090442

Barnes, S. J. (2020). Information management research and practice in the post-COVID-19 world. *International Journal of Information Management*, *55*, 102175. https://doi.org/10.1016/j.ijinfomgt.2020.102175

Barsotti, F., & Koçer, R. G. (2022). MinMax fairness: From Rawlsian theory of justice to solution for algorithmic bias. *AI & Society*, *39*(3), 961–981. https://doi.org/10.1007/s00146-022-01577-x

Binns, R. (2020). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*, *16*(1), 197–212. https://doi.org/10.1111/rego.12358

Binns, R., van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). "It's reducing a human being to a percentage"; Perceptions of justice in algorithmic decisions. *arXiv (Cornell University)*. http://arxiv.org/abs/1801.10408

Bleher, H., & Braun, M. (2023). Reflections on putting AI ethics into practice: How three AI ethics approaches conceptualize theory and practice. *Science and Engineering Ethics*, *29*(3), 1–23. https://doi.org/10.1007/s11948-023-00443-3

Brendel, A., Mirbabaie, M., Lembcke, T.-B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability*, *13*(4), 1974. https://doi.org/10.3390/su13041974

Buyl, M., & De Bie, T. (2024). Inherent limitations of AI fairness. *Communications of the ACM*, *67*(2), 48–56. https://doi.org/10.1145/3624700

Cañas, J. J. (2022). AI and ethics when human beings collaborate with AI agents. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.836650

Chen, N., Hu, M., & Li, W. (2022). Algorithmic decision-making safeguarded by human knowledge. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4281867

Chen, V., Liao, Q. V., Vaughan, J. W., & Bansal, G. (2023). Understanding the role of human intuition on reliance in human-AI decision-making with explanations. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–26. https://doi.org/10.1145/3610219

Chen, X. (2023). Ethical governance of AI: An integrated approach via human-in-the-loop machine learning. *Communications in Mathematical Sciences and Fluid Mechanics*, *29*, 1–13. https://doi.org/10.3390/cmsf2023008029

Chun, J., de Witt, C. S., & Elkins, K. (2025). Comparative global AI regulation: Policy perspectives from the EU, China, and the US. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5104429

Corrêa, N. K., Galvão, C., Santos, J. W., Pino, C. D., Pinto, E. P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & Oliveira, N. de. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, *4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Cremer, D. D. (2024). Opinion piece: On the ethics of a pending AI crisis in business. *AI and Ethics*, *5*(1), 101–105. https://doi.org/10.1007/s43681-024-00551-1

Cremer, D. D., & Narayanan, D. (2023). How AI tools can—and cannot—help organizations become more ethical. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1093712

Curto, G., & Comim, F. (2023). SAF: Stakeholders' agreement on fairness in the practice of machine learning development. *Science and Engineering Ethics*, *29*(4), 1–28. https://doi.org/10.1007/s11948-023-00448-y

d'Aquin, M., Troullinou, P., O'Connor, N. E., Cullen, A., Faller, G., & Holden, L. (2018). Towards an "ethics by design" methodology for AI research projects. *Proceedings of the 54th Annual Convention of the Association for Computing Machinery*, *54*, 1–6. https://doi.org/10.1145/3278721.3278765

Dé, R., Pandey, N., & Pal, A. (2020). Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International Journal of Information Management*, *55*, 102171. https://doi.org/10.1016/j.ijinfomgt.2020.102171

Decker, M. C., Wegner, L., & Leicht-Scholten, C. (2024). Procedural fairness in algorithmic decision-making: The role of public engagement. *Ethics and Information Technology*, *27*(1), 1–13. https://doi.org/10.1007/s10676-024-09811-4

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Demartini, G., Roitero, K., & Mizzaro, S. (2023). Data bias management. *Communications of the ACM*, *67*(1), 28–30. https://doi.org/10.1145/3611641

Deming, D., Ong, C., & Summers, L. H. (2025). Technological disruption in the labor market. *National Bureau of Economic Research Working Paper*, 33323. https://doi.org/10.3386/w33323

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J. S., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2019). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2023). The extent of algorithm aversion in decision-making situations with varying gravity. *PLoS ONE*, *18*(2), e0278751. https://doi.org/10.1371/journal.pone.0278751

Fu, R., Huang, Y., & Singh, P. V. (2020). AI and algorithmic bias: Source, detection, mitigation and implications. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3681517

García-Murillo, M., & MacInnes, I. (2019). The impact of AI on employment: A historical account of its evolution. *RePEc: Research Papers in Economics*. https://www.econstor.eu/bitstream/10419/205178/1/Garcia-Murillo-MacInnes.pdf

Garibay, Ö. Ö., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C. K., Falco, G., Fiore, S. M., Garibay,

I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J. T., Konstan, J. A., Koon, S., López-González, M., Maifeld-Carucci, I., & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human-Computer Interaction*, *39*(3), 391–411. https://doi.org/10.1080/10447318.2022.2153320

Geng, B., & Varshney, P. K. (2022). Human-machine collaboration for smart decision making: Current trends and future opportunities. *Proceedings of the 61st Conference on Information Sciences and Systems*, *61*, 1–6. https://doi.org/10.1109/cic56439.2022.00019

Gkeredakis, E., Swan, J., Nicolini, D., & Tsoukas, H. (2023). What is the right thing to do? The constitutive role of organizational ethical frameworks in collective ethical sensemaking. *Human Relations*, *77*(11), 1650–1679. https://doi.org/10.1177/00187267231205165

Grgić-Hlača, N., Castelluccia, C., & Gummadi, K. P. (2022). Taking advice from (dis)similar machines: The impact of human-machine similarity on machine-assisted decision-making. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *10*(1), 74–83. https://doi.org/10.1609/hcomp.v10i1.21989

Guan, H., Dong, L., & Zhao, A. (2022). Ethical risk factors and mechanisms in artificial intelligence decision making. *Behavioral Sciences*, *12*(9), 343. https://doi.org/10.3390/bs12090343

Guszcza, J., Danks, D., Fox, C. R., Hammond, K. J., Ho, D. E., Imas, A., Landay, J. A., Levi, M., Logg, J. M., Picard, R. W., Raghavan, M., Stanger, A., Ugolnik, Z., & Woolley, A. W. (2022). Hybrid intelligence: A paradigm for more responsible practice. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4301478

Hao, X., Demir, E., & Eyers, D. (2024). Exploring collaborative decision-making: A quasi-experimental study of human and generative AI interaction. *Technology in Society*, *78*, 102662. https://doi.org/10.1016/j.techsoc.2024.102662

Helberger, N., Araujo, T., & de Vreese, C. H. (2020). Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review*, *39*, 105456. https://doi.org/10.1016/j.clsr.2020.105456

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human-AI collaboration: Concept, sources, and evidence. *European Journal of Information Systems*, *1*, 1–21. https://doi.org/10.1080/0960085x.2025.2475962

Hossain, S., Fernando, M., & Akter, S. (2025). The influence of artificial intelligence-driven capabilities on responsible leadership: A future research agenda. *Journal of Management & Organization*, *1*, 1–18. https://doi.org/10.1017/jmo.2025.10010

Janhonen, J. (2023). Socialisation approach to AI value acquisition: Enabling flexible ethical navigation with built-in receptiveness to social influence. *AI and Ethics*, *3*(4), 1361–1379. https://doi.org/10.1007/s43681-023-00372-8

Jauernig, J., Uhl, M., & Walkowitz, G. (2022). People prefer moral discretion to algorithms: Algorithm aversion beyond intransparency. *Philosophy & Technology*, *35*(1), 1–22. https://doi.org/10.1007/s13347-021-00495-y

Jedličková, A. (2024). Ethical approaches in designing autonomous and intelligent systems: A comprehensive survey towards responsible development. *AI & Society*, *39*(4), 1639–1665. https://doi.org/10.1007/s00146-024-02040-9

Jonker, C. M., Siebert, L. C., & Murukannaiah, P. K. (2024). Reflective hybrid intelligence for meaningful human control in decision-support systems. In *Edward Elgar Publishing eBooks* (pp. 188–208). Edward Elgar Publishing. https://doi.org/10.4337/9781802204131.00019

Kandasamy, U. C. (2024). Ethical leadership in the age of AI challenges, opportunities and framework for ethical leadership. *arXiv (Cornell University)*. https://doi.org/10.48550/ARXIV.2410.18095

Karakuş, N., Gedik, K., & Kazazoğlu, S. (2025). Ethical decision-making in education: A comparative study of teachers and artificial intelligence in ethical dilemmas. *Behavioral Sciences*, *15*(4), 469. https://doi.org/10.3390/bs15040469

Kiškis, M. (2023). Legal framework for the coexistence of humans and conscious AI. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.1205465

Koponen, J., Julkunen, S., Laajalahti, A., Turunen, M., & Spitzberg, B. H. (2023). Work characteristics needed by middle managers when leading AI-integrated service teams. *Journal of Service Research*, *26*(4), 570–590. https://doi.org/10.1177/10946705231220462

Kuppler, M., Kern, C., Bach, R. L., & Kreuter, F. (2022). From fair predictions to just decisions? Conceptualizing algorithmic fairness and distributive justice in the context of data-driven decision-making. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.883999

Lai, S., Kim, J., Kunievsky, N., Potter, Y., & Evans, J. (2025). Biased AI improves human decision-making but reduces trust.

arXiv (Cornell University). https://doi.org/10.48550/ARXIV.2508.09297

Lai, V., Chen, C., Smith, A., Liao, Q. V., & Tan, C. (2023). Towards a science of human-AI decision making: An overview of design space in empirical human-subject studies. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 1369–1379. https://doi.org/10.1145/3593013.3594087

Leitão, D., Saleiro, P., Figueiredo, M. A. T., & Bizarro, P. (2022). Human-AI collaboration in decision-making: Beyond learning to defer. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2206.13202

Levi-Faur, D., Kariv-Teitelbaum, Y., & Medzini, R. (2021). Regulatory governance: History, theories, strategies, and challenges. In *Oxford Research Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.1430

Leyer, M., & Schneider, S. (2021). Decision augmentation and automation with artificial intelligence: Threat or opportunity for managers? *Business Horizons*, *64*(5), 711–720. https://doi.org/10.1016/j.bushor.2021.02.026

Liao, S. M. (2023). Ethics of AI and health care: Towards a substantive human rights framework. *Topoi*, *42*(3), 857–865. https://doi.org/10.1007/s11245-023-09911-8

Lou, B., Lu, T., Raghu, T. S., & Zhang, Y. (2025). Unraveling human-AI teaming: A review and outlook. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5211067

Maiti, M., Kayal, P., & Vujko, A. (2025). A study on ethical implications of artificial intelligence adoption in business: Challenges and best practices. *Future Business Journal*, *11*(1). https://doi.org/10.1186/s43093-025-00462-5

Minssen, T., Smith, H. L., & Mann, S. P. (2025). The European AI Act: Promises, perils, & trade-offs in a complex regulatory ecosystem. In A. O. Garcia & J. F. Negrete (Eds.), *Inteligencia Artificial: Hacia una Nueva Era en la Historia de la Humanidad* (1st ed., pp. 330–372). Instituto para el Desarrollo Industrial y la Transformación Digital A.C. (INADI). https://www.inadi.mx/_files/ugd/527291_c90b1434a1014e2686e235c9e4b7eb73.pdf

Müller, J. (2021). Menschliche intuition als kreatives Korrektiv für KI. *Controlling & Management Review*, *65*(1), 62–67. https://doi.org/10.1007/s12176-020-0363-4

Mulligan, D. K., Kroll, J. A., Kohli, N., & Wong, R. Y. (2019). This thing called fairness: Disciplinary confusion realizing a value in technology. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–33. https://doi.org/10.1145/3359221

Nakao, Y., Strappelli, L., Stumpf, S., Naseer, A., Regoli, D., & Gamba, G. D. (2022). Towards responsible AI: A design space exploration of human-centered artificial intelligence user interfaces to investigate fairness. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2206.00474

Nasir, S., Khan, R. A., & Bai, S. (2024). Ethical framework for harnessing the power of AI in healthcare and beyond. *IEEE Access*, *12*, 31014–31023. https://doi.org/10.1109/access.2024.3369914

Noriega, P., Verhagen, H., Padget, J., & d'Inverno, M. (2021). Ethical online AI systems through conscientious design. *IEEE Internet Computing*, *25*(6), 58–66. https://doi.org/10.1109/mic.2021.3098324

Oppioli, M., Sousa, M. J., Sousa, M., & de Nuccio, E. (2023). The role of artificial intelligence for management decision: A structured literature review. *Management Decision*, *61*(11), 3230–3253. https://doi.org/10.1108/md-08-2023-1331

Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Tal, A. S., AlanHartman, & Kuflik, T. (2021). Mitigating bias in algorithmic systems—A fish-eye view. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2103.16953

Osasona, F., Amoo, O. O., Atadoga, A., Abrahams, T. O., Farayola, O. A., & Ayinla, B. S. (2024). Reviewing the ethical implications of AI in decision making processes. *International Journal of Management & Entrepreneurship Research*, *6*(2), 322–338. https://doi.org/10.51594/ijmer.v6i2.773

Peng, A., & Simard-Halm, M. (2020). The perils of objectivity: Towards a normative framework for fair judicial decision-making. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (p. 343). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375869

Pflanzer, M., Traylor, Z., Lyons, J. B., Dubljević, V., & Nam, C. S. (2022). Ethics in human–AI teaming: Principles and perspectives. *AI and Ethics*, *3*(3), 917–932. https://doi.org/10.1007/s43681-022-00214-z

Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*, *3*(3), 699–712. https://doi.org/10.1007/s43681-023-00258-9

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., Antonelli, G., Awadie, H., Bernhofer, S., Carballal, S., Dinis-Ribeiro, M., Fernández-Clotet, A., Fernández-Esparrach, G., Gralnek, I. M., Higasa, Y., Hirabayashi, T., Hirai, T., Iwatate, M., Kawano, M., Mader, M., ... Cherubini, A. (2022). Experimental evidence of effective human–AI

collaboration in medical decision-making. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-18751-2

Risser, L., Picard, A. M., Hervier, L., & Loubes, J. (2022). A survey of identification and mitigation of machine learning algorithmic biases in image analysis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2210.04491

Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2022). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*, *33*(1), 100925. https://doi.org/10.1016/j.hrmr.2022.100925

Saghafian, S., & Idan, L. (2024). Effective generative AI: The human-algorithm centaur. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2406.10942

Salman, C. (2024). The universal fix: Computing pacification in times of crisis. *Deep Blue (University of Michigan)*. https://doi.org/10.7302/23788

Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding human-centred AI: A review of its defining elements and a research agenda. *Behaviour and Information Technology*, *44*(1), 1–24. https://doi.org/10.1080/0144929x.2024.2448719

Seeamber, R., & Badea, C. (2023). If our aim is to build morality into an artificial agent, how might we begin to go about doing so? *IEEE Intelligent Systems*, *38*(6), 35–41. https://doi.org/10.1109/mis.2023.3320875

Seufert, S., & Meier, C. (2023). Hybride Intelligenz: Zusammenarbeit mit KI-Assistenzsystemen in wissensintensiven Bereichen. *HMD Praxis der Wirtschaftsinformatik*, *60*(6), 1194–1209. https://doi.org/10.1365/s40702-023-01012-9

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence*, *6*. https://doi.org/10.3389/frai.2023.976887

Söllner, M., Arnold, T., Benlian, A., Bretschneider, U., Knight, C. L., Ohly, S., Rudkowski, L., Schreiber, G., & Wendt, D. H. (2025). ChatGPT and beyond: Exploring the responsible use of generative AI in the workplace. *Business & Information Systems Engineering*, *67*(2), 1–13. https://doi.org/10.1007/s12599-025-00932-8

Spitale, G. (2020). Making sense in the flood. How to cope with the massive flow of digital information in medical ethics. *Heliyon*, *6*(7), e04426. https://doi.org/10.1016/j.heliyon.2020.e04426

Stahl, B. C., Andreou, A. G., Brey, P., Hatzakis, T., Kirichenko, A. S., Macnish, K., Shaelou, S. L., Patel, A., Ryan, M., & Wright, D. (2020). Artificial intelligence for human flourishing–Beyond principles for machine learning. *Journal of Business Research*, *124*, 374–385. https://doi.org/10.1016/j.jbusres.2020.11.030

Szafran, D., & Bach, R. L. (2024). "The human must remain the central focus": Subjective fairness perceptions in automated decision-making. *Minds and Machines*, *34*(3), 1–28. https://doi.org/10.1007/s11023-024-09684-y

Tahaei, M., Constantinides, M., & Quercia, D. (2023). A systematic literature review of human-centered, ethical, and responsible AI. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2302.05284

Trunk, A., Birkel, H., & Hartmann, E. (2020). On the current state of combining human and artificial intelligence for strategic organizational decision making. *BuR - Business Research*, *13*(3), 875–911. https://doi.org/10.1007/s40685-020-00133-x

Uddin, A. S. M. A. (2023). The era of AI: Upholding ethical leadership. *Open Journal of Leadership*, *12*(4), 400–416. https://doi.org/10.4236/ojl.2023.124019

Vakali, A., & Tantalaki, N. (2024). Rolling in the deep of cognitive and AI biases. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2407.21202

Valença, L., & Santos, R. de S. (2025). Justiça algorítmica: Instrumentalização, limites conceituais e desafios na engenharia de software. In *Anais do XXII Simpósio Brasileiro de Sistemas de Informação* (p. 225). https://doi.org/10.5753/wics.2025.8032

Vincent, V. (2021). Integrating intuition and artificial intelligence in organizational decision-making. *Business Horizons*, *64*(4), 425–436. https://doi.org/10.1016/j.bushor.2021.02.008

Waller, M., Rodrigues, O., & Cocarascu, O. (2023). Bias mitigation methods for binary classification decision-making systems: Survey and recommendations. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2305.20020

Wang, D., Churchill, E. F., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From human-human collaboration to human-AI collaboration. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3334480.3381069

Wu, W., & Liu, S. (2023a). A comprehensive review and systematic analysis of artificial intelligence regulation policies.

arXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.12218

Wu, W., & Liu, S. (2023b). Dilemma of the artificial intelligence regulatory landscape. *Communications of the ACM*, *66*(9), 28–30. https://doi.org/10.1145/3584665

Xu, W., & Gao, Z. (2024). An intelligent sociotechnical systems (iSTS) framework: Enabling a hierarchical human-centered AI (hHCAI) approach. *arXiv (Cornell University)*. https://doi.org/10.48550/ARXIV.2401.03223

Xu, W., Gao, Z., & Dainoff, M. (2023). An HCAI methodological framework (HCAI-MF): Putting it into action to enable human-centered AI. *arXiv (Cornell University)*. https://doi.org/10.48550/ARXIV.2311.16027

Yamamoto, H., & Suzuki, T. (2025). Exploring condition in which people accept AI over human judgements on justified defection. *Scientific Reports*, *15*(1), 1–12. https://doi.org/10.1038/s41598-025-87170-w

Yue, Y., & Shyu, J. Z. (2023). An overview of research on human-centered design in the development of artificial general intelligence. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2309.12352

Zaidan, E., & Ibrahim, I. A. (2024). AI governance in a complex and rapidly changing regulatory landscape: A global perspective. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03560-x

Zhang, H., Wang, L., Sheng, Y., Xu, X., Mankoff, J., & Dey, A. K. (2023). A framework for designing fair ubiquitous computing systems. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–28. https://doi.org/10.1145/3594739.3610677