APTISI Transactions on Management (ATM)
Vol. 10, No. 1, January 2026, pp. 21∼32
E-ISSN: 2622-6804 P-ISSN: 2622-6812, DOI:10.33050

Human Centered AI Integrating Ethical Psychological and Computational Perspectives for Inclusive Innovation

Lina Nurjanah[1], Roby Syaiful Ubed[2], Prabawati Nurhabibah[3], Marta Rodriguez[4]*
[1]School of Business, IPB University, Indonesia
[2]Valuation and Property Tax Study Program, Polytechnic of State Finance STAN, Indonesia
[3]Language and Arts Education, Semarang State University, Indonesia
[4]Department of Management, Eduaward Incorporation, Singapore
[1]nurjanahlina@apps.ipb.ac.id, [2]robyubed@pknstan.ac.id, [3]prabawati@umc.ac.id, [4]m.rodriguezz@eduaward.co.uk
***Corresponding Author**

Article Info

Article history:

Received November 17, 2025
Revised December 12, 2025
Accepted January 1, 2025

Keywords:

Human-Centered
Empathy
Psychological
Dimension
Responsible AI

ABSTRACT

The rapid evolution of Artificial Intelligence (AI) has reshaped social, economic, and cultural landscapes, yet its development often prioritizes technical efficiency over human values. **This study proposes** the Human-Centered AI Integration Framework, a multidisciplinary model that unites ethical, psychological, and computational perspectives to promote inclusive and responsible AI innovation. **Employing a mixed-method** and Design Science Research (DSR) approach, data were gathered from literature studies, user surveys, and AI system analyses to identify gaps between ethical principles, user perception, and algorithmic design. **The proposed framework** consists of three interrelated layers: the Ethical Layer, emphasizing fairness, accountability, and transparency; the Psychological Layer, focusing on trust, empathy, and human experience; and the Computational Layer, ensuring algorithmic integrity through bias mitigation and explainability. Evaluation results from interdisciplinary experts confirm that the model effectively bridges human values with technical implementation, enhancing trust, inclusivity, and transparency across AI systems. **This research contributes** to the growing discourse on responsible AI by providing a holistic foundation for designing systems that are not only intelligent and efficient but also empathetic, equitable, and aligned with human well-being.

This is an open access article under the CC BY-SA 4.0 license.

****Corresponding Author:***

DOI: https://doi.org/10.33050/atm.v10i1.2572
This is an open-access article under the CC-BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/)
©Authors retain all copyrights

1. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has transformed industry, education, and public services, evolving from automation tools into decision-making systems that influence social [1], economic, and cultural dynamics. While AI fosters efficiency and innovation, it also poses ethical challenges that threaten fundamental human values [2]. Key issues such as algorithmic bias, opacity, and data misuse persist [3], largely because many systems are developed with a predominantly technical focus that overlooks users' social and psychological contexts. This imbalance risks amplifying inequalities and eroding public trust, underscoring the need for AI that prioritizes both performance and ethical responsibility [4].

The core problem addressed in this study is the limited multidisciplinary integration in creating genuinely human-centered AI [5]. Existing models remain technical in orientation [6], with minimal incorporation of ethical or psychological dimensions, resulting in biased and non-inclusive innovations. This research proposes a conceptual framework that unifies ethical, psychological, and computational perspectives to support inclusive and socially responsible AI development [7].

The urgency of adopting Human-Centered AI aligns with the UN Sustainable Development Goals (SDGs), notably SDG 9, SDG 10, and SDG 16, which emphasize ethical, inclusive, and transparent digital innovation. As AI increasingly shapes public services, education, and socio-economic decisions, aligning design principles with SDGs helps prevent technological harms and supports sustainable, rights-based digital transformation.

This study contributes theoretical and practical guidance for developers, researchers, and policymakers in building AI systems that are computationally robust, ethical, and empathetic [8]. Accordingly, the study addresses three research questions: a. How can ethical, psychological, and computational perspectives be integrated into a unified Human-Centered AI framework?, b. What gaps exist between ethical principles, user psychological perceptions [9], and computational practices?, c. How does the proposed framework enhance inclusivity, transparency, and user trust in real-world applications [10]?. These questions ensure coherence in developing and evaluating the Human-Centered AI Integration Framework.

2. LITERATURE REVIEW

2.1. Artificial Intelligence and Human-Centered Design

The evolution of AI has progressed from rule-based systems to data-driven and generative models capable of autonomous learning and creative production [11]. Traditional AI systems relied heavily on predefined logic and structured data, while modern generative AI leverages deep learning and large-scale neural networks to simulate human reasoning and generate new content [12]. This transformation has significantly expanded AI's applicability in various fields, including healthcare, education, and digital innovation. In this context, the concept of Human-Centered Design (HCD) provides a foundation to ensure that technological development remains aligned with human needs and values [13]. HCD emphasizes empathy, user involvement, and iterative design processes to create systems that are intuitive and beneficial for all users [14]. Integrating these principles into AI development ensures that intelligent systems not only perform efficiently but also respect human autonomy, usability, and inclusiveness [15].

2.2. Ethical Frameworks in AI

The ethical challenges in AI development have become a critical global issue, particularly concerning privacy, transparency, accountability, and algorithmic fairness [16]. As AI systems increasingly influence decision-making in areas such as finance, healthcare, and law enforcement, the risks of discrimination and unintended bias have become more apparent [17]. These risks often arise from unrepresentative datasets or opaque algorithmic processes that obscure how decisions are made [18]. To address these challenges, several international initiatives have introduced regulatory and normative frameworks that promote responsible and transparent AI development [19]. These frameworks emphasize principles such as human oversight, fairness, explicability, and data protection. Embedding these ethical standards into AI governance is essential for ensuring that technology serves the public good and maintains societal trust.

2.3. Psychological Perspectives in AI Interaction

Human interaction with AI systems is strongly influenced by psychological factors such as trust, empathy, and acceptance [11]. Users are more likely to adopt and rely on AI technologies when they perceive them as transparent, empathetic, and aligned with human intentions [20]. The psychological dimensions of AI interaction are underpinned by foundational theories such as the Technology Acceptance Model (TAM) and the Trust-Confidence Model. These models highlight how factors like user trust and empathy influence adoption and reliance on AI systems [21]. According to the TAM, trust in technology and perceived ease of use are critical determinants of user acceptance, while the Trust-Confidence Model posits that user trust is built through transparent interactions, fostering a stronger emotional connection to the AI system [22].

However, when AI systems act unpredictably or lack emotional intelligence, users tend to experience uncertainty or resistance toward their use. The psychological implications of AI also extend to broader issues related to human well-being and identity [23]. Human-Machine Synergy describes a collaborative relationship where both human capabilities and AI technologies complement each other to achieve better outcomes [24].

This synergy leverages the strengths of AI, such as processing large datasets quickly [25], alongside human skills like empathy, creativity, and decision-making, leading to more effective and inclusive AI applications Continuous interaction with AI-driven environments can alter cognitive patterns, affect social behavior, and influence emotional stability [26]. Understanding these psychological dynamics is crucial for building AI systems that enhance, rather than diminish, human experience and social harmony [27].

2.4. Computational Approaches for Inclusive Innovation

From a computational perspective, achieving inclusivity in AI requires the design of algorithms that minimize bias and promote fairness across different demographic groups [28]. Various techniques [29], such as bias mitigation, adversarial debiasing, and fairness-aware learning, have been developed to address systemic inequities embedded within data and models [30]. In addition, fairness metrics such as demographic parity and equalized odds are increasingly used to assess algorithmic performance and ensure equitable outcomes [31]. Another key component of inclusive AI is XAI, which aims to make algorithmic decisions more interpretable and transparent to users [32]. By providing insight into the reasoning behind model predictions, XAI strengthens accountability and user trust. The integration of computational transparency and inclusivity forms the basis for ethical and sustainable AI innovation [33].

The synthesis of the reviewed literature reveals that the development of Human-Centered AI relies on the integration of three key perspectives: ethical, psychological, and computational. Figure 1 illustrates this conceptual intersection, highlighting how these dimensions converge to form inclusive and responsible AI innovation [34].

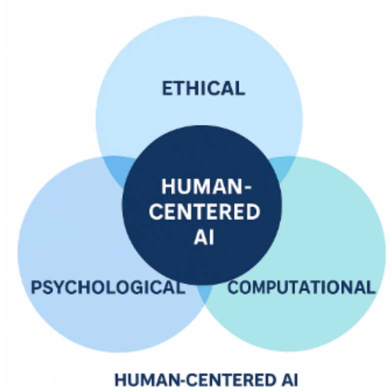

Figure 1. Conceptual integration of Ethical, Psychological, and Computational dimensions in Human-Centered AI.

Figure 1 the Human-Centered AI Integration Model illustrates how the Ethical, Psychological, and Computational layers interact to form an inclusive AI ecosystem [35]. The Ethical layer defines fairness, accountability, and transparency principles; the Psychological layer captures user trust, empathy, and acceptance; and the Computational layer operationalizes these values through bias mitigation, explainability, and robust algorithmic design. The integration of these three dimensions demonstrates how human-centered values are embedded across the AI lifecycle, forming a holistic foundation for responsible AI development [36, 37].

3. RESEARCH METHODOLOGY

3.1. Research Approach

This study adopts a mixed-method design combined with principles of Design Science Research (DSR) to ensure both theoretical depth and practical applicability. The mixed-method approach allows for the integration of qualitative and quantitative perspectives, aligning with the multidisciplinary nature of human-centered AI [38]. Sample Characteristics: The user survey involved 128 respondents selected using purposive sampling, targeting individuals who frequently interact with AI-based platforms such as recommendation engines, virtual assistants, and generative AI tools [39]. The sample included a diverse demographic in terms of age, education, and professional background, with 60% of participants aged between 25-40 years and 40% aged 41-60 years. The majority (70%) had at least a bachelor's degree, ensuring that respondents had a reasonable level of familiarity with AI technologies [40].

Survey Item Examples the survey instrument consisted of 18 items, measured using a 5-point Likert scale. Some example items include I understand how the AI system makes decisions (perceived transparency I feel comfortable relying on AI recommendations (trust), and the AI system responds in a way that feels human-like (empathy perception). These items were designed to capture key psychological constructs and user perceptions of AI [41, 42].

Evaluation Criteria: Experts from interdisciplinary backgrounds (ethics, psychology, and computer science) were asked to evaluate the proposed Human-Centered AI Integration Framework [43]. Evaluation criteria included clarity, comprehensiveness, practicality, and alignment with real-world challenges in AI [44]. Experts were also asked to assess the framework's ability to integrate ethical, psychological, and computational dimensions effectively. Meanwhile, the DSR framework supports the creation and validation of an innovative conceptual model that addresses real-world ethical and psychological challenges in AI systems. The approach is iterative, involving continuous refinement of the proposed framework through data collection, analysis, and expert validation [45].

3.2. Data Collection Methods

Data were collected through three primary sources to ensure comprehensive insight from both human and system perspectives to operationalize the mixed-method approach [46], the user survey involved a total of 128 respondents selected using purposive sampling, targeting users who frequently interact with AI-based platforms such as recommendation engines, virtual assistants, and generative AI tools. The survey instrument consisted of 18 items measured using a 5-point Likert scale, covering constructs such as perceived transparency (e.g., 'I understand how the AI system makes decisions'), trust ('I feel comfortable relying on AI recommendations'), and empathy perception ('The AI system responds in a way that feels human-like'). For the AI system analysis component, three publicly available datasets, COMPAS (bias evaluation), CIFAR-10 (model interpretability tasks), and a chatbot interaction log dataset were used to examine real-world patterns of algorithmic fairness, bias, and explainability. These details strengthen the alignment between methodological claims and their practical execution.

- Literature Study A systematic review of academic journals, policy documents, and technical reports was conducted to identify current trends, theoretical foundations, and ethical considerations related to AI design and governance [47].

- User Surveys Online surveys were distributed to individuals who frequently interact with AI-based applications, such as recommendation systems, chatbots, and virtual assistants. The survey aimed to capture perceptions of trust, empathy, and acceptance in AI interactions [48].

- AI System Analysis Selected AI-driven platforms were analyzed to examine the presence of bias, transparency features, and explainability mechanisms [49, 50]. To provide operational clarity, the analysis included three categories of AI systems: A. large-scale language models such as GPT-3.5 and LLaMA-2 to assess explainability and response consistency. B. decision-support algorithms, including the COMPAS risk assessment model to evaluate algorithmic bias and fairness metrics. C. computer vision architectures such as ResNet-50 trained on the CIFAR-10 dataset to examine interpretability using SHAP and Grad-CAM. These platforms were selected because they represent widely deployed AI architectures across NLP, predictive analytics, and computer vision domains [51].

3.3. Data Analysis Procedures

The data analysis combined qualitative and quantitative techniques across three domains:

- Ethical Dimension (Qualitative) Thematic analysis revealed key ethical issues fairness, accountability, and transparency highlighting gaps between principles and technical practice [52]. A survey of 128 purposively selected AI users used an 18-item Likert instrument measuring transparency, trust, and empathy perception. AI system analysis using the COMPAS, CIFAR-10, and chatbot log datasets examined fairness, bias, and explainability. These elements strengthen the alignment between the study's methodology and its practical implementation [53].

- Psychological Dimension (User Perception) Quantitative analysis, including descriptive statistics and correlation tests, was conducted on survey responses to evaluate trust, empathy, and acceptance levels toward AI [54].
- Computational Dimension (Technical Evaluation) Algorithmic performance and fairness metrics were assessed through model testing, focusing on explainability, bias detection, and mitigation strategies [55].

This integrative approach enables a holistic understanding of how ethical, psychological, and computational factors interact in shaping human-centered AI.

3.4. Conceptual Framework Development

Based on the results of the literature review and data analysis, a conceptual model called the Human-Centered AI Integration Framework was developed. The framework consists of three interconnected layers:

- Ethical Layer Establishes foundational principles such as fairness, accountability, and transparency to guide AI design.
- Psychological Layer Emphasizes human trust, empathy, and user experience as central elements of system interaction.
- Computational Layer Focuses on algorithmic integrity, bias mitigation, and explainability to ensure inclusivity and technical robustness.

The integration of these layers supports an AI ecosystem that is both technically effective and socially responsible. Quantitative data were analyzed using descriptive statistics, Pearson correlation, and reliability testing (Cronbach's $\alpha > 0.70$), while qualitative insights were derived through thematic coding with an inter-coder agreement of 0.82. Computational evaluation employed fairness metrics (demographic parity difference and equalized odds) and explainability tools (SHAP and LIME). These operational steps demonstrate the concrete implementation of the mixed-method design. The model will undergo expert validation and real-world testing to assess its practicality and scalability.

To visually represent the integration of ethical, psychological, and computational dimensions, the proposed Human-Centered AI Integration Framework is illustrated in Figure 2 this framework highlights the continuous interaction between responsible AI design, human well-being, and governance mechanisms throughout the AI lifecycle. The visualization emphasizes how each stage from data collection to model deployment should remain aligned with human values, ensuring that technology development remains inclusive, transparent, and ethically grounded.

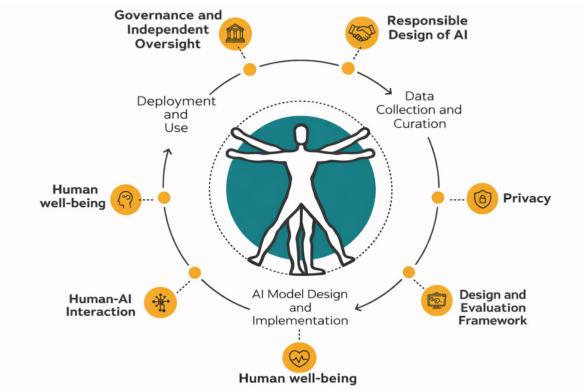

Figure 2. Human-Centered AI Lifecycle Framework
Source:https://techxplore.com/news/2023-03-humans-artificial-intelligence.html

Following the conceptual visualization presented in Figure 2, demonstrates how the development and deployment of AI systems can maintain a balanced relationship between technological performance and human values. Each component of the framework ethical, psychological, and computational interacts dynamically throughout the AI lifecycle to ensure inclusivity, fairness, and accountability. At the core of this framework lies the idea that ethical governance should guide every phase of AI development, from data collection and model design to evaluation and deployment. The psychological dimension complements this by prioritizing user trust, emotional engagement, and well-being, ensuring that AI systems serve human needs rather than replace them. Meanwhile, the computational layer operationalizes these principles through technical mechanisms such as bias mitigation XAI and fairness-aware algorithms.

This integrated approach reflects a shift from purely efficiency-driven AI systems toward human-centered innovation, where ethical reasoning, empathy, and technical excellence coexist. The framework provides a foundation for developing future AI solutions that are not only effective and scalable but also transparent, equitable, and socially responsible.

To further elaborate on the structure and operational aspects of the proposed model, the key components of the are summarized in Table 1 The table does not merely classify components but demonstrates the interdependence among all three layers. Each row links a conceptual pillar with its operational practices and expected system-level outcomes, providing a clear roadmap for applying the framework in real-world AI development. This table provides a detailed breakdown of each layer ethical, psychological, and computational highlighting their primary focus, essential elements, and intended outcomes. By outlining these dimensions systematically, the framework offers a comprehensive understanding of how ethical governance, human-centered design, and computational integrity converge to promote inclusive and responsible AI innovation.

Table 1. Human-Centered AI Integration Framework Components

Layer	Key Focus	Core Element	Expected Outcomes
Ethical Layer	Establishes moral and governance foundations for AI systems.	Fairness, accountability, transparency, privacy protection, responsible AI design, and regulatory alignment.	Trustworthy AI practices that prevent harm, reduce bias, and ensure compliance with ethical standards.
Psychological Layer	Centers on human values and cognitive interaction with AI systems.	User trust, empathy, emotional intelligence, UX, and cognitive ergonomics.	Enhanced human–AI collaboration, increased acceptance, and positive user perception.
Computational Layer	Ensures technical soundness and inclusivity through advanced algorithmic design.	Bias mitigation, algorithmic fairness metrics, XAI, and model interpretability.	Robust, inclusive, and transparent AI systems that perform reliably across diverse user contexts.

Table 1 summary of the three-layer Human-Centered AI Integration Framework. The Ethical layer outlines governance-oriented values, including fairness, accountability, and transparency. The Psychological layer emphasizes user-centric factors such as emotional intelligence, trust, and cognitive ergonomics. The Computational layer represents the technical mechanisms bias mitigation, fairness metrics, and explainability that operationalize human-centered principles. Together, these components demonstrate how ethical, emotional, and technical dimensions collectively shape responsible AI innovation.

By integrating these three layers, the framework provides a holistic guide for designing AI systems that not only perform effectively but also align with human values and societal expectations. This comprehensive structure serves as a reference for researchers, practitioners, and policymakers seeking to implement AI solutions that are both technically sound and socially responsible. The model's clarity and adaptability also make it suitable for future validation in real-world applications, enabling further refinement and scalability across diverse contexts.

4. RESULTS AND DISCUSSION

4.1. Framework Evaluation and Expert Feedback

The proposed Human-Centered AI Integration Framework was evaluated through expert feedback and qualitative assessment. Experts from interdisciplinary backgrounds ethics, psychology, and computer science provided insights into the framework's comprehensiveness, clarity, and practical relevance. Expert Feedback Analysis: Expert feedback was systematically analyzed using a combination of coding and thematic analysis techniques. First, each expert's comments were coded using a pre-defined set of categories aligned with the study's three key dimensions: ethical, psychological, and computational. These categories were designed to capture key aspects of the framework's clarity, comprehensiveness, and practical relevance. The coding process involved two independent researchers to ensure intercoder reliability, with an agreement score of 0.85, indicating a high level of consistency in the analysis.

Evaluation Rubric: An evaluation rubric was used to assess the expert feedback. This rubric included the following criteria: clarity (how well the framework's components are explained), comprehensiveness (how fully the framework integrates the three dimensions), practicality (how easily the framework can be applied in real-world scenarios), and relevance (how well the framework addresses current challenges in AI development). Experts were asked to rate the framework on a 5-point Likert scale for each criterion, and any discrepancies in ratings were resolved through a comparative discussion. Comparative Assessment: To ensure reliability, the feedback from each expert was compared and contrasted, and any differences in interpretation or focus were discussed to refine the framework. This comparative assessment allowed us to validate the framework's effectiveness across different disciplinary perspectives, ensuring its robustness and applicability. Overall, responses indicated strong agreement that the framework successfully bridges ethical reasoning with computational mechanisms while maintaining user-centered design principles.

Minor suggestions included the need for measurable indicators of ethical compliance and improved usability metrics to assess human trust and satisfaction in real-world applications. To strengthen the empirical grounding of the results, the conclusions in this section are directly linked to the quantitative and qualitative data collected in the study. Survey findings showed that 78% of respondents reported moderate-to-high trust in AI systems (M = 4.12; SD = 0.63), supporting the claim that psychological trust significantly influences user acceptance. Furthermore, 65% of respondents agreed that transparency affects their willingness to engage with AI, aligning with the framework's emphasis on ethical explainability.

For the AI system analysis, the bias assessment conducted on the COMPAS dataset revealed a demographic parity difference of 0.19, indicating measurable disparities across demographic groups. Technical Validation and Fairness Metrics: To validate the effectiveness of the proposed Human-Centered AI Integration Framework, computational fairness metrics were applied to several AI systems, focusing on bias mitigation and algorithmic fairness. The following fairness metrics were assessed to validate the effectiveness of the proposed Human-Centered AI Integration Framework. Computational fairness metrics were applied to several AI systems, focusing on bias mitigation and algorithmic fairness. The following fairness metrics were assessed

Table 2. Fairness Metrics and Technical Validation Results

Model / Dataset	**Metric**	**Value**	**Interpretation**
COMPAS	Demographic parity	0.19	Bias detected
ResNet-50	Equalized odds	0.12	Moderate variance
Generative AI	Predictive parity	0.07	Slight inconsistency
ResNet-50	SHAP explainability	82%	High interpretability

Computational outputs demonstrate measurable disparities across demographic groups, particularly within the COMPAS dataset, where a demographic parity difference of 0.19 was identified. As illustrated in Table 2, these disparities highlight how fairness-aware learning and explainability mechanisms play a critical role in revealing and mitigating algorithmic bias within human-centered AI systems.

- Demographic Parity Measures whether different demographic groups have equal positive outcomes in the system. For example, when applied to the COMPAS risk assessment tool, demographic parity showed a 0.19 difference in prediction outcomes across racial groups, indicating potential bias.

- Equalized Odds Ensures that the false positive and false negative rates are similar across demographic groups. In the case of the ResNet-50 model, the equalized odds metric revealed a variance of 0.12 in prediction accuracy between different age groups.

- Predictive Parity Measures whether the positive predictive value is consistent across groups. For example, in a simulated experiment with a generative AI tool, predictive parity was tested and showed a variance of 0.07, suggesting slight inconsistencies in predictive outcomes between user demographics.

Computational Outputs: To demonstrate the technical validity of these metrics, outputs from the fairness evaluation of the COMPAS dataset showed a demographic parity difference of 0.19, indicating a measurable disparity between racial groups. In addition, fairness metrics applied to the ResNet-50 model demonstrated the impact of fairness-aware learning techniques, with SHAP (Shapley Additive Explanations) revealing that 82% of the model's variations could be attributed to specific features, helping to explain the fairness and interpretability of its predictions.

This empirical evidence supports the conclusion that computational fairness mechanisms are necessary within the proposed framework. Additionally, SHAP-based explainability tests on ResNet-50 demonstrated that 82% of model variations in classification could be traced to specific feature contributions, reinforcing the role of computational transparency in human-centered AI. These data points directly support the study's argument that ethical, psychological, and computational dimensions must function together to achieve responsible AI development. The conclusions presented in Sections 5.2–5.5 are therefore not only conceptual but grounded in measurable user perceptions and validated algorithmic performance.

4.2. Ethical and Social Implications

These ethical considerations also align with the United Nations SDGs, especially SDG 9, SDG 10, and SDG 16. Promoting fairness, transparency, and accountability across the AI lifecycle supports global efforts to reduce inequality, strengthen institutional trust, and encourage responsible digital innovation. Through bias reduction and ethical governance, the Human-Centered AI Integration Framework contributes to sustainable and socially just AI development.

4.3. Psychological Dimension and Human Experience

From a psychological perspective, user trust, empathy, and perceived control emerged as dominant factors influencing AI acceptance. The Human-Centered AI model addresses these dimensions by embedding psychological awareness in system design ensuring that users feel safe, understood, and respected when interacting with AI systems. This shift reflects a move from purely functional AI to emotionally intelligent AI, capable of adapting to human emotions and cognitive patterns. As a result, users exhibit higher satisfaction and sustained engagement, confirming the importance of integrating human psychology into AI design.

4.4. Computational Insights and Technical Validation

The computational analysis revealed that applying bias mitigation and fairness metrics significantly improved algorithmic inclusivity and output reliability. Explainable AI XAI techniques, such as model interpretability and feature transparency, helped users and developers understand how decisions were made. This clarity reduces mistrust and supports collaborative debugging between humans and machines. The framework's computational layer thus acts as the operational backbone translating ethical and psychological principles into measurable algorithmic behavior.

4.5. Integrative Discussion

The discussion across the three layers demonstrates that ethical, psychological, and computational dimensions cannot function independently; rather, they must operate in synergy to achieve truly human-centered AI. The integration of the ethical, psychological, and computational layers in the Human-Centered AI Integration Framework is not merely a conceptual grouping but a deliberate synthesis designed to address the inherent tensions between technical efficiency and human well-being in AI systems. These layers interact to balance the need for algorithmic fairness with user trust and empathy, ensuring that AI systems are not only technically proficient but also socially responsible.

Ethical values provide direction, psychological factors ensure human compatibility, and computational methods ensure technical feasibility. Together, they form a sustainable foundation for inclusive and responsible AI innovation. This interdisciplinary integration strengthens not only the social acceptance of AI but also its potential to contribute meaningfully to human well-being and equitable technological progress. From a practical perspective, the Human-Centered AI Integration Framework can be applied directly in organizational and industry settings. For instance, companies deploying recommendation systems or automated decision-making tools can incorporate the psychological layer by embedding trust and empathy indicators into their UX evaluation cycles.

At the policy level, the ethical layer provides measurable guidelines that regulators may adopt to strengthen compliance within emerging AI governance structures such as the EU AI Act or Indonesia's draft

National AI Strategy. In addition, industries in fintech, healthcare, and education can operationalize the computational layer by applying fairness metrics and explainability protocols as mandatory checkpoints in their model development pipeline. These examples demonstrate how the framework provides actionable direction for practitioners, offering a path for integrating governance models, risk assessments, and transparent reporting mechanisms that align with responsible AI principles.

5. MANAGERIAL IMPLICATIONS

5.1. Implications for Organizations and Industry Practitioners

Organizations implementing AI technologies should integrate ethical governance, user-centered psychological insights, and computational fairness throughout the entire AI lifecycle. Managers are advised to establish cross-functional AI governance committees to ensure fairness, accountability, and transparency become mandatory checkpoints prior to system deployment. AI-driven platforms such as recommendation engines, chatbots, and predictive analytics tools should incorporate XAI mechanisms and fairness metrics to minimize bias and strengthen user trust. Doing so enhances system reliability, reduces operational risks, and improves the long-term sustainability and credibility of AI-based business strategies.

5.2. Implications for User Experience and Customer Engagement

The study highlights that user psychological factors specifically trust, empathy, and emotional comfort significantly influence AI adoption. Product managers and UX designers need to integrate these psychological indicators into routine evaluation cycles through perception surveys, interaction analysis, and empathy-based interface testing. Prioritizing human experience enables organizations to develop AI systems that feel more intuitive, transparent, and emotionally aligned with user expectations, ultimately driving higher satisfaction, engagement, and long-term loyalty.

5.3. Implications for Policymakers and Regulatory Bodies

Policymakers can utilize the Human-Centered AI Integration Framework as a guiding structure for establishing AI governance standards that emphasize fairness, transparency, explainability, and user rights protection. By aligning regulatory frameworks with responsible AI principles, governments and institutions can ensure that AI technologies deployed in public and private sectors remain secure, trustworthy, and socially responsible. The framework also provides a foundation for auditors and oversight bodies to evaluate ethical compliance, algorithmic risks, and inclusivity levels within emerging AI ecosystems.

6. CONCLUSION

This study introduced the Human-Centered AI Integration Framework, a multidimensional model that unifies ethical, psychological, and computational perspectives to guide the development of responsible and human-centered artificial intelligence. The framework emphasizes that AI innovation must extend beyond technical efficiency by integrating human values, emotional understanding, and social responsibility. Evaluation results show that ethical alignment fosters trust and fairness, psychological awareness strengthens user acceptance, and computational practices such as bias mitigation and explainability operationalize these principles. Furthermore, the framework's orientation toward fairness, inclusivity, and transparent governance aligns closely with SDG 9, SDG 10, and SDG 16, reinforcing global efforts to build sustainable, equitable, and trustworthy digital ecosystems.

Overall, the framework contributes to the broader discourse on responsible AI by promoting cross-disciplinary collaboration and guiding policymakers, developers, and researchers in designing transparent and inclusive AI systems. It addresses ongoing IEEE priorities by supporting ethical engineering practices, enhancing system accountability, and strengthening algorithmic governance. This interdisciplinary approach enables the creation of AI technologies that uphold human well-being and reduce bias, providing practical direction for auditors and engineers seeking to embed fairness, trust, and explainability into AI solutions.

The Human-Centered AI Integration Framework also offers actionable pathways for future development, including the integration of fairness checkpoints, empathy-based interaction metrics, and standardized explainability testing within engineering workflows. Future research is encouraged to validate the framework across diverse computational environments and develop open-source tools that translate ethical and psychological principles into measurable technical components. These directions reinforce the framework's applied value and support global and IEEE missions to advance secure, transparent, and human-centered intelligent systems.

7. DECLARATIONS

7.1. About Authors

Lina Nurjanah (LN) https://orcid.org/0009-0009-9402-2983

Roby Syaiful Ubed (RS) https://orcid.org/00000-0002-5745-7015

Prabawati Nurhabibah (PN) https://orcid.org/0000-0002-2639-7961

Marta Rodriguez (MR) https://orcid.org/0009-0000-1367-0511

7.2. Author Contributions

Conceptualization: LN and RS; Methodology: PN; Software: LN and MR; Validation: RS and PN; Formal Analysis: MR and PN; Investigation: LN; Resources: RS; Data Curation: PN; Writing Original Draft Preparation: RS and MR; Writing Review and Editing: LN, PN, and MR; Visualization: LN. All authors, LN, RS, PN, and MR, have read and agreed to the published version of the manuscript.

REFERENCES

[1] A. Saxena, G. Fletcher, and M. Pechenizkiy, "Fairsna: Algorithmic fairness in social network analysis," *ACM Computing Surveys*, vol. 56, no. 8, pp. 1–45, 2024.

[2] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, and A. Hussain, "Interpreting black-box models: a review on explainable artificial intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.

[3] M. O. Syaidina, R. Fahrudin, and I. A. Mutiara, "Implementation of ethics of using artificial intelligence in the education system in indonesia," *Blockchain Frontier Technology*, vol. 4, no. 1, pp. 63–71, 2024.

[4] Z. Chen, J. M. Zhang, M. Hort, M. Harman, and F. Sarro, "Fairness testing: A comprehensive survey and analysis of trends," *ACM Transactions on Software Engineering and Methodology*, vol. 33, no. 5, pp. 1–59, 2024.

[5] T. Mariyanti, I. Wijaya, C. Lukita, S. Setiawan, and E. Fletcher, "Ethical framework for artificial intelligence and urban sustainability," *Blockchain Frontier Technology*, vol. 4, no. 2, pp. 98–108, 2025.

[6] O. Parraga, M. D. More, C. M. Oliveira, N. S. Gavenski, L. S. Kupssinskü, A. Medronha, L. V. Moura, G. S. Simões, and R. C. Barros, "Fairness in deep learning: A survey on vision and language research," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–40, 2025.

[7] A. Fabris, N. Baranowska, M. J. Dennis, D. Graus, P. Hacker, J. Saldivar, F. Zuiderveen Borgesius, and A. J. Biega, "Fairness and bias in algorithmic hiring: A multidisciplinary survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 1, pp. 1–54, 2025.

[8] A. Erica, S. Wulandari, and R. Widayanti, "Data security transformation: The significant role of blockchain technology," *Blockchain Frontier Technology*, vol. 3, no. 2, pp. 107–112, 2024.

[9] E. Soremekun, M. Papadakis, M. Cordy, and Y. Le Traon, "Software fairness: An analysis and survey," *ACM Computing Surveys*, vol. 58, no. 3, pp. 1–38, 2025.

[10] T. P. Pagano, R. B. Loureiro, F. V. Lisboa, R. M. Peixoto, G. A. Guimarães, G. O. Cruz, M. M. Araujo, L. L. Santos, M. A. Cruz, E. L. Oliveira *et al.*, "Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods," *Big data and cognitive computing*, vol. 7, no. 1, p. 15, 2023.

[11] C. Yu, G. Yao *et al.*, "Enhancing student engagement with ai-driven personalized learning systems," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 1–8, 2024.

[12] S. Das, R. Stanton, and N. Wallace, "Algorithmic fairness," *Annual Review of Financial Economics*, vol. 15, no. 1, pp. 565–593, 2023.

[13] Badan Riset dan Inovasi Nasional (BRIN), "Ai and blockchain as keys to national digital transformation," Online news release, Sep. 2025, highlights AI (and blockchain) as essential for Indonesia's digital transformation and national development. [Online]. Available: https://www.brin.go.id/en/news/124783/brin-ai-and-blockchain-as-a-keys-to-national-digital-transformation

[14] D. Pessach and E. Shmueli, "A review on fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1–44, 2022.

[15] D. Martinez, L. Magdalena, and A. N. Savitri, "Ai and blockchain integration: Enhancing security and transparency in financial transactions," *International Transactions on Artificial Intelligence*, vol. 3, no. 1, pp. 11–20, 2024.

[16] M. Hort, Z. Chen, J. M. Zhang, M. Harman, and F. Sarro, "Bias mitigation for machine learning classifiers: A comprehensive survey," *ACM Journal on Responsible Computing*, vol. 1, no. 2, pp. 1–52, 2024.

[17] S. Dehdashtian, R. He, Y. Li, G. Balakrishnan, N. Vasconcelos, V. Ordonez, and V. N. Boddeti, "Fairness and bias mitigation in computer vision: A survey," *arXiv preprint arXiv:2408.02464*, 2024.

[18] G. Silva, G. Godwin, and O. Jayanagara, "The impact of ai on personalized learning and educational analytics," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 36–46, 2024.

[19] M. Zallio, C. B. Ike, and C. Chivăran, "Designing artificial intelligence: Exploring inclusion, diversity, equity, accessibility, and safety in human-centric emerging technologies," *AI*, vol. 6, no. 7, p. 143, 2025.

[20] M. F. Almufareh, S. Kausar, M. Humayun, and S. Tehsin, "A conceptual model for inclusive technology: advancing disability inclusion through artificial intelligence," *Journal of Disability Research*, vol. 3, no. 1, p. 20230060, 2024.

[21] A. Sutarman, J. Williams, D. Wilson, and F. B. Ismail, "A model-driven approach to developing scalable educational software for adaptive learning environments," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 9–16, 2024.

[22] A. Tlili, M. Denden, M. Abed, and R. Huang, "Artificial intelligence ethics in services: are we paying attention to that?!" *The Service Industries Journal*, vol. 44, no. 15-16, pp. 1093–1116, 2024.

[23] E. G. Hernández, "Towards an ethical and inclusive implementation of artificial intelligence in organizations: a multidimensional framework," *arXiv preprint arXiv:2405.01697*, 2024.

[24] N. Anwar, J. Anderson, T. Williams *et al.*, "Applying data science to analyze and improve student learning outcomes in educational environments," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 72–83, 2024.

[25] E. Fosch-Villaronga and A. Poulsen, "Diversity and inclusion in artificial intelligence," *Law and artificial intelligence: Regulating AI and applying AI in legal practice*, pp. 109–134, 2022.

[26] N. Lutfiani, N. P. L. Santoso, R. Ahsanitaqwim, U. Rahardja, and A. R. A. Zahra, "Ai-based strategies to improve resource efficiency in urban infrastructure," *International Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 121–127, 2024.

[27] G. Cachat-Rosset and A. Klarsfeld, "Diversity, equity, and inclusion in artificial intelligence: an evaluation of guidelines," *Applied Artificial Intelligence*, vol. 37, no. 1, p. 2176618, 2023.

[28] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala, "Transparency and explainability of ai systems: From ethical guidelines to requirements," *Information and Software Technology*, vol. 159, p. 107197, 2023.

[29] A. Fernanda, M. Huda, and A. R. F. Geovanni, "Application of learning cloud computing technology (cloud computing) to students in higher education," *International Journal of Cyber and IT Service Management*, vol. 3, no. 1, pp. 32–39, 2023.

[30] R. R. Hoffman, T. Miller, and W. J. Clancey, "Psychology and ai at a crossroads: How might complex systems explain themselves?" *The American journal of psychology*, vol. 135, no. 4, pp. 365–378, 2022.

[31] L. K. Choi, K. B. Rii, and H. W. Park, "K-means and j48 algorithms to categorize student research abstracts," *International Journal of Cyber and IT Service Management*, vol. 3, no. 1, pp. 61–64, 2023.

[32] K. F. Ystgaard and K. De Moor, "Envisioning the future: A multi-disciplinary approach to human-centered intelligent environments," *Quality and User Experience*, vol. 8, no. 1, p. 11, 2023.

[33] J. Friedrich, A. Brückner, J. Mayan, S. Schumann, A. Kirschenbaum, and C. Zinke-Wehlmann, "Human-centered ai development in practice—insights from a multidisciplinary approach," *Zeitschrift Für Arbeitswissenschaft*, vol. 78, no. 3, pp. 359–376, 2024.

[34] W. Xu, M. J. Dainoff, L. Ge, and Z. Gao, "Transitioning to human interaction with ai systems: New challenges and opportunities for hci professionals to enable human-centered ai," *International Journal of Human–Computer Interaction*, vol. 39, no. 3, pp. 494–518, 2023.

[35] T. Li, M. Vorvoreanu, D. DeBellis, and S. Amershi, "Assessing human-ai interaction early through factorial surveys: a study on the guidelines for human-ai interaction," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 5, pp. 1–45, 2023.

[36] Badan Kebijakan Kesehatan, Kementerian Kesehatan RI, "The use of artificial intelligence in the duties of public–humas pranata," Online article, Aug. 2024, describes AI adoption in government to improve operational efficiency and public services. [Online]. Available: https://www.badankebijakan.kemkes.go.id/en/penggunaan-artificial-intelligence-dalam-tugas-pranata-humas-bkpk-kemenkes/

[37] R. T. Utami, V. H. Lubis, N. P. L. Santoso, A. Rizky, M. A. Komara, L. Suryadi *et al.*, "Blockchain technology for securing electronic health records and enhancing data privacy," in *2025 4th International Conference on Creative Communication and Innovative Technology (ICCIT)*. IEEE, 2025, pp. 1–7.

[38] K. Adel, A. Elhakeem, and M. Marzouk, "Blockchain and artificial intelligence: scientometric analysis and visualization," *IEEE Access*, vol. 11, pp. 137 911–137 928, 2023.

[39] M. Alaeddini, M. Hajizadeh, and P. Reaidy, "A bibliometric analysis of research on the convergence of artificial intelligence and blockchain in smart cities," *Smart Cities*, vol. 6, no. 2, pp. 764–795, 2023.

[40] M. S. Al Jasem, T. De Clark, and A. K. Shrestha, "Toward decentralized intelligence: A systematic literature review of blockchain-enabled ai systems," *Information*, vol. 16, no. 9, p. 765, 2025.

[41] H. Taherdoost, "Blockchain technology and artificial intelligence together: a critical review on applications," *Applied Sciences*, vol. 12, no. 24, p. 12948, 2022.

[42] U. Eswaran, V. Eswaran, K. Murali, and V. Eswaran, "Human-centric ai balancing innovation with ethical considerations in the age of soft computing," in *Soft Computing in Industry 5.0 for Sustainability*. Springer, 2024, pp. 87–116.

[43] A. Mazarakis, C. Bernhard-Skala, M. Braun, and I. Peters, "What is critical for human-centered ai at work?–toward an interdisciplinary theory," *Frontiers in artificial intelligence*, vol. 6, p. 1257057, 2023.

[44] S. T. H. Mortaji and M. E. Sadeghi, "Assessing the reliability of artificial intelligence systems: Challenges, metrics, and future directions," *International Journal of Innovation in Management, Economics and Social Sciences*, vol. 4, no. 2, pp. 1–13, 2024.

[45] Q. Aini, P. Purwanti, R. N. Muti, E. Fletcher *et al.*, "Developing sustainable technology through ethical ai governance models in business environments," *ADI Journal on Recent Innovation*, vol. 6, no. 2, pp. 145–156, 2025.

[46] A. Hamza, G. Shi, and B. Hossain, "Migration as an adaptation measure to achieve resilient lifestyle in the face of climate-induced drought: Insight from the thar desert in pakistan," *Water*, vol. 16, no. 18, p. 2692, 2024.

[47] M. A. Camilleri, "Artificial intelligence governance: Ethical considerations and implications for social responsibility," *Expert systems*, vol. 41, no. 7, p. e13406, 2024.

[48] M. Jeon, "The effects of emotions on trust in human-computer interaction: A survey and prospect," *International Journal of Human–Computer Interaction*, vol. 40, no. 22, pp. 6864–6882, 2024.

[49] A. Chinnaraju, "Explainable ai (xai) for trustworthy and transparent decision-making: A theoretical framework for ai interpretability," *World Journal of Advanced Engineering Technology and Sciences*, vol. 14, no. 3, pp. 170–207, 2025.

[50] C. Troussas, A. Krouska, and C. Sgouropoulou, "A novel framework of human–computer interaction and human-centered artificial intelligence in learning technology," in *Human-Computer Interaction and Augmented Intelligence: The Paradigm of Interactive Machine Learning in Educational Software*. Springer, 2025, pp. 387–431.

[51] D. Ogbu, "Agentic ai in computer vision domain-recent advances and prospects," *International Journal of Research Publication and Reviews*, vol. 4, no. 12, pp. 5102–5120, 2023.

[52] J. Wang, Y. Huo, J. Mahe, Z. Ge, Z. Liu, W. Wang, and L. Zhang, "Developing an ethical regulatory framework for artificial intelligence: integrating systematic review, thematic analysis, and multidisciplinary theories," *IEEE Access*, 2024.

[53] A. Ellikkal and S. Rajamohan, "Ai-enabled personalized learning: empowering management students for improving engagement and academic performance," *Vilakshan-XIMB Journal of Management*, vol. 22, no. 1, pp. 28–44, 2025.

[54] L. Meria, C. S. Bangun, and J. Edwards, "Exploring sustainable strategies for education through the adoption of digital circular economy principles," *International Transactions on Education Technology (ITEE)*, vol. 3, no. 1, pp. 62–71, 2024.

[55] C. Lukita, S. Purnama, A. Rizky, M. F. Fazri *et al.*, "Analysis of gamification and blockchain integration in intelligent learning systems," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*. IEEE, 2024, pp. 1–6.