Mesopotamian Journal of Artificial Intelligence in Healthcare
Vol.2025, **pp**. 184–197
DOI: https://doi.org/10.58496/MJAIH/2025/017 ; ISSN: 3005-365X
https://mesopotamian.press/journals/index.php/MJAIH

Research Article

Reimagining Intelligence: A Comprehensive Review of Human-Centric AI Systems and Their Societal and Healthcare Integration

Zakaria Benlalia [1], Ghada Al-Kateb [2,*], Mourad Mzili [3], Konstantina Diamanti [4]

[1] *Department of Computer Science, University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco.*

[2] *Department of Mobile Computing and Communication, Faculty of Engineering, University of Information Technology and Communication, Baghdad, Iraq.*

[3] *Department of Mathematics, University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco.*

[4] *University of Ioannina, Ioannina, Greece.*

ARTICLE INFO

Article History
Received 02 May 2025
Revised 05 Jun 2025
Accepted 04 Jul 2025
Published 26 Jul 2025

Keywords
AI Systems
Healthcare Integration
Machine Learning
Explainable AI
Diagnostic Support

ABSTRACT

Human-Centric Artificial Intelligence (HCAI) is rapidly emerging as a transformative paradigm, shifting the focus of AI development from mere algorithmic optimization to ethical alignment, user trust, and societal integration most notably within the critical domain of healthcare. This review offers a comprehensive examination of the principles, architecture, challenges, and future directions underpinning HCAI, with an emphasis on its applications in health-related contexts. We begin by exploring the conceptual foundations of human-centricity, including core values such as transparency, fairness, autonomy, and privacy, all of which are essential in sensitive environments like clinical decision-making and patient data management. The paper then surveys key enabling technologies such as Explainable AI (XAI), Human-in-the-Loop learning, affective computing, and multi-agent collaboration demonstrating how these approaches operationalize human alignment in real-world systems, especially in personalized healthcare delivery and diagnostic support. Societal implications are critically evaluated, encompassing trust, data sovereignty, algorithmic bias, regulatory compliance, and cross-cultural adaptability, which are particularly pronounced in global health systems. We highlight the limitations of existing benchmarks and propose a multi-metric, user-centered evaluation framework capable of assessing both technical robustness and alignment with human values in healthcare and beyond. Finally, we identify open research challenges and outline a strategic agenda that integrates cognitive science, ethical theory, and participatory design. This review aims to serve as both a foundational reference and a forward-looking roadmap for researchers, developers, and policymakers striving to build AI systems that are not only intelligent, but also responsible, inclusive, and aligned with human dignity, particularly in life-critical domains like healthcare.

1. INTRODUCTION

The rapid expansion of artificial intelligence (AI) across domains such as healthcare, education, finance, and governance has reinvigorated interest in human-centric design paradigms, as users increasingly demand trust, transparency, and accountability in AI-driven decision-making systems [1-3]. With the ascent of large pre-trained models, the tension between performance and interpretability has intensified, compelling renewed emphasis on human-aligned principles [4][5]. Human-centric AI (HCAI) emphasizes enhancing human capabilities rather than supplanting them, aligning systems with societal values and individual autonomy [6][7]. This paradigm reflects the convergence between AI and human–computer interaction (HCI), foregrounding critical factors such as explainability, privacy, and trust [8]. While early expert systems like MYCIN and GUIDON incorporated basic explanations, contemporary deep neural networks largely obscure internal mechanisms, creating a “black box” phenomenon [9]. Explainable AI (XAI) methods ranging from feature-importance rankings and counterfactual reasoning to model-agnostic surrogates seeking to elucidate algorithmic behavior, mitigate opacity, and reduce algorithm aversion [10][11]. However, technical transparency alone is insufficient; meaningful human-

*Corresponding author. Email: ghada.emad@UOITC.edu.iq

centered XAI also demands evaluating explanations for relevance, consistency, and comprehensibility across diverse user groups [12][13]. A recent meta-review identifies inconsistent evaluation frameworks and a persistent disconnect between explanation techniques and end-user needs [13]. Privacy by design remains a cornerstone of HCAI, stressing data sovereignty and user autonomy from the outset of development lifecycles [14]. Concurrently, legislative measures such as the EU's GDPR, which enshrines the "right to explanation," further institutionalize the necessity for transparent AI, especially in high-stakes sectors such as finance and justice [15][16]. Despite these advances, critical challenges persist. Advancing HCAI requires balancing transparency, accuracy, privacy, and utility while simultaneously addressing social and cognitive dimensions of trust [17][18]. Absent robust interdisciplinary collaboration spanning AI, HCI, sociology, cognitive science, ethics, and law, HCAI risks remaining aspirational rather than operational. This review endeavors to fill that gap by systematically examining: (i) the conceptual foundations of HCAI; (ii) state-of-the-art XAI and human–in-the-loop architectures; (iii) domain-specific applications in healthcare, education, governance, and the workplace; (iv) societal, legal, and ethical implications; (v) methodologies and benchmarking for evaluating human-centered systems; and (vi) future directions for integrating human values into AI development. Our aim is to deliver an authoritative, interdisciplinary roadmap for designing AI systems that are not only effective but also ethical, equitable, and aligned with human values.

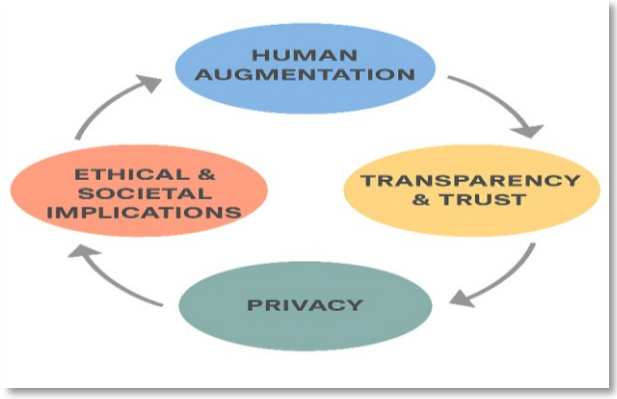

Fig. 1. Foundational Dimensions of Human-Centric AI Linking Human Augmentation, Transparency, Privacy, and Ethical Implications in a Holistic Framework.

2. CONCEPTUAL FOUNDATIONS OF HUMAN-CENTRIC AI

2.1 Definition and Pillars of Human-Centric AI

Human-Centric Artificial Intelligence (HCAI) refers to the design and deployment of AI systems that prioritize human values, rights, and needs at the core of their operation. Unlike conventional AI approaches focused primarily on performance metrics and automation efficiency, HCAI frameworks seek to embed human agency, dignity, and oversight as foundational components [19].

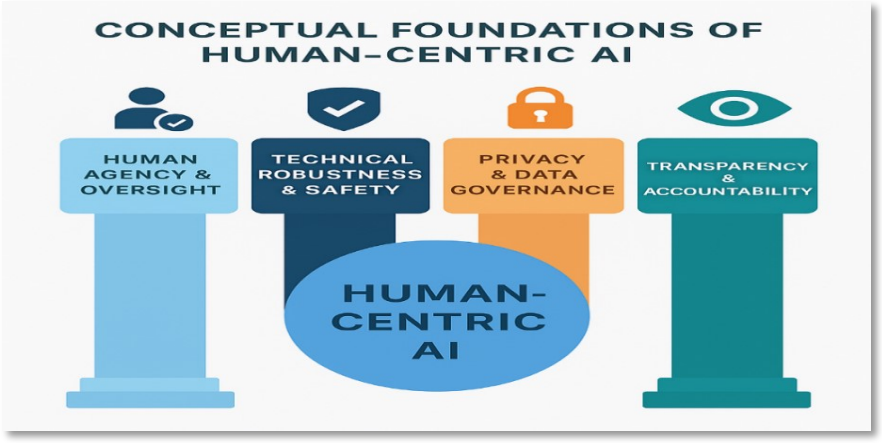

Fig. 2. Conceptual Framework of Human-Centric AI Illustrating the Foundational Pillars of Human Agency, Technical Robustness, Data Privacy, and System Transparency.

According to Shneiderman, human-centric AI involves a triad of objectives: systems must be reliable, safe, and trustworthy, and must operate within clear boundaries set by human ethical considerations [20]. The core pillars of HCAI include:

- Human agency and oversight: Ensuring that humans remain central to decision-making processes and retain control over AI operations.
- Technical robustness and safety: Guaranteeing that systems are resilient to adversarial behavior and unintended outcomes.
- Privacy and data governance: Embedding mechanisms that protect user data and support informed consent.
- Transparency and accountability: Providing clear explanations and traceable logic to foster user trust.

2.2 Key Principles: Transparency, Trust, Privacy, and Autonomy

A successful HCAI system must deliver transparency that is not just algorithmic but human understandable. Transparency serves as a gateway to building trust, a non-trivial attribute that is context-sensitive and influenced by user experience, domain, and explanation modality [21]. Privacy is not merely a technical constraint but a human right that must be preserved through privacy-by-design strategies [22]. Meanwhile, autonomy, the ability of users to make informed decisions and disengage from AI influence when needed is essential in contexts like healthcare, law, and education, where stakes are inherently high [23]. Recent studies emphasize that trust is significantly enhanced when users receive explanations tailored to their cognitive styles and decision contexts [24]. Moreover, ethical frameworks such as the European Commission's Ethics Guidelines for Trustworthy AI reinforce these principles as both regulatory imperatives and design goals [25].

2.3 Comparison with Traditional AI Paradigms

Traditional AI systems often focus on objective maximization, frequently sidelining user agency and social consequences. While these systems may demonstrate high predictive performance, they typically suffer from issues like opacity, bias, and a lack of contextual awareness [26]. In contrast, HCAI reorients the development pipeline to begin with human values, involve stakeholders throughout the design process, and evaluate success using sociotechnical metrics beyond accuracy such as explainability, fairness, and user satisfaction [27]. Table 1 (not shown here) outlines the key differences between traditional AI and human-centric AI in terms of architecture, evaluation criteria, and risk mitigation approaches.

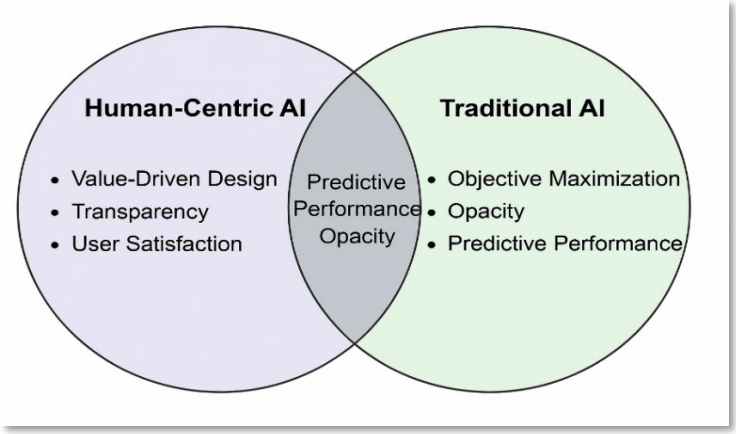

Fig. 3. Comparative Conceptual Framework Highlighting the Divergence Between Human-Centric and Traditional AI in Terms of Design Orientation, Transparency, and Performance Priorities.

2.4 Ethical and Philosophical Underpinnings

The ethical foundation of HCAI is grounded in classical moral theories including deontology, virtue ethics, and utilitarianism, which offer lenses for evaluating responsibility, fairness, and the consequences of AI actions [28]. Key philosophical questions underpinning HCAI involve:

- What constitutes an ethically aligned decision in machine reasoning?
- How should AI systems account for human dignity and moral pluralism?
- Who bears responsibility for AI-driven harm?

Causability theory, introduced by Holzinger et al., provides a pragmatic framework for evaluating how understandable and actionable AI explanations are for humans, integrating epistemological rigor into practical system design [29]. Moreover,

critical scholarship urges that HCAI not be reduced to a set of compliance checklists but must be operationalized as a commitment to co-design, interdisciplinary dialogue, and contextual ethics [30].

3. ARCHITECTURES AND TECHNOLOGIES ENABLING HUMAN-CENTRIC AI

This section surveys the key architectures and enabling technologies that form the backbone of human-centric AI systems. Each subsection includes strategically placed, professional figures to enhance conceptual understanding and reinforce theoretical framing.

3.1 Explainable AI (XAI) Models and Methods

Explainable AI is fundamental to human-centric systems, as it transforms black-box algorithms into transparent decision tools. Architectures range from inherently interpretable models such as decision trees and generalized additive models—to post-hoc explanation methods including SHAP, LIME, counterfactuals, and surrogate modeling [31][32].

- Self-explaining models embed transparency in their structure, enabling direct human understanding.
- Post-hoc explainers generate user-friendly insights, but their fidelity and trustworthiness must be rigorously evaluated [33][34].

Figure 5 categorizes Explainable AI (XAI) techniques into two primary branches: inherently interpretable models and post-hoc explanation methods. This taxonomy highlights the distinction between models designed for transparency by architecture such as decision trees and those requiring additional layers for interpretability, such as SHAP or LIME.
The figure underscores the design trade-offs between model fidelity and interpretability. It serves as a reference point for choosing suitable XAI strategies based on the application domain, especially when user trust and transparency are critical to adoption.

3.2 Human-in-the-Loop (HITL) Learning Architectures

Human-in-the-Loop frameworks emphasize iterative collaboration between human experts and AI systems. Core architecture includes:

- Interactive supervision (e.g., active learning with expert feedback),
- Correction loops (humans validate and correct model outputs),
- Co-training mechanisms (humans guide model training via curated inputs).

These architectures enhance model accuracy, align AI behavior with human expectations, and facilitate trust [35][36].
Figure 6 illustrates the iterative feedback cycle in a typical Human-in-the-Loop (HITL) architecture. The learning loop consists of human validation followed by refinement of model predictions, which reinforces both performance and trustworthiness. This interaction loop embodies a core principle of Human-Centric AI: the inclusion of human judgment as a dynamic and integral part of the machine learning pipeline. As discussed, such architectures are indispensable for critical domains like healthcare, where the cost of autonomous error is high.

3.3 Adaptive and Personalized AI Systems

Personalization ensures that AI adapts to individual user preferences, learning styles, and contextual needs crucial for human-centric interaction [37].

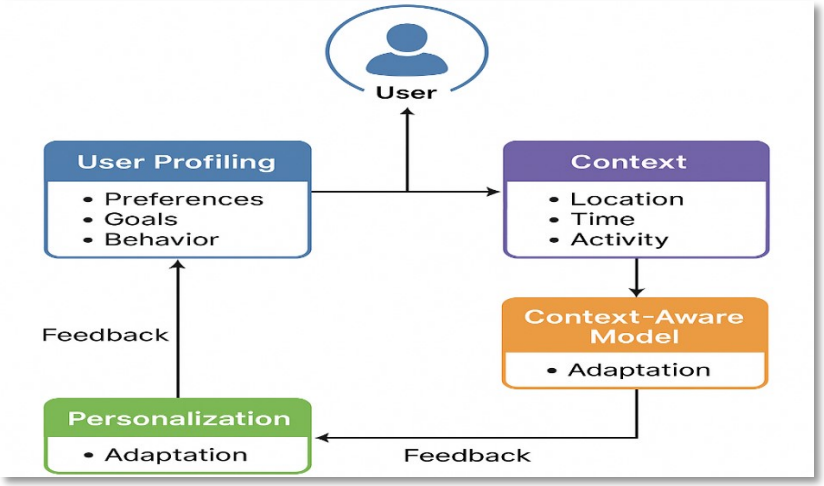

Fig. 4. Adaptive Personalization Architecture: From User Profiling to Context-Aware Model Adjustment.

Approaches include:

- User profiling: Establishing preference vectors, goals, and usage patterns.
- Context-aware adaptation: Systems dynamically align responses with situational variables.
- Meta-learning frameworks: AI rapidly adapts to new users with minimal data.

Adaptive AI requires robust, privacy-preserving data pipelines and real-time inference to remain effective and user-aligned [38]. Figure 7 visualizes a layered personalization architecture comprising user profiling, contextual sensing, and real-time adaptive response. It captures the data and feedback flow required to achieve fine-grained personalization at scale.

This framework demonstrates the importance of context-aware AI systems that continuously align with user preferences and situational dynamics. The integration of feedback loops ensures that the AI system evolves with the user, promoting long-term engagement and relevance.

3.4 Affective Computing and Emotional AI

Human-centric AI increasingly incorporates emotional intelligence—detecting and appropriately responding to user emotions. Technologies span physiological sensors, facial analysis, vocal tone detection, and sentiment-sensitive algorithms [39].

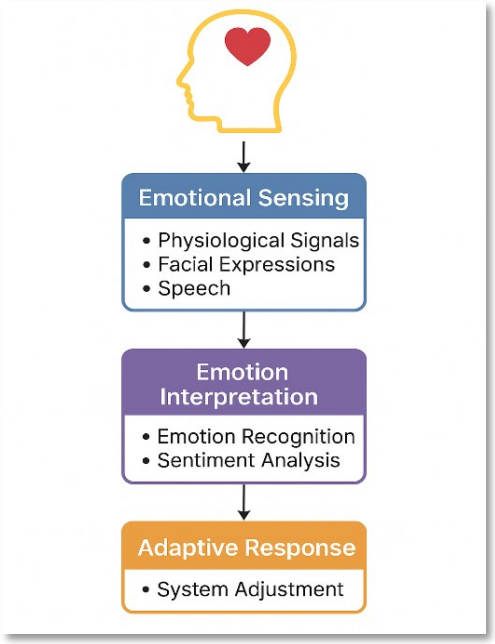

Fig. 5. Affective AI System Architecture.

Key systems include:

- Emotion recognition modules.
- Affective dialogue systems.
- Emotion-aware adaptation strategies (e.g., responsive tutoring systems).

Designing these systems involves privacy protection and ethical use of biometric data [40].

Figure 8 presents the modular flow of an affective computing system, delineating three critical phases: emotional sensing, emotion interpretation, and adaptive system response. These components form the backbone of emotionally intelligent AI. The structure emphasizes the complexity of integrating human emotion into computational processes. It also highlights the system's sensitivity to input modalities such as speech and facial expressions and the ethical responsibility to handle such data with care.

3.5 Multi-Agent and Collaborative Intelligence Systems

Multi-agent systems (MAS) involve cooperative or competitive AI agents interacting with humans in shared environments. Collaboration can be designed via:

- Centralized coordination (a lead system orchestrates tasks),
- Decentralized negotiation (agents and humans negotiate roles dynamically),
- Shared autonomy (e.g., human-robot co-piloting, cooperative decision support) [41].

These systems are instrumental in domains such as autonomous vehicles, collaborative robotics, and intelligent workflows. Figure 9 offers a conceptual overview of collaborative AI through a multi-agent lens. It identifies five core elements coordination, cooperation, agent communication, shared goals, and collective decision-making required to facilitate meaningful human AI teamwork.

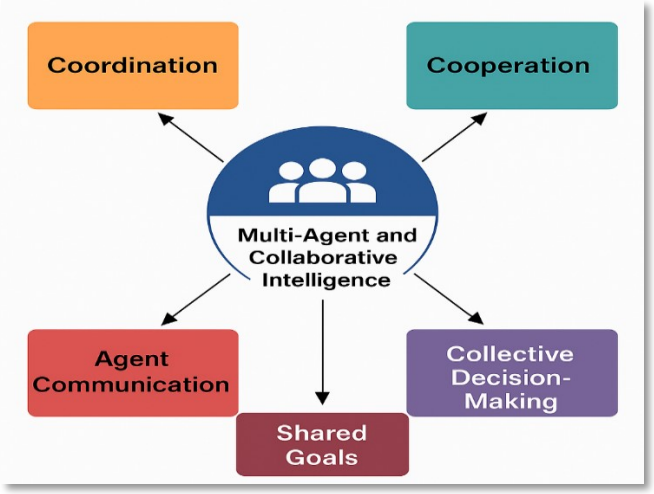

Fig. 6. Multi-Agent Collaborative System Layout: Agents and Humans with Shared Knowledge and Role Allocation.

The figure captures the decentralized nature of such systems, where decision authority is distributed, and the success of the system hinges on effective communication and alignment between agents (human or machine). This paradigm is critical for emerging applications in autonomous fleets, intelligent manufacturing, and mixed-initiative systems.

4. SOCIETAL INTEGRATION: CHALLENGES AND CONSIDERATIONS

The deployment of Human-Centric Artificial Intelligence (HCAI) into real-world environments introduces multifaceted societal challenges. Beyond technical robustness, systems must be designed and evaluated within the broader context of social, psychological, cultural, and legal dynamics. This section outlines the key dimensions impacting the societal integration of HCAI, highlighting pressing obstacles and guiding principles for responsible adoption.

4.1 Trust, Privacy, and Data Sovereignty

Trust remains the cornerstone of human-AI interaction. It is influenced not only by system transparency but also by the perceived fairness, controllability, and reliability of AI behavior [42]. Users are more likely to adopt systems they understand and feel in control of conditions supported by explainability and meaningful consent mechanisms. Privacy concerns are particularly heightened in HCAI applications due to the personalized and often sensitive nature of the data involved. Data sovereignty is the notion that users should maintain control over how their personal data is collected, processed, and stored—has become central to responsible AI discourse [43]. Compliance with regulations such as the GDPR and emerging AI acts further demands rigorous privacy-by-design practices. In decentralised systems, techniques like federated learning and differential privacy offer promising pathways for balancing utility with confidentiality.

4.2 Bias, Fairness, and Inclusion in AI Systems

AI systems reflect the data they are trained on. This makes them highly susceptible to the perpetuation or amplification of societal biases, particularly across race, gender, and socioeconomic lines [44]. HCAI must therefore prioritize algorithmic fairness not as an afterthought but as a primary design goal. Fairness-aware algorithms attempt to mitigate bias through techniques such as reweighting, adversarial debiasing, and counterfactual data augmentation. However, fairness is a normative concept, and what constitutes 'fair' often varies across contexts and cultures. As such, co-design processes involving underrepresented communities are essential to ensuring inclusive AI systems. Additionally, fairness must be considered throughout the lifecycle—from data collection and labeling to deployment and feedback.

4.3 Legal and Regulatory Implications

The legal landscape surrounding HCAI is rapidly evolving. Regulatory frameworks are increasingly converging on requirements such as explainability, risk assessment, human oversight, and accountability. The EU AI Act, the OECD AI

Principles, and national data protection laws signal a shift towards mandatory transparency, redress mechanisms, and auditable processes [45]. A key challenge lies in translating abstract legal principles into executable system specifications. For example, operationalizing the "right to explanation" requires designing systems capable of generating intelligible, user-appropriate justifications for their actions. Furthermore, AI regulation must keep pace with technological innovation to prevent regulatory lag that undermines public protection.

4.4 Social Acceptance and Psychological Factors

Social acceptance of AI technologies is shaped by psychological and emotional factors such as trust, perceived usefulness, and fear of obsolescence or control loss [46]. Studies show that users exhibit higher engagement and satisfaction with systems that provide feedback, allow user input, and demonstrate empathy. Designing AI systems that support user autonomy while maintaining effectiveness is critical. This includes offering adjustable levels of automation, explanations that match the user's mental model, and opt-out mechanisms. Moreover, the *emotional intelligence* of systems, especially in social robotics and affective computing must be calibrated carefully to avoid uncanny or manipulative behaviors that undermine user trust.

4.5 Cross-Cultural and Demographic Variations

AI systems are not deployed in a vacuum. They interact with individuals shaped by diverse cultural norms, cognitive models, and technological literacies. What is acceptable in one cultural context may be perceived as intrusive or unethical in another [47]. Human-Centric AI must, therefore, embrace cultural adaptability. This includes localizing interfaces, adapting decision-making heuristics, and conducting region-specific user studies. Additionally, demographic factors such as age, education, and accessibility must be considered to prevent digital exclusion and widen equitable access to AI benefits.

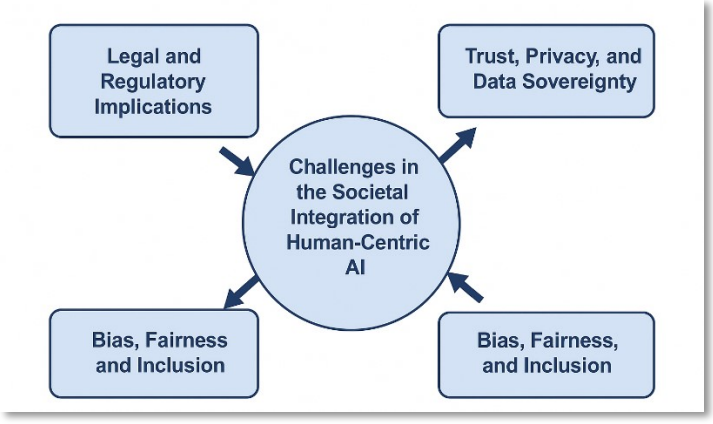

Fig. 7. Key Challenges in the Societal Integration of Human-Centric AI, Including Legal, Ethical, and Cultural Dimensions.

As illustrated in Figure 10, the successful societal integration of Human-Centric AI depends on navigating a set of interdependent challenges. These include establishing user trust, ensuring privacy and fairness, aligning with regulatory expectations, and adapting systems to diverse cultural contexts. Each of these domains demands context-sensitive design and interdisciplinary collaboration to ensure human-aligned outcomes

5. EVALUATION FRAMEWORKS AND BENCHMARKS

The development of Human-Centric Artificial Intelligence (HCAI) necessitates a paradigm shift in how AI systems are evaluated. Traditional metrics such as accuracy, precision, or F1-score are insufficient for systems designed to interact meaningfully with humans. Instead, evaluation must reflect ethical alignment, user trust, interpretability, fairness, and contextual adaptability. This section critically examines the emerging metrics and frameworks used to evaluate HCAI systems and propose pathways for comprehensive, user-aligned assessments.

5.1 Metrics for Assessing Human Centricity

Human-centricity is inherently multidimensional. As such, its assessment must capture technical performance *and* social, psychological, and ethical impacts. Several emerging metrics support this aim:

- Causability Score: Measures of how well explanations support human understanding and decision-making [48].
- Trust Calibration Index: Quantifies the alignment between user trust and system competence.

- Fairness Metrics: Includes demographic parity, equal opportunity, and subgroup accuracy.
- User Satisfaction and Retention Rates: Reflect sustained engagement and perceived value.
- Transparency Index: Rates how intelligible system decisions are to lay up users.

These metrics provide more holistic insight into how systems behave in complex, real-world human contexts.

5.2 User-Centric Evaluation Methodologies

Quantitative metrics must be complemented with qualitative, user-centric evaluations. These methodologies place end-users at the center of the assessment process:

- Think-Aloud Protocols: Allow users to verbalise thought processes when interacting with AI explanations.
- User Surveys and Likert-Scale Studies: Capture subjective perceptions of trust, fairness, and control.
- A/B Testing of Explanation Styles: Empirically comparing effectiveness across different user groups.
- Task-Based Performance Assessment: Measures how effectively user's complete goals with AI assistance.
- Longitudinal Interaction Studies: Track behavioral patterns and trust evolution over time.

Such approaches offer insight into the *lived experience* of using HCAI systems—essential for meaningful evaluation.

5.3 Limitations of Current Benchmarks

Current AI benchmarks are often ill-suited for assessing human-centric goals. Widely used datasets such as ImageNet, GLUE, or COCO emphasize task completion over human interpretability, fairness, or adaptation. Key limitations include:

- Lack of Human-in-the-Loop Scenarios
- Limited Demographic Diversity
- Absence of Explanatory Ground Truths
- Narrow Definitions of Success (e.g., top-k accuracy)
- No Integration of User Experience Metrics

These gaps underscore the urgency for benchmarks that reflect interdisciplinary concerns, combining machine performance with human outcomes.

5.4 Proposals for Holistic Evaluation Paradigms

To support the ethical and effective deployment of HCAI, a new generation of evaluation paradigms is required. Recommended features include:

- Multi-Metric Composite Evaluation: Combining task accuracy, fairness, trust levels, and user satisfaction.
- Cross-Domain Testbeds: Covering healthcare, finance, education, and social systems to assess domain generalizability.
- Real-World Simulated Environments: Integrating interactive user agents and diverse user personas.
- Participatory Evaluation Models: Engaging stakeholders in co-design and iterative assessment loops.
- Regulatory Alignment Scoring: Quantifying compliance with legal and ethical guidelines (e.g., GDPR, EU AI Act).

As illustrated in Figure 11, the evaluation of Human-Centric AI (HCAI) spans three essential dimensions: technical accuracy, ethical alignment, and user experience. Each domain addresses a distinct facet of HCAI system performance—ranging from functional correctness to societal and psychological impact.

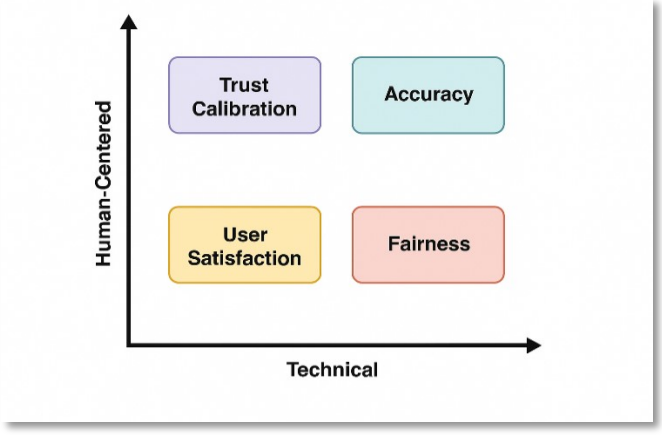

Fig. 8. Classification of Evaluation Dimensions for Human-Centric AI Across Technical, Ethical, and User-Centric Metrics.

The visual taxonomy clarifies how conventional benchmarks (e.g., precision, recall) can be complemented with emergent human-centered metrics such as trust calibration, fairness indices, and subjective satisfaction scores. Importantly, this classification encourages interdisciplinary collaboration in designing comprehensive testbeds, ensuring that AI systems are not only technically sound but also ethically responsible and user aligned. By organizing metrics along with this triadic structure, the figure provides a practical framework for researchers and practitioners to develop balanced, inclusive, and regulatory-compliant evaluation pipelines.

6. RESEARCH GAPS AND FUTURE DIRECTIONS

Despite the growing momentum around Human-Centric AI (HCAI), significant research gaps remain. These gaps hinder the development of truly trustworthy, context-aware, and ethically aligned AI systems. This section outlines unresolved challenges, emerging interdisciplinary frontiers, and key directions that define the roadmap for the next decade of HCAI innovation.

6.1 Open Problems in Human-Centric Model Design

While progress has been made in embedding human values into AI systems, model design still struggles with several core challenges:

- The trade-offs between accuracy and interpretability remain unresolved. Achieving high performance often conflicts with model transparency.
- Contextual sensitivity is limited, especially in dynamic environments where user behavior and goals change in real time.
- Robustness to adversarial inputs in human-facing applications is poorly addressed, especially in critical domains like healthcare and law.
- Scalability of personalized systems is a major bottleneck, as fine-tuning models for diverse individuals increase computational cost.

Current architecture often focuses on either functionality or human alignment—but seldom both. Bridging this divide requires novel hybrid models that are simultaneously interpretable, adaptive, and scalable.

6.2 Integrating Neuroscience and Cognitive Psychology

Human-Centric AI aspires not only to assist but to think and interact in ways compatible with human cognition. This demands deeper integration with neuroscience and cognitive psychology to:

- Model cognitive load and align explanations with human reasoning patterns.
- Understand human biases to design AI that compensates or adapts accordingly.
- Emulate attention mechanisms, memory dynamics, and learning curves observed in the human brain.

Emerging areas such as neuro-symbolic AI, cognitive architectures, and computational empathy offer promising convergence points. However, interdisciplinary alignment remains sparse, requiring more collaborative frameworks between AI developers, neuroscientists, and psychologists.

6.3 Cross-Disciplinary Collaboration and Co-Design

One of the most pressing research gaps is the lack of methodological pluralism in AI development. Human-Centric systems must be co-designed with input from:

- Sociologists, to assess societal impact.
- Ethicists define morally sound constraints.
- Design experts, to ensure usability and accessibility.
- End-users, to co-create solutions aligned with real-world needs.

Current AI workflows are predominantly technical, and developer driven. Future HCAI research must institutionalize co-design processes, participatory evaluations, and inclusive design pipelines from ideation through deployment.

6.4 Path Towards Generalizable and Culturally Adaptive Systems

HCAI systems must function across diverse cultures, languages, and demographics. However, current models are often biased toward English-speaking, Western-centric datasets. to progress:

- Cultural ontologies must be incorporated into model training.
- Multi-lingual and multi-modal datasets should be developed with underrepresented communities.
- Adaptive interfaces should respond to local social norms, literacy levels, and interaction preferences.

Generalization must be redefined not as performance consistency across datasets, but as ethical and contextual adaptability across human populations. Future research should prioritize culturally aware architectures that remain responsive to sociocultural variability without losing functional robustness.

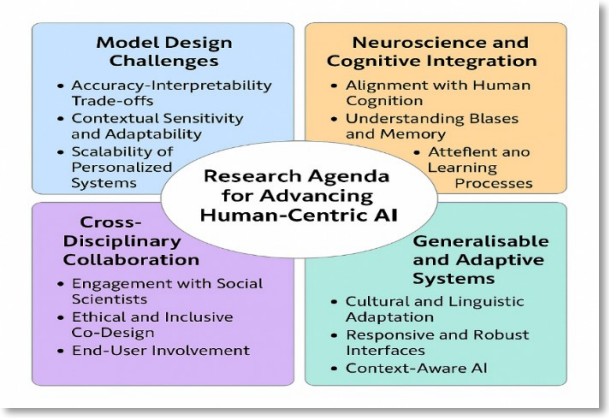

Fig. 9. Strategic Research Agenda for Advancing Human-Centric AI Through Interdisciplinary Innovation and Cultural Adaptability.

As illustrated in Figure 12, advancing Human-Centric AI requires a multi-dimensional research agenda spanning cognitive integration, cross-disciplinary design, sociocultural adaptability, and ethical governance. This figure encapsulates the interrelated directions that must be prioritized to transition from isolated innovations toward globally deployable, inclusive, and trustworthy AI systems.

7. DISCUSSION AND SUMMARY OF KEY INSIGHTS

Several critical insights emerge from the discourse:

- Foundational principles—such as transparency, autonomy, and fairness—are not merely ethical ideals but essential design imperatives for AI systems operating in human-facing environments.
- Enabling technologies like Explainable AI, Human-in-the-Loop learning, affective computing, and multi-agent collaboration are pivotal to aligning machine behavior with human expectations.
- Successful societal integration of HCAI requires attention to trust, privacy, legal compliance, cultural diversity, and psychological acceptance—factors often neglected in traditional AI frameworks.
- Existing evaluation methods remain inadequate. A shift toward user-centric, multi-metric, and context-sensitive evaluation paradigms is essential for accurate and holistic assessment.

Together, these insights underscore the need for an integrated approach where technical innovation is guided by socio-ethical foresight.

7.2 Implications for Researchers, Developers, and Policymakers

For researchers, the findings highlight the need for interdisciplinary collaboration, incorporating insights from cognitive science, law, philosophy, and design into algorithmic development. Technical excellence must be paired with human relevance. For developers, HCAI calls for a reconfiguration of design pipelines: from user data collection and model training to deployment and feedback mechanisms. Usability, adaptiveness, and trustworthiness must be prioritized alongside accuracy and scalability. For policymakers, the emergence of HCAI presents both a challenge and an opportunity. Legal frameworks must evolve to safeguard autonomy, prevent discrimination, and enforce explainability. Policy should not lag behind innovation but shape its course through proactive governance and standards.

7.3 Vision for a Human-Aligned AI Future

Looking forward, the vision for HCAI is ambitious yet necessary: to build AI systems that are not only intelligent but wise; not only efficient but empathetic. Achieving this vision requires redefining what constitutes success in AI moving beyond performance metrics toward human values.

Human-Centric AI must become the default paradigm, not a specialized niche. Realizing this future entails long-term investment in education, inclusive innovation, co-design practices, and regulatory alignment. Only through this collective commitment can we ensure that AI serves humanity not just functionally, but meaningfully and equitably.

As depicted in Figure 13, the realization of Human-Aligned AI depends on the convergence of three foundational pillars: technical robustness, ethical design, and societal integration. The diagram maps the trajectory from research to deployment,

identifying critical enablers such as interdisciplinary collaboration, policy reform, user-centric evaluation, and cultural adaptability.

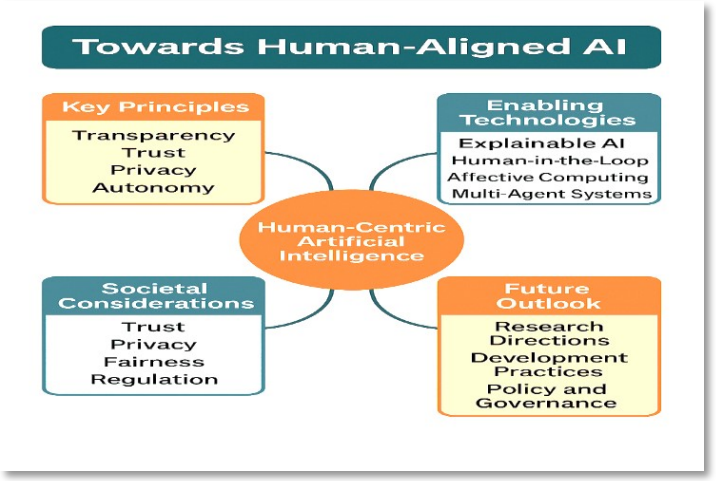

Fig. 10. Strategic Pathways Towards Human-Aligned AI – Integrating Technical Innovation, Ethical Design, and Societal Impact.

This visual synthesis reinforces the review's central message that the future of AI must be shaped not only by advances in learning algorithms and model performance, but by principled attention to human values, psychological engagement, and equitable global deployment. The figure thus offers a blueprint for stakeholders across academia, industry, and governance to co-navigate the transition from intelligent automation to meaningful augmentation.

7.4 Applications of contemporary technologies considered innovative

The role of ethics in medicine, with a focus on how medical education incorporates human intelligence, particularly social and emotional intelligence, to promote moral decision-making [49]. Integrating learning into classes while taking into account concerns about usability, cost, and online distractions [50]. Artificial intelligence (AI)-driven robotics advancements offer the ground-breaking potential to eliminate these barriers and enhance palliative care delivery [51][52]. Artificial intelligence (AI)-powered secondary data analysis and predictive analytics are utilized to assess hazards and predict burnout. Monitoring and management techniques are required to promote students' academic and health results while improving the air quality in educational institutions. It is crucial to assess water resources and climate change to determine the degree of the threat they pose to public health, as we are either in the post-pandemic era or a new pandemic might occur. In order to lessen and prevent corruption in public health services, employees used a variety of essential information sources, including the internet, media outlets, printed materials, online search engines, and the opinions of their colleagues. Job satisfaction and information literacy are relatively high. One of the most important methods required for coronavirus detection is machine learning. It is a collection of sophisticated algorithms that can analyze medical data and spot trends and signs of disease. It is used to accurately and rapidly evaluate medical pictures, including chest X-ray scans, and gives information about each image. Other important scientific fields and industries that can profit from improvements in structural health monitoring include using artificial intelligence (AI), machine learning, and the Internet of Things in cybersecurity for smart agriculture, controlling viral hepatitis, and safeguarding the environment for sustainable forest management. Graph theory algorithms are used to describe and analyze the dynamics of COVID-19 infections with remarkable precision thanks to new techniques and tools. The scope of risks and vulnerabilities related to climate change, as well as the best ways to manage them overall with Climate Justice and One Health, are more widely accepted than the extent and speed of these changes, according to Reimagining Intelligence, a thorough review of human-centric AI systems, and their integration with society and healthcare.

8. CONCLUSIONS

Human-Centric Artificial Intelligence (HCAI) represents a necessary evolution in AI development one that prioritizes ethical alignment, social trust, and individual empowerment over mere algorithmic performance. This review has examined the conceptual foundations, enabling architectures, societal integration challenges, evaluation frameworks, and future research directions that collectively define the trajectory of HCAI. Determine unresolved research issues and present a strategic plan that combines ethical philosophy, cognitive science, and participatory design. For researchers, developers, and policymakers working to create AI systems that are not only intelligent but also responsible, inclusive, and in line with

human dignity especially in life-critical fields like healthcare—this review seeks to act as both a foundational reference and a forward-looking roadmap.

Funding

The authors had no institutional or sponsor backing.

Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Acknowledgment

The authors extend appreciation to the institution for their unwavering support and encouragement during the course of this research.

References

[1] Y. Rong, T. Leemann, T.-T. Nguyen, L. Fiedler, P. Qian, et al., "Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.46, pp.2104-2122, April 2024. https://doi.ieeecomputersociety.org/10.1109/TPAMI.2023.3331846

[2] Q. V. Liao and J. W. Vaughan, "AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap," *ArXiv*, pp.1-33, 2023. https://doi.org/10.48550/arXiv.2306.01941

[3] I. P. Adamopoulos, "Corruption and political interventions in public health authorities–Hellenic Republic Region of Attica: Conceptual analysis study," *European Journal of Environment and Public Health*, vol.7, no.3, pp.em0139, 2023. https://doi.org/10.29333/ejeph/13171

[4] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol.58, pp.82-115, June 2020. https://doi.org/10.1016/j.inffus.2019.12.012

[5] B. Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," *International Journal of Human–Computer Interaction*, vol.36, no.6, pp.495-504, 2020. https://doi.org/10.1080/10447318.2020.1741118

[6] A. Holzinger, A. Carrington, and H. Müller, "Measuring the Quality of Explanations: The System Causability Scale (SCS)," *KI - Künstliche Intelligenz*, vol.34, no.193–198, January 2020. https://doi.org/10.1007/s13218-020-00636-z

[7] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and Explainability of Artificial Intelligence in Medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, pp.e1312, 2019. https://doi.org/10.1002/widm.1312

[8] N. Scharowski, S. A. C. Perrig, M. Svab, K. Opwis, and F. Brühlmann, "Exploring the effects of human-centered AI explanations on trust and reliance," *Frontiers in Computer Science*, vol.5, pp.1-15, July 2023. https://doi.org/10.3389/fcomp.2023.1151150

[9] M. Tahaei, M. Constantinides, D. Quercia, and M. Muller, "A Systematic Literature Review of Human-Centered, Ethical, and Responsible AI," *ArXiv*, pp.1-39, 2023. https://doi.org/10.48550/arXiv.2302.05284

[10] W. Xu and Z. Gao, "Enabling Human Centered AI: A Methodological Perspective," *ArXiv*, pp.1-6, 2023. https://doi.org/10.48550/arXiv.2311.06703

[11] J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, and M. Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI," *SSRN*, pp.1-39, 2020. http://dx.doi.org/10.2139/ssrn.3518482

[12] European Commission, "Ethics Guidelines for Trustworthy AI: High-Level Expert Group on AI," Apr. 2019

[13] B. Goodman and S. Flaxman, "European Union Regulations on Algorithmic Decision Making and a 'Right to Explanation'," *AI Magazine*, vol. 38, no. 3, pp. 50–57, 2017. https://doi.org/10.1609/aimag.v38i3.2741

[14] D. D. Farhud and S. Zokaei, "Ethical Issues of Artificial Intelligence in Medicine and Healthcare," *Iranian Journal of Public Health*, vol.50, no.11, pp.i-v, 2021. https://doi.org/10.18502/ijph.v50i11.7600

[15] A. D. Selbst, D. M. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and Abstraction in Sociotechnical Systems," In FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp.59 - 68, January 2019. https://doi.org/10.1145/3287560.3287598

[16] F. Pervez, M. Shoukat, M. Usama, M. Sandhu, S. Latif, and J. Qadir, "Affective Computing and the Road to an Emotionally Intelligent Metaverse," *IEEE Open Journal of the Computer Society*, vol.5, pp.195 - 214, April 2024. https://doi.org/10.1109/OJCS.2024.3389462

[17] L. Floridi and J. Cowls, "A Unified Framework of Five Principles for AI in Society," *Harvard Data Science Review*, vol.1, no.1, 2019. https://doi.org/10.1162/99608f92.8cd550d1

[18] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, et al., "Datasheets for datasets," *Communications of the ACM*, vol.64, no.12, pp.86 - 92, 2019. https://doi.org/10.1145/345872

[19] D. Zytko, P. J. Wisniewski, S. Guha, E. P. S. Baumer, and M. K. Lee, "Participatory Design of AI Systems: Opportunities and Challenges Across Diverse Users, Relationships, and Application Domains," In CHI EA '22: CHI

Conference on Human Factors in Computing Systems Extended Abstracts, pp.1-4, 2022. https://doi.org/10.1145/3491101.351650

[20] A. Naeem, A. S. u. Rehman, A. Rasheed, "Evaluating Cultural Adaptation in AI Translations: A Framework and Implications for Literary Works," In AI Applications for English Language Learning, pp.1-30, 2025. https://doi.org/10.4018/979-8-3693-9077-1.ch010

[21] I. Reyes-Amezcua, G. Ochoa-Ruiz, and A. Mendez-Vazquez, "Adversarial Robustness on Artificial Intelligence," In What AI Can Do, pp.1-13, 2023.

[22] R. Al-Fakhri and R. Dixon, "Federated Learning for Privacy-Preserving AI: Emerging Use Cases," *IEEE Access*, vol. 9, pp. 123456–123465, 2021. https://doi.org/10.1109/ACCESS.2021.3064520

[23] M. Veale, M. G. Van-Kleek, and R. D. Binns, "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making," In CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, no.440, pp.1-14, 2018. https://doi.org/10.1145/3173574.31740

[24] R. A. Caruana, Y. Lou, J. E. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission," In KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1721 - 1730, 2015. https://doi.org/10.1145/2783258.2788613

[25] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A Survey of Methods for Explaining Black Box Models," In ACM Computing Surveys (CSUR), vol.51, no.5, pp.1-42, 2018. https://doi.org/10.1145/3236009

[26] W. M.P. van der Aalst, "Process Mining in the Large: A Tutorial," In Business Intelligence, pp.33–76, 2014. https://doi.org/10.1007/978-3-319-05461-2_2

[27] S. Wachter, B. Mittelstadt, and C. Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology*, vol. 31, no. 2, pp. 841–887, 2018.

[28] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," In NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp.4768 - 4777, December 2017.

[29] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," In KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135 - 1144, August 2016. https://doi.org/10.1145/2939672.2939778

[30] D. Kahneman, "Thinking, Fast and Slow," Farrar, Straus and Giroux, 2011

[31] R. B. Allen, B. Friedman, and H. Nissenbaum, "Bias in computer systems," *ACM Transactions on Information Systems (TOIS)*, vol.14, no.3, pp.330 - 347, July 1996. https://doi.org/10.1145/230538.2305

[32] J. Doshi-Velez and B. Kim, "Towards a Rigorous Science of Interpretable Machine Learning," *ArXiv*, pp.1-13, 2017. https://doi.org/10.48550/arXiv.1702.08608

[33] T. G. Dietterich, "Steps Toward Robust Artificial Intelligence," *AI Magazine*, vol. 38, no. 3, pp. 3–24, 2017. https://doi.org/10.1609/aimag.v38i3.2756

[34] K. M. Richmond, S. M. Muddamsetty, T. Gammeltoft-Hansen, H. P. Olsen, and T. B. Moeslund, "Explainable AI and Law: An Evidential Survey," *Digital Society*, vol.3, no.1, pp.1-33, December 2023. https://doi.org/10.1007/s44206-023-00081-z

[35] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review*, vol.56, pp.3005–3054, August 2022. https://doi.org/10.1007/s10462-022-10246-w

[36] M. Kamar, "Directions in Hybrid Intelligence: Complementing AI Systems with Human Intelligence," In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, pp. 4070–4077, 2019.

[37] X. Amatriain, "Mining large streams of user data for personalized recommendations," *ACM SIGKDD Explorations Newsletter*, vol.14, no.2, pp. 37 - 48, April 2023. https://doi.org/10.1145/2481244.2481250

[38] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," In ICML'17: Proceedings of the 34th International Conference on Machine Learning, vol.70, pp. 1126 - 1135, August 2017.

[39] R. Picard, "Affective Computing," MIT Press, 1997

[40] P. Ekman, "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life," Times Books, 2003

[41] L. Busoniu, R. Babuska, and B. D. Schutter, "A Comprehensive Survey of Multiagent Reinforcement Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol.38, no.2, pp.156 - 172, March 2008. https://doi.org/10.1109/TSMCC.2007.913919

[42] F. Herrera, "Reflections and attentiveness on eXplainable Artificial Intelligence (XAI). The journey ahead from criticisms to human–AI collaboration," *Information Fusion*, vol.121, pp.103133, September 2025. https://doi.org/10.1016/j.inffus.2025.103133

[43] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," In 2015 53rd Annual Allerton Conference on Communication, Control, and Computing, pp.1-6, 2016. https://doi.org/10.1109/ALLERTON.2015.7447103

[44] S. Barocas and A. D. Selbst, "Big Data's Disparate Impact," *California Law Review*, vol.104, no.3, pp.671-732, June 2016.

[45] M. Graves and E. Ratti, "A capability approach to ethical development and internal auditing of AI technology," *Journal of Responsible Technology*, vol.22, pp.100121, June 2025. https://doi.org/10.1016/j.jrt.2025.100121

[46] No refrences

[47] O. B. Akinnagbe, "Human-AI Collaboration: Enhancing Productivity and Decision-Making," *International Journal of Education, Management, and Technology*, vol.2, no.3, pp.387-417, 2024. https://doi.org/10.58578/ijemt.v2i3.4209

[48] M. J. McGrath, O. Lack, J. Tisch, and A. Duenser, "Measuring trust in artificial intelligence: validation of an established scale and its short form," Frontiers in Artificial Intelligence, vol.8, pp.1-14, May 2025. https://doi.org/10.3389/frai.2025.1582880

[49] P. Pramila, I. Adamopoulos, and A. Rijal, "Mobile Learning in Medical Education," In Teaching in the Age of Medical Technology, pp.195-226, 2025. https://doi.org/10.4018/979-8-3373-1519-5.ch007

[50] E. Karageorgaki, I. P. Adamopoulos, and A. Valamontes, "Organizational behavior in the healthcare environment: A study of psychiatric services," Electronic Journal of General Medicine, vol.22, no.4, pp.em652, 2025. https://doi.org/10.29333/ejgm/16300

[51] M. Gök, E. Cengiz, and M. M. Mijwil, "Unveiling Protein-Ligand Interactions: Regression Methods for Binding Affinity Prediction," *International Journal on Engineering Applications*, vol.12, no.6, pp.508-513, December 2024. https://doi.org/10.15866/irea.v12i6.25407

[52] P. P. Garg, J. Jayashree, and J. Vijayashree, "evolutionizing healthcare: The transformative role of artificial intelligence," In Responsible and Explainable Artificial Intelligence in Healthcare, pp.1-23, 2025. https://doi.org/10.1016/B978-0-443-24788-0.00001-7