JOURNAL OF
Emerging Paradigms
in Computing Systems
JEPCS

Journal of Emerging Paradigms in Computing Systems (JEPCS) | Volume 03, Issue No. 01, 2026

ARTICLE TYPE: Review Article

Article ID: JEPCS-2026-0301-0039

Human-Centered Artificial Intelligence for Intelligent Decision Support

Harshwardhansinh Chauhan[1,*], Rocky Upadhyay[1], Sanjay Sharma[2], Shivam Upadhyay[1]

[1] Department of CSE, Drs. Kiran & Pallavi Patel Global University (KPGU), Vadodara, Gujarat, India.
[2] Sigma University, Vadodara, Gujarat, India
*Corresponding author: hwkc392@gmail.com
Contributing authors: hwkc392@gmail.com, rko75316060@gmail.com, drsanjaysharma8388@gmail.com, shivam91092@gmail.com

Received: May 11, 2026 **Accepted:** June 27, 2026

Abstract- Artificial Intelligence (AI) is now becoming prevalent in other important sectors such as healthcare, finance, governance, industry, etc., where decision support systems (DSS) are utilized. AI models can analyze massive amounts of data and identify patterns that are not easily noticeable to humans, but their black-box nature can make them hard to understand, explain, and trust. Human-Centered Artificial Intelligence (HCAI) has come to be a design paradigm where humans, not algorithms, are at the heart of the decision-making process with a focus on explainability, calibrated trust, meaningful human control, and collaborative human–AI interaction. In this paper, a theoretical and conceptual research on the HCAI for Intelligent Decision Support Systems (IDSS) is presented. It summarizes the basic concepts of HCAI, provides a discussion of mechanisms of explainability and trust calibration, and presents a five-layer conceptual framework (Data and Knowledge, AI Reasoning, Decision Logic, Human Interaction, Governance and Trust layers) for designing human-centric intelligent decision support architectures. Prominent design principles, typical application scenarios, and open research questions are explored, as well as future tracts such as neuro-symbolic reasoning, reinforcement learning from human feedback, and agentic orchestration with human supervision. This paper does not present an implementation or empirical evaluation, but is more of a conceptual synthesis to support researchers and practitioners in designing intelligent decision support systems that are more transparent, trustworthy, and human-aligned.

Keywords- Human-Centered AI; decision support systems; explainable AI; trust calibration; intelligent systems

1. Introduction

In the past few years, the use of Artificial Intelligence for human decision making has increased at a drastic rate in almost every sector of the society. In the last decade, investments in AI have skyrocketed worldwide, and companies from all regions are already using AI tools to boost their operations. Over the last decade, global investment in AI has grown significantly, with AI tools reported to be widely used by companies across regions. Use of intelligent decision support systems (IDSS) that integrate AI-based inference with interactive interfaces is growing to assist clinicians in disease diagnosis, assist operators of critical infrastructure, assist in resource allocation and assist managers for financial risk assessment. These systems offer quicker analysis, greater scope of potential solutions, and less reliance on scarce human resources.

While AI-based decision support can offer these benefits, it also has its drawbacks. Many successful predictive models are actually black box models, which give out predictions without any explanation or justification of how or why the prediction was made. This opacity has led to overreliance and underreliance in trust with users relying too much on the AI recommendation and rejecting useful suggestions [1], [5]. Moreover, biases found in the training data of AI systems can

be amplified, raising ethical concerns when they make mistakes, and leaving users uneasy when their interactions with the system are beyond their control and without explanation. In high-stakes environments like healthcare, criminal justice, and safety critical systems, a poorly researched recommendation can have significant repercussions, which is particularly troublesome.

Human-Centered Artificial Intelligence (HCAI) has become a solution to these challenges. HCAI is not about full automation; it is about having systems that have high automation levels, with high levels of human control and human judgment, values and oversight, so that human judgment, values and oversight become part of the decision process [2] and [3]. When it comes to decision support, this approach is not to view AI as a decision maker, but as an assistant that supports the human thought process and keeps the human well informed, accountable, and in charge of the decision.

This paper provides theoretical and conceptual analysis of HCAI for intelligent decision making support. There is no implementation, prototype, or empirical evaluation offered; the goal is to take concepts from the HCI, EAI, and decision support fields and combine them into a unified concept model. The rest of the paper is organized as follows. Section 2 presents the review methodology, followed by related work in Section 3. Section 4 presents the foundations of HCAI. Explainability, trust, and human–AI collaboration are discussed in Section 5. Section 6 presents the five-layer conceptual model for HCAI-based intelligent decision support. The design principles are described in Section 7, representative application areas in Section 8, open challenges in Section 9, future research directions in Section 10, and the conclusion in Section 11.

2. Review Methodology

2.1. Review Design and Scope

This article is positioned as a conceptual review and theoretical synthesis rather than an empirical research article. The review was designed to identify recurring architectural patterns, human roles, explainability mechanisms, trust and reliance issues, and governance requirements relevant to Human-Centered Artificial Intelligence (HCAI) and Intelligent Decision Support Systems (IDSS). The literature scope primarily covers 2015–2026, while seminal earlier publications are retained where they provide foundational concepts.

2.2. Literature Search Strategy

The literature search used scholarly databases and publisher/index search interfaces covering computer science, human-computer interaction, artificial intelligence, decision support, responsible AI, and governance. Search strings combined the following concepts: (“human-centered AI” OR HCAI) AND (“decision support” OR “intelligent decision support”); (“human-AI collaboration” OR “human-AI teaming”) AND (trust OR reliance OR calibration); (“explainable AI” OR XAI) AND (decision making OR decision support); and (“AI governance” OR “trustworthy AI”) AND (human oversight OR accountability). Reference lists of highly relevant reviews and foundational studies were additionally examined to identify directly relevant literature.

2.3. Inclusion and Exclusion Criteria

Studies were included when they were peer-reviewed journal/conference papers, authoritative books or reports, systematic reviews, frameworks, taxonomies, design guidelines, or empirical studies directly addressing HCAI, human-AI interaction, explainability, trust/reliance, intelligent decision support, governance, or human oversight. Studies were excluded when they focused only on algorithmic accuracy without a human decision component, were duplicates, were promotional/non-scholarly material, or were outside the conceptual scope of human-centered decision support.

2.4. Screening and Thematic Analysis

Screening was conducted in two stages: title/abstract relevance screening followed by full-text assessment. Relevant literature was coded against five analytical dimensions: (1) architecture and system layers, (2) human role and intervention points, (3) explainability and transparency, (4) trust, reliance and uncertainty, and (5) governance, accountability and lifecycle controls. The coded findings were compared thematically and synthesized into the proposed five-layer reference architecture and the operational design principles. Because this is a conceptual review rather than a meta-analysis, no pooled statistical effect estimate is claimed.

2.5. Review Questions

- RQ1: What architectural and interaction patterns characterize HCAI-based decision support?
- RQ2: How are explainability, uncertainty, trust and reliance addressed?
- RQ3: How and where are humans involved in AI-supported decisions?
- RQ4: Which governance and accountability mechanisms are required across the lifecycle?
- RQ5: What research gaps remain for high-stakes and emerging agentic decision-support systems?

Table 1: *Comparison of prior HCAI, human-AI interaction, explainability, and governance frameworks.*

Framework / study	Layers / structure	Human role	Explainability	Governance	Limitation
Shneiderman HCAI [2],[3]	Control–automation framing	High human control with automation	Understandable and reliable interaction	Organizational responsibility and trustworthiness	High-level; not a detailed IDSS architecture
Amershi et al. [7]	Human-AI interaction guidelines	Observe, correct, control, feedback	Behavior/feedback communication	Lifecycle-oriented guidance	Interaction-focused
Wang et al. [4]	Clinical CDSS design themes	Clinician-centered workflow	Explainability and usability	Clinical sociotechnical constraints	Domain-specific
Gomez et al. [5]	Interaction-pattern taxonomy	Often final accept/reject; richer patterns proposed	Interaction-dependent	Limited institutional governance	Taxonomy rather than unified architecture
Arrieta et al. [12]	XAI taxonomy	Human interpretation	Interpretable and post-hoc methods	Responsible-AI linkage	Explainability-centered
HLEG Trustworthy AI [13]	Seven trustworthy-AI requirements	Human agency/oversight	Transparency	Strong governance/accountability	Not a detailed IDSS architecture
Proposed HCAI-IDSS	Five functional layers + feedback	Question, correct, reject, escalate	Evidence, uncertainty, contextual explanations	Audit, accountability, bias, compliance	Conceptual; needs empirical validation

3. Related Work

Over the last few years, there have been a number of investigations combining HCD with AI decision support. This paper is a systematic review of ninety studies published from 2015 to 2025 and proposed a multi-layer architectural taxonomy which maps the development of decision support systems from simple black-box generative models to fully orchestrated, agentic architecture with human oversight. The review also identified the contributions related to HCAI and linked them to the UN Sustainable Development Goals, showing that these contributions tend to be focused on health, urban, and industrial areas and that there is comparatively little in terms of equity-focused goals. Table 1 shows the comparison with prior studies.

In the clinical environment, Wang et al., analyzed the design and assessment of clinical decision support systems with AI, and found that while the predictive power of these models can be high, they can be hindered by recurring sociotechnical barriers such as workflow mismatch, attitudinal barriers, and usability issues, which hinder their adoption [4]. Work by Panigutti et al. on co-design research approaches for developing explainable AI decision support systems in health care settings highlights the importance of explainability as a key consideration that must be discussed with the end user and is not a pre-established technical design component [6].

At a more general level of human-AI interaction, Gomez et al. developed a taxonomy of human-AI interaction patterns in the context of an AI system assisting in decision-making, and found that the majority of systems currently in use only involve humans at the end of the decision-making process, that is, when humans choose to accept or reject the AI's advice [5]. This is echoed in various fields, and leads to the search for design frameworks to address this by involving the human in the process, not merely an add-on to the AI output.

Although these works give meaningful insights into domain-specific and empirical aspects, there is a need for a compact, theory-driven conceptual model that combines architectural layering, explainability mechanisms, trust calibration and governance in a single reference model. This is why the present paper seeks to fill up the gap by making a synthesis that is theoretical alone.

Emerging agentic AI adds a further architectural challenge. When AI systems can plan, invoke tools, coordinate subtasks and execute actions, human oversight must move from a single final approval point toward risk-sensitive decision gates. This strengthens the need for explicit intervention, logging, escalation and authority boundaries in HCAI-based decision support.

Research on trustworthy and responsible AI further indicates that transparency, human agency, robustness, fairness, privacy and accountability are interdependent rather than isolated requirements. This supports treating governance as a cross-cutting architectural concern. In the proposed framework, governance controls are therefore connected to data, reasoning, decision logic and interaction rather than being placed only at the end of the pipeline.

Human-AI teaming research reinforces this distinction. Effective collaboration depends on complementary capabilities, appropriate allocation of tasks, user understanding of system limitations, and mechanisms for recovering from AI error. Appropriate reliance is therefore a better target than maximum reliance. In decision-support settings, the user should be able to recognize when the system is likely to be useful and when domain knowledge should override it.

Recent HCAI literature also emphasizes that human-centeredness must be evaluated at the level of the whole sociotechnical system. Frameworks that concentrate only on model interpretability do not necessarily specify how users should intervene, how responsibility should be allocated, or how system behavior should be monitored after deployment. Conversely, governance frameworks can define accountability requirements without describing the interaction mechanisms through which a user questions or corrects an AI recommendation.

4. Foundations of Human-Centered Artificial Intelligence

HCAI is a design philosophy and engineering discipline which considers human needs, values and capabilities as a primary focus during the development of AI systems, not as an "after-thought" once technical system performance has been optimized. The aim of well-designed systems is to attain simultaneously high levels of human control and high levels of computer automation, and it is neither necessary nor desirable for most consequential applications to pursue full automation at the cost of human oversight [2, 3], according to Shneiderman's HCAI framework. The framework also notes that design based on HCAI will be more likely to create reliable, safe, and trustworthy systems, which also relies on other aspects of the design such as organizational practices, governance structures, and interface design.

4.1. From Automation-Centered to Human-Centered Design

The design of traditional AI systems has tended to focus on the three key metrics of success: predictive accuracy, computational efficiency and autonomy, with usability and interpretability being secondary priorities. In HCAI, the ordering is reversed, with human performance, satisfaction, self-efficacy, and understanding being first-class design goals, and technical performance being measured with respect to these human performance goals. This does not mean that model accuracy is not a concern, but rather that the key is that a highly accurate model may be wrong because users may not be able to properly adjust their faith in the model's advice, may not be able to see when the model is incorrect, or may not be able to combine the output from a model with the decisions they are making.

5. Trust, Explainability, and Human-AI Collaboration

5.1. Explainable Artificial Intelligence

Explainable AI (XAI) is the collection of methods and design principles for rendering the actions of AI models comprehensible to human users. Model-agnostic methods like LIME approximate the local behavior of a complex model with a more interpretable surrogate model [8] or methods like SHAP [9] draw the conclusions of a model to individual input features using game-theoretic principles. In addition to offering a broad overview of the various types of XAI techniques, Arrieta et al. propose a taxonomy that classifies them into interpretable models and post-hoc explanation methods for black-box models, and expands the understanding of explainability to the broader concepts of trustworthy and responsible AI [12]. The DARPA Explainable AI program also made explainability an essential factor for successful human-machine teaming in real-world scenarios, not just an academic one [15].

5.2. Trust and Calibrated Reliance

Giving an explanation alone is not sufficient to ensure that users trust and use a system properly. Miller shows that useful explanations should be based on social science knowledge of the processes by which humans build, and assess, explanations in their everyday reasoning, and that this should be based on technical concepts of feature importance [10] alone is insufficient. Doshi-Velez and Kim also argue for a formal science of interpretability that bridges the gap between explanation techniques and measurable human understanding and downstream decision quality, rather than assuming that interpretability is a desirable property. In the realm of decision support specifically, evidence indicates that the effect of explanations on reliance is context-dependent; that is, explanations can lead to decreased over-reliance in some contexts, have minimal effect in others, and even increase miscalibrated trust in some contexts, suggesting that explanations should be tailored to the user, task and risk context, rather than applied in a generic manner [1].

The way humans and AI systems collaborate to share information and tasks affects the success of decision support, and it is difficult to explain. Gomez et al. recognize common interaction patterns, from basic accept/reject decisions to more complex patterns of interaction, such as repeated rounds of feedback, negotiation, and problem reformulation, where most existing systems fall on the simpler end of the spectrum [5]. Collaboration is a recurring theme in the literature of HCAI, and it is a major motivation behind the proposed framework of Section 6, which is centered on richer collaboration, that is, the human is able to question, adjust, or redirect the AI's reasoning, rather than simply approving its final output.

6. Conceptual Framework for HCAI-Based Intelligent Decision Support

6.1. Overview

To build on the principles presented above, a conceptual framework for intelligent decision support systems with HCAI is proposed in this paper, consisting of five layers. The framework is designed as a reference model and not as an implementation per se; it outlines the functional layers which should be incorporated by a human-centred IDSS and the relationship between these functional layers. The five layers are: (i) Data and Knowledge Layer, (ii) AI Reasoning Layer, (iii) Decision Logic Layer, (iv) Human Interaction Layer, and (v) Governance and Trust Layer. Unlike traditional pipeline designs, where data is passed from input to output, the proposed framework is designed as a cycle, with data from the Data and Knowledge and AI Reasoning layers propagating back to the Human Interaction and Governance layers, where it can be used to continually improve the system's behavior over time. As shown in Figure 1, the proposed architecture

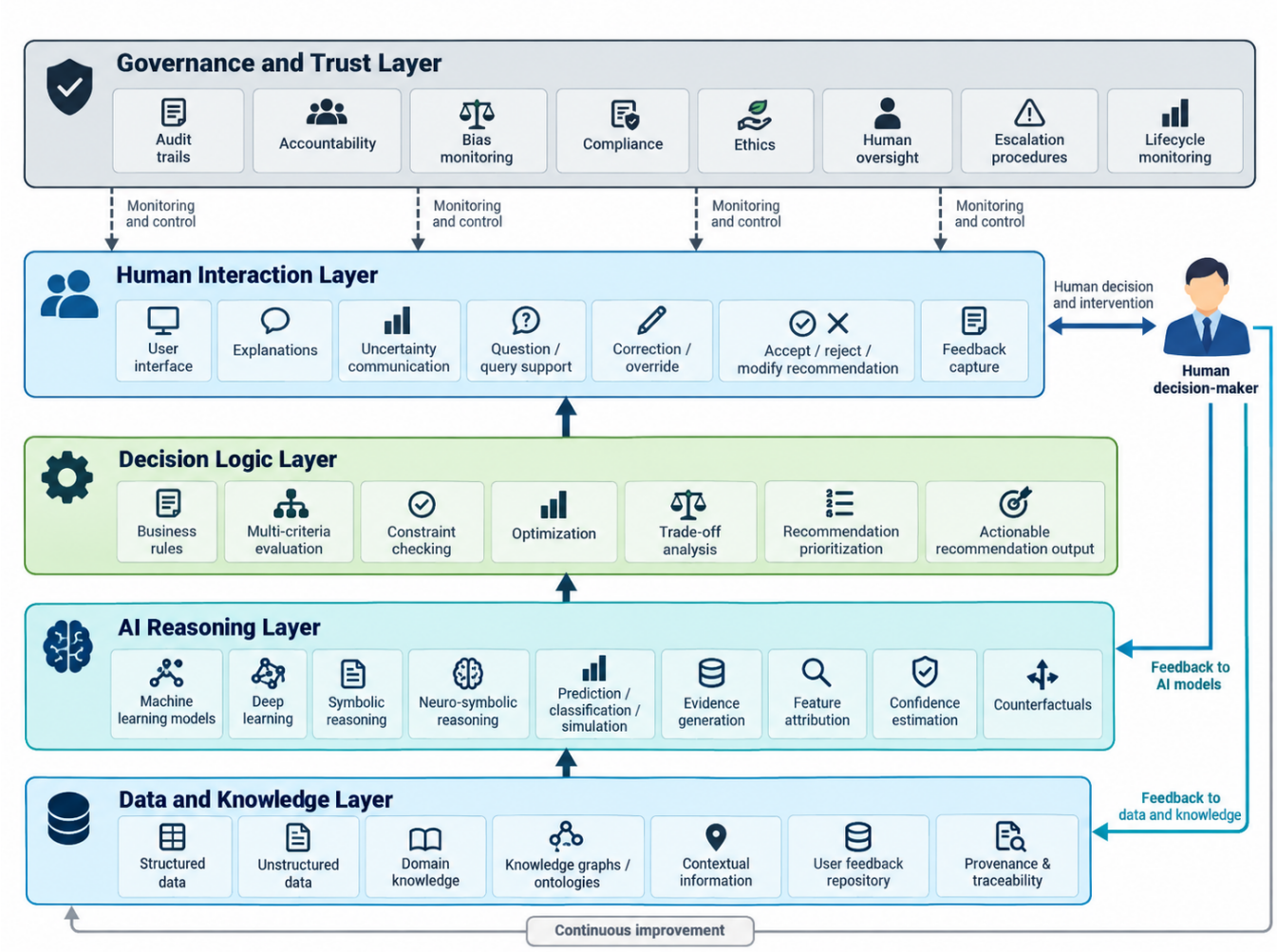


Figure 1: *Five-layer conceptual architecture for HCAI-based intelligent decision support.*

does not treat the five layers as isolated functional blocks. Instead, the layers are connected through decision flow, human intervention, governance controls, and feedback mechanisms.

6.2. Data and Knowledge Layer

The main role of this layer is to collect, process and store structured and unstructured data, domain knowledge and contextual information that is used to support the reasoning of the system. It involves heterogeneous data sources, knowledge graphs and ontologies for domain knowledge, and means to collect user feedback in the form of new labeled examples. The layer is concerned with the integrity of the data and knowledge used as its basis, as well as with provenance and traceability, where these are treated as first-class concerns, as the trustworthiness of the recommendations downstream relies on the trustworthiness of the data and knowledge used as a basis.

6.3. AI Reasoning Layer

The AI Reasoning Layer is used to apply the predictive, diagnostic, or generative models derived from machine learning, deep learning, symbolic reasoning, or hybrid neuro-symbolic approaches to generate candidate outputs, like predictions, classifications, or scenario simulations. This layer is also associated with explanation-generation mechanisms to ensure that it does not make unqualified decisions, but rather the output statements are backed by evidence, such as feature attributions, confidence measures, or counterfactual alternatives.

6.4. Decision Logic Layer

This layer takes the results of the AI Reasoning Layer and outputs actionable recommendations using decision rules, multi-criteria evaluation, constraint satisfaction or optimization procedures. It must consider trade-offs, impose domain-specific constraints, and prioritize and/or filter candidate solutions prior to giving them to the human decision-maker. By

Table 2: *Operational assessment criteria for the proposed HCAI design principles.*

Principle	Operational requirement	Example assessment criterion
Human authority	A named human decision owner exists for consequential decisions.	Documented owner, override path and escalation procedure.
Context-sensitive explanation	Explanation matches user expertise, task and risk.	User evaluation confirms that users can identify rationale and limitations.
Explicit uncertainty	Recommendation includes confidence/uncertainty and known limitations.	Low-confidence or out-of-distribution cases trigger a defined response.
Graceful degradation	System fails safely outside expected operating conditions.	Fallback, escalation or safe-stop behavior is documented and tested.
Corrective feedback	Users can submit traceable corrections.	Feedback is logged, reviewed and linked to update procedures.
Calibrated reliance	System aims for appropriate acceptance/rejection, not maximal trust.	Evaluation measures decisions on both correct and incorrect AI advice.
Auditability	Recommendations and interventions can be reconstructed.	Relevant input, model/version, output, user action and policy checks are logged.
Continuous user participation	Representative users participate throughout lifecycle.	Evidence exists at requirements, evaluation, deployment and post-deployment review.

maintaining this layer explicit and accessible for inspection (not folding the logic in an opaque end-to-end model), users and auditors can see how and why a specific recommendation is chosen over other options.

6.5. Human Interaction Layer

The Human Interaction Layer is the layer through which all communications between the system and its human users pass. It provides recommendations and explanations via suitable interfaces, solicits user feedback and corrections and communicates uncertainty in a manner that is appropriate to the user's level of expertise and the consequences of the decision. This layer also provides the facility for the user to ask a question or to reject the system's suggestions, thus making the human the last decision-maker in important decisions.

6.6. Governance and Trust Layer

The last layer focuses on organizational and societal aspects of trustworthiness that transcend any particular transaction, including auditability, accountability for incorrect recommendations, monitoring for bias, regulatory compliance and adherence to ethical standards like those outlined for trustworthy AI [13]. It's not just a technical component, it's a series of institutional practices, review boards, audit trails, escalation procedures that enable the system to hold people accountable throughout the system's life cycle.

7. Design Principles for Human-Centered Intelligent Decision Support Systems

Based upon the principles of human-AI interaction [7] and human-centered computing [14] the following design principles are proposed as a guideline on how to build intelligent decision support systems within the framework in Section 6:

- Maintain human relevance and control over consequential decisions, in which the system is used to assist and not replace humans.
- Give explanation that is specific to the user's level of expertise, task context, and stakes of the decision, and not just one generic explanation.
- Be explicit about uncertainty have confidence levels and known limitations along with any recommendation.
- Enable graceful degradation and provide clear escalation procedures when the system's confidence is low or when input to the system is outside of the range of trained input.
- Allow users to feed back corrective information that will be collected and utilized to improve future system operation.
- Design for calibrated reliance, not maximal reliance, and actively discouraging over-trust and unwarranted dismissal of system output.
- Ensure auditing and tracking capabilities for recommendation, for accountability and post hoc review.
- Engage representative end users early in the design process and throughout the design life cycle, not just at deployment and evaluation time.

Operational assessment criteria for the principles are summarized in Table 2. These criteria are intended as design-review checks rather than validated universal thresholds.

8. Application Domains

The decision support for HCAI has been investigated in various application areas. Explainedness is key for the safe use of clinical decision support systems in healthcare, such as in diagnostic imaging, treatment planning and triage [1, 4, 6]. HCAI-based systems can be used in industry and production to assist control room operators in handling complex production processes; they can facilitate the sharing of human–machine situational awareness, and provide interpretable

alerts to help reduce operational risk [1]. Algorithmic decision support can be used in governance and public administration for processes like allocating resources and/or assessing risks, where human supervision can protect from unchecked automation and can help maintain democratic accountability [1]. Other areas include mobility and transportation, where the shared control and supervisory interfaces assist the pilots and vehicle operators in the process; environmental and agricultural management, where decision support interprets the information from sensors and human knowledge for irrigation and resource planning; and finance, where the decision support interprets the information and can be used in risk assessment without impacting the regulatory requirements. In all these areas, there's a common theme: in order for the system to be adopted, the users must be able to see, challenge, and properly trust its behavior.

9. Challenges and Open Issues

Several important issues remain unanswered, although there is a growing interest in HCAI.

- First, explanation alone is not sufficient for the appropriate level of reliance, some users might over-trust or under-trust the system even with the most convincing explanations and the link between explanation and reliance is sensitive and context-dependent [1, 5].
- Second, many of the existing systems have a very limited view of the human role as merely a last-pass approval step instead of a human-in-the-loop, involved throughout the reasoning process [5].
- Third, data and knowledge curation for human-centred systems, and especially gathering high quality human feedback for ongoing learning, is still expensive and not scalable [1].
- Fourth, when moving from conceptual or prototype systems to sustained operational deployment, it is often overlooked in the research literature that there will be organizational issues related to workflow integration, training and accountability of staff [1].
- Last but not least, as trustworthy AI [13] is designed to meet regulatory and ethical standards, it is important to have continuous interaction between technical designers, domain experts, and policy stakeholders, which is not always reflected in current practices in the development of AI.

10. Future Research Directions

There are a number of directions that are promising in the field of intelligent decision support advancement based on HCAI. Neuro-symbolic approaches, which are able to merge the pattern-recognition ability of learned models with the interpretability and constraint-handling of symbolic reasoning, could provide a way to ensure the verifiability of the domain knowledge behind AI recommendations while maintaining explainability [1]. Reinforcement learning from human feedback provides an approach that continually tunes the system's recommendations to the human judgment and values, which is moving beyond static reward functions and towards true reinforcement learning [1]. A further step in the evolution of the layered architecture introduced in Section 6 is the agentic architectures, in which multiple specialized reasoning and decision components are orchestrated by an orchestration layer, but a human is also allowed to inspect and intervene at critical points [1] and thus is not precluded from being held accountable for the final decisions. Last but not least, adaptive and context-sensitive explanation policies, which adjust the form, timing and depth of explanations to the individual user, task and risk level, offer a promising change from the approach of explaining as a static property to explaining as a dynamically manageable property of the human–AI interaction.

11. Conclusion

In this paper, a theoretical and conceptual discussion of Human-Centered Artificial Intelligence for Intelligent Decision Support has been provided. A conceptual framework with five layers (Data and Knowledge, AI Reasoning, Decision Logic, Human Interaction and Governance and Trust layers) was then suggested as a reference model for designing human-centered decision support architectures, after reviewing the foundational principles of HCAI and the mechanisms of explainability, trust calibration, and human–AI collaboration. The design principles, example application areas, and open questions were discussed, as well as the future directions of research such as: neuro-symbolic reasoning, reinforcement learning from human feedback, and human supervised agentic orchestration. The key takeaway from this synthesis is that technical accuracy does not automatically result in effective or trustworthy decision support, but that trust must be built and earned through the design of systems that are transparent, accountable, and collaborative, and that human oversight is needed to make effective use of AI.

Author Declarations

Author Contributions (CRediT): Harshwardhansinh Chauhan: Conceptualization; Methodology; Investigation; Literature Review; Writing - Original Draft. Rocky Upadhyay: Methodology; Supervision; Validation; Writing - Review & Editing. Sanjay Sharma: Investigation; Formal Analysis; Validation; Writing - Review & Editing. Shivam Upadhyay:

Investigation; Literature Review; Validation; Visualization; Writing - Review & Editing. All authors have read and approved the submitted version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

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

Ethics Statement: Not applicable.

Declaration of Generative AI: During the preparation of this manuscript, the authors used LLM tools for language refinement, grammatical correction, and improvement of readability. The authors subsequently reviewed and verified the resulting content and take full responsibility for the accuracy, integrity, and content of the manuscript.

References

[1] M. Fanfani, L. A. Ipsaro Palesi, and P. Nesi, "Human-Centered AI for Decision Support Systems: A systematic review of application domains, architecture designs, current trends and future directions," *Big Data and Cognitive Computing*, vol. 10, no. 6, Art. no. 186, 2026, doi: 10.3390/bdcc10060186.

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

[3] B. Shneiderman, *Human-Centered AI*. Oxford, U.K.: Oxford University Press, 2022, doi: 10.1093/oso/9780192845290.001.0001.

[4] L. Wang, Z. Zhang, D. Wang, W. Cao, X. Zhou, P. Zhang, J. Liu, X. Fan, and F. Tian, "Human-Centered design and evaluation of AI-empowered clinical decision support systems: A systematic review," *Frontiers in Computer Science*, vol. 5, Art. no. 1187299, 2023, doi: 10.3389/fcomp.2023.1187299.

[5] C. Gomez, S. M. Cho, S. Ke, C.-M. Huang, and M. Unberath, "Human-AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review," *Frontiers in Computer Science*, vol. 6, Art. no. 1521066, 2025, doi: 10.3389/fcomp.2024.1521066.

[6] C. Panigutti, A. Beretta, D. Fadda, F. Giannotti, D. Pedreschi, A. Perotti, and S. Rinzivillo, "Co-design of human-centered, explainable AI for clinical decision support," *ACM Transactions on Interactive Intelligent Systems*, vol. 13, no. 4, Art. no. 21, pp. 1–35, 2023, doi: 10.1145/3587271.

[7] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz, "Guidelines for human-AI interaction," in *Proc. CHI Conf. Human Factors in Computing Systems*, 2019, Art. no. 3, pp. 1–13, doi: 10.1145/3290605.3300233.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?: Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[10] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019, doi: 10.1016/j.artint.2018.07.007.

[11] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017, doi: 10.48550/arXiv.1702.08608.

[12] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[13] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*. Brussels, Belgium: European Commission, 2019, doi: 10.2759/346720.

[14] W. Xu, "Toward human-centered AI: A perspective from human-computer interaction," *Interactions*, vol. 26, no. 4, pp. 42–46, 2019, doi: 10.1145/3328485.

[15] D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) program," *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019, doi: 10.1609/aimag.v40i2.2850.

[16] D. Wang, C. Li, and J. Guo, "Transitioning to Human-Centered AI: A systematic review of theories, scenarios, and hypotheses in human-AI interactions," *Proceedings of the Association for Information Science and Technology*, vol. 61, no. 1, pp. 673–678, 2024, doi: 10.1002/pra2.1078.

[17] A. Klingbeil, C. Grützner, and P. Schreck, "Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI," *Computers in Human Behavior*, vol. 160, Art. no. 108352, 2024, doi: 10.1016/j.chb.2024.108352.

[18] R. Fok and D. S. Weld, "In search of verifiability: Explanations rarely enable complementary performance in AI-advised decision making," *AI Magazine*, vol. 45, no. 3, pp. 317–332, 2024, doi: 10.1002/aaai.12182.

[19] J. Cecil, E. Lermer, M. F. C. Hudecek, J. Sauer, and S. Gaube, "Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task," *Scientific Reports*, vol. 14, Art. no. 9736, 2024, doi: 10.1038/s41598-024-60220-5.

[20] S. Mehrotra, C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman, "A systematic review on fostering appropriate trust in human-AI interaction: Trends, opportunities and challenges," *ACM Journal on Responsible Computing*, vol. 1, no. 4, Art. no. 26, pp. 1–45, 2024, doi: 10.1145/3696449.

[21] C. Holland, G. Perry, and H. F. Neyedli, "Calibrating trust, reliance and dependence in variable-reliability automation," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 68, no. 1, pp. 604–610, 2024, doi: 10.1177/10711813241277531.

[22] M. Kumar and D. Singh, "Balancing autonomy and oversight in reliable agentic artificial intelligence through adaptive human interaction architectures," *Discover Artificial Intelligence*, vol. 6, Art. no. 709, 2026, doi: 10.1007/s44163-026-01373-2.

[23] E. Tabassi, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2023, doi: 10.6028/NIST.AI.100-1.

[24] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*. Paris, France: United Nations Educational, Scientific and Cultural Organization, adopted Nov. 23, 2021.