

Hybrid AI Decision Support in Healthcare: Insights into Usability and Trust

Olivier BLANSON HENKEMANS^{a,1} Bodhi ALLARD^a, Liesbeth VAN DEN BERG^b
^aTNO, Department of Child health

^bAmsterdam UMC, University of Amsterdam, Department of Medical Informatics

ORCID ID: Olivier Blanson Henkemans <https://orcid.org/0000-0002-8931-9207>, Bodhi Allard, <https://orcid.org/0009-0004-5316-492X>, Liesbeth van den Berg <https://orcid.org/0000-0001-6832-4069>

Abstract. Decision Support Systems (DSS) increasingly integrate Artificial Intelligence (AI) to enhance clinical reasoning, yet adoption depends on professionals' trust and usability. This study explored how hybrid AI decision support, combining rule-based guideline logic with generative reasoning, can be designed and evaluated in practice. Using a human-centred design approach within preventive child healthcare, a prototype integrating Large Language Models (LLMs) into an existing guideline-based DSS was developed and tested. Usability and trust were evaluated with nine healthcare professionals through task observations, usability scales, and interviews. The system showed good usability and moderate trust; highest scores were for traceability, workflow integration, and autonomy, while confidence indicators and contextual reasoning needed improvement. The study demonstrates the feasibility and acceptability of hybrid AI DSS in child healthcare and provides initial design requirements for usable, trustworthy, and human-centred AI integration across clinical domains.

Keywords. Clinical Decision Support Systems, hybrid AI, human-centred design (HCD), usability, trust

1. Introduction

Decision support systems (DSS) are increasingly used in healthcare to support consistent, transparent, and evidence-based decision-making. In preventive child healthcare (PCH), professionals make complex, multi-factorial decisions that combine guideline knowledge with growth and developmental assessments and parental narratives. These decisions are often made under time pressure and uncertainty, making effective decision support particularly relevant in this domain.

In the Netherlands, approximately 50% of PCH organizations (serving more than a million of children) use the Smart Guideline Module (SRM) as DSS.[1] It translates guideline logic into structured decision rules. The SRM is part of the I-JGZ platform, a modular digital infrastructure designed to connect and support different health systems. The platform combines shared standards with local flexibility, making it easier to scale up, exchange data securely, and adapt to regional needs. It provides a reliable foundation for integrating decision-support tools, such as the SRM, directly within EHRs.

¹ Corresponding Author: Olivier Blanson Henkemans, TNO Child Health, Sylviusweg 71, 2333 BE, Leiden, the Netherlands. Olivier.BlansonHenkemans@TNO.nl

Like most rule-based DSS, the SRM promotes standardization and guideline adherence, but it struggles to interpret unstructured and narrative information dominating daily practice, such as free-text notes, contextual observations, and parental concerns.

Recent advances in Artificial Intelligence (AI), particularly Large Language Models (LLMs), offer new possibilities for interpreting unstructured clinical information and navigating complex guideline texts. LLMs have demonstrated strong alignment with clinical guidelines, but also raise concerns related to transparency, reliability, and professional autonomy, especially in domains like PCH, where contextual interpretation and uncertainty are inherent to care.[2]

The aim of this study is to investigate how a hybrid AI decision support system can be designed and evaluated to support usability and trust among PCH professionals. Using the SRM as a representative case, we apply a human-centred design approach to derive design insights for usable and trustworthy AI-supported decision support in healthcare.

2. Design and Evaluation of a Hybrid AI Decision-Support System

Hybrid AI decision support systems (hybrid AI-DSS) build on data-driven infrastructures and the principle of hybrid intelligence, which combines the analytical consistency of AI with professionals' contextual, ethical, and empathetic judgment.[3] Studies show that LLM-based DSS align well with clinical guidelines, but diverge when local context or intuition is required, highlighting the need for human-in-the-loop systems that ensure traceable reasoning and professional accountability.[2-5] Compared to traditional rule-based DSS, the hybrid design combines deterministic guideline logic with generative, interactive reasoning.[6]

The design process was started with a narrative review, which identified *trust* and *usability* as predominant factors for the use and adoption of LLM-supported decision support systems, and therefore guided the focus of the subsequent human-centred design and evaluation.[7] Trust in AI-driven decision support depends on transparency and explainability. Professionals must understand how outputs are generated and be able to verify their sources. Explainable-AI techniques that provide selective, referenced explanations help build confidence and accountability while supporting responsible implementation.[8–10] As a core principle of human-centred design (HCD), usability ensures that AI complements professional workflows rather than disrupts them. Yet, few AI-DSS projects apply HCD systematically, despite evidence that co-development with users enhances acceptance and adoption.[7–9]

2.1. Human Centred Design of a Hybrid AI-DSS

To understand user needs for AI-based decision support, we applied a human-centred design process to develop a hybrid AI-DSS prototype for PCH. Physicians, nurses, and domain experts participated in a survey (N = 153) and co-design workshops (N = 6), identifying key barriers such as handling narrative input, interpretation differences, and time spent locating guideline information. These insights were consolidated into 15 user requirements, as listed in Table 1, covering functional, interaction, explainability, personalization, and professional–ethical aspects, which directly informed the design of the AI-enhanced SRM.

The resulting prototype combines deterministic guideline logic with contextual analysis of unstructured input using retrieval-augmented generation (RAG). Integrated

into an electronic health record EHR, the SRM presents structured recommendations alongside concise, referenced AI-generated summaries and supports follow-up questions via a chat interface (Figure 1). When AI-generated advice diverges from guideline logic, the rule-based output remains leading and differences are explicitly shown, ensuring transparency and preserving professional judgment.

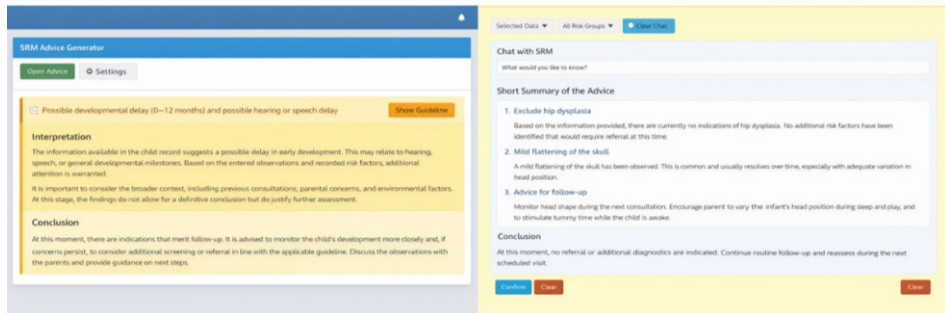


Figure 1. Hybrid AI Decision Support System interface, with LLM-generate advice, based on unstructured EHR-data and deterministic guidelines logic (Left), and chat-function for introducing additional contextual information and posing follow up questions (Right).

2.2. Evaluation

The evaluation examined how PCH professionals perceive the usability and trustworthiness of a hybrid AI-DSS, as these dimensions emerged as key determinants from the narrative review. We tested a prototype of the SRM, enhanced LLM technology, that combines rule-based guideline logic with generative reasoning to assist professionals. Nine PCH professionals (six physicians, three nurses) completed realistic case-based tasks (e.g., hip dysplasia, skin conditions, growth monitoring) in a simulated consultation setting.[2] A mixed-methods design was used: *usability* was assessed with the Healthcare System Usability Scale (HSUS) and *trust* with the AI Attitude Scale (TAI). [11–12] Observations, think-aloud sessions, and interviews provided qualitative insights into reliability, explainability, and autonomy, which were integrated with quantitative results to assess alignment with professional reasoning.

3. Results: Usability and Trust

Overall usability of the LLM-supported SRM-prototype was good to very good (meanHSUS=3.6/5≈usability score of 72%). Participants found the interface intuitive, familiar, and well aligned with their workflow. The inclusion of AI-generated summaries directly within the SRM view reduced the need to consult lengthy guideline texts and improved efficiency (meanHSUS item “Work Effectiveness”=3.4/5). Task-completion rates exceeded 85%, suggesting that the prototype could be used smoothly. The highest ratings were given for workflow integration, side-by-side display of AI and rule-based outputs, and traceability to guideline logic.

Trust in the AI-system’s output was moderate but positive (meanTAI=3.3/5). Factors contributing to confidence included explicit references to guideline sections, alignment with the professional’s reasoning, and concise, readable explanations. Trust decreased when contextual details, such as prior visits or regional referral rules, were

missing, or when statements were overly general. Traceability emerged as the strongest determinant of trust: direct guideline citations increased perceived reliability, while non-referenced text was treated with caution. Participants emphasized that AI should “advise, not decide”, valuing transparent reasoning and control over editable outputs. Confidence indicators and contextual sensitivity were identified as priorities for improvement.

Table 1. Requirements Derived from the Co-Design Phase, Rational and Evaluation

Category / Requirement	Rationale and Evaluation Summary
<i>Functional</i>	
1. Integration with existing workflow	Professionals requested minimal workflow disruption; the interface was intuitive and well aligned with existing systems.
2. Consistent linkage to guideline logic	Traceability to official guidelines was essential for accountability; users valued embedded references that increased reliability.
3. Automatic retrieval of relevant text fragments	Quick access to guideline paragraphs improved efficiency by reducing manual searches.
4. Support for unstructured input	The system partially supported free-text entries such as parental concerns, though contextual interpretation remained limited.
5. Display of both AI- and rule-based outputs	Side-by-side presentation enabled verification and learning, allowing users to compare deterministic and generative reasoning.
<i>Interaction & Explainability</i>	
6. Transparent reasoning	Professionals emphasised clear, “visible” reasoning; traceability of evidence sources was critical for trust.
7. Confidence indicators and limits of AI reasoning	Users wanted clarity about the system’s certainty and contextual limitations; absence of explicit confidence cues reduced trust.
8. Selective explanations on demand	Participants preferred concise advice with optional detail to balance efficiency and understanding.
9. Editable, confirmable recommendations	Editable outputs reinforced autonomy and responsibility, allowing users to control final decisions.
10. Highlight differences of AI and guideline logic	Users wanted deviations clearly marked; differences were not always explicit and require improvement.
<i>Professionalism & Ethics</i>	
11. Preserve professional autonomy	The “advise, not decide” principle was upheld; users retained control over all recommendations.
12. Support learning and reflection	The system was viewed as potentially helpful for reflective practice but not yet realised in use.
13. Ensure data privacy and sovereignty	Privacy-by-design processing within secure environments ensured full compliance with governance requirements.
14. Avoid bias and regional mismatch	Alignment with regional policies was recognised as necessary but not yet implemented.
15. Enable feedback loops for improvement	Professionals wanted their corrections to inform model refinement; such feedback channels are a key next step.

4. Discussion and conclusion

This study examined how hybrid AI-DSS, combining rule-based guideline logic with generative LLM reasoning, can be designed and evaluated to support professional reasoning and trust in preventive child healthcare. Using a human-centred design process, user needs related to workflow integration, explainability, and autonomy were translated into a structured set of functional and interaction requirements that directly informed system design.

The evaluation showed that the hybrid AI-DSS is technically feasible and professionally acceptable when embedded in existing EHRs. Usability was rated high, reflecting effective implementation of requirements related to workflow fit, efficiency, and concise information presentation. Trust was moderate but positive and more sensitive to incomplete requirement coverage. Trust depended strongly on traceability: explicit guideline references, side-by-side presentation of rule-based and AI-generated outputs, and editable recommendations reinforced accountability and professional control. Requirements related to confidence indicators, contextual sensitivity, and feedback mechanisms were only partially realised, helping to explain lower trust scores.

These findings show that usability and trust emerge from how specific design requirements are implemented. Explainability should operate at the interface level, supporting professional judgment under uncertainty rather than relying on opaque model-level explanations. Thus, five priorities emerge: linking AI outputs to identifiable sources, embedding decision support in workflows, incorporating contextual factors, enabling user feedback, and positioning hybrid AI as a learning and reflection tool.

Limitations include the small sample and simulated rather than live consultations. Future research should use real-world data under privacy-preserving conditions and test adaptive feedback mechanisms for continuous co-learning.

In conclusion, the study demonstrates that hybrid AI-DSS can be designed to meet professional needs and strengthen guideline-based reasoning without losing autonomy. When developed for transparency, traceability, and human control, such systems offer a path toward responsible, context-aware human-AI collaboration in healthcare.

References

- [1] Huizing AHJ, Eekhout I, van Buuren S, Blanson Henkemans OA. Data-driven healthcare innovations in a fragmented healthcare system: a modular approach. *Stud Health Technol Inform.* 2025;327:328–332.
- [2] Moeskops I, van den Broek T, Huizing A, Blanson Henkemans OA. AI-beslisondersteuning in de zorg: De waarde van Large Language Models voor jeugdartsen: Een verkenning. *Tijdschr Hum Factors.* 2025;50(1).
- [3] Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. *BISE.* 2021;63(5):513–523.
- [4] Gaber F, Shaik M, Allegra F, et al. Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. *NPJ Digit Med.* 2025;8:263. doi:10.1038/s41746-025-01684-1
- [5] Knop M, Weber S, Mueller M, Niehaves B. Human factors and technological characteristics influencing interaction with AI-enabled CDSS. *JMIR Hum Factors.* 2022;9(1):e.
- [6] Domingues NS. A hybrid decision support system using rule-based and AI methods: the OnCATs knowledge-based framework. *Int J Med Inform.* 2026;206:106144.
- [7] Allard BM. User-centered design of an LLM-supported decision support system for clinical guidelines [master's thesis]. Amsterdam: Faculty of Medicine, University of Amsterdam; 2025.
- [8] Wang L, Zhang Z, Wang D, et al. Human-centered design and evaluation of AI-empowered clinical decision support systems: A systematic review. *Front Comput Sci.* 2023;5:1187299.
- [9] Bienefeld N, Boss JM, Lüthy R, et al. Solving the explainable-AI conundrum by bridging clinicians' needs and developers' goals. *NPJ Digit Med.* 2023;6(1):1–7.
- [10] He X, Zheng X, Ding H, Liu Y, Zhu H. AI-CDSS design guidelines and practice verification. *Int J Hum Comput Interact.* 2023;39(12):2231–2246.
- [11] Fjeld J, Achten N, Hilligoss H, Nagy AC, Srikumar M. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication. 2020;2020(1).
- [12] Ghorayeb A, Darbyshire JL, Wronikowska MW, Watkinson PJ. Design and validation of a new Healthcare Systems Usability Scale (HSUS) for clinical decision support systems: a mixed-methods approach. *BMJ Open.* 2023;13(1):e066972.
- [13] Scharowski N, Perrig SA, Aeschbach LF, von Felten N, Opwis K, Wintersberger P, Brühlmann F. To trust or distrust trust measures: Validating questionnaires for trust in AI. *arXiv [Preprint].* 2024.