Year 2 Study - Agentic AI-Powered CliniKiosk: Expanding Healthcare Access with Explainable and Empathetic AI

Problem/Question

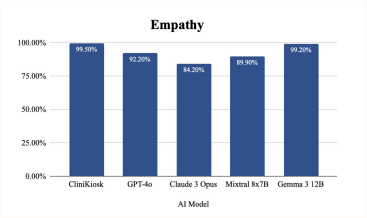
- Healthcare access gaps are worsening (workforce shortages, hospital closures) for rural communities
- Trust gaps exist in AI healthcare tools – especially low empathy, unclear reasoning/transparency, and language quality
- Research question: Can a multilingual, agentic, retrieval-augmented digital CliniKiosk deliver accurate, empathetic, explainable, and linguistically rich health recommendations?

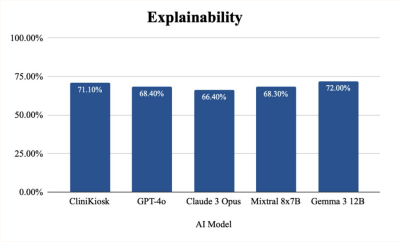
Framework

- System: React/TypeScript, Supabase, agentic workflow (router → planner → executor → manager), Whisper-1 voice input, i18next multilingual output
- Evidence grounding: RAG pipeline retrieving from PubMed, CrossRef, ClinicalTrials.gov, AHA/ACP/NICE
- Evaluation dataset: 500 clinician-verified simulated cases with gold-standard
- CliniKiosk was benchmarked against GPT-4o, Claude 3 Opus, Mixtral 8×7B, and Gemma 3 12B

Findings

- CliniKiosk delivered triage and treatment recommendations on par with best performing open AI models and achieved the highest empathy scores among all systems evaluated and second highest explainability





Conclusions

- A multilingual, RAG-based, agentic CliniKiosk can deliver empathetic, evidence-based health recommendations and triage decisions with accuracy comparable to leading LLMs
- CliniKiosk prioritizes human-centered communication, trust, and transparency

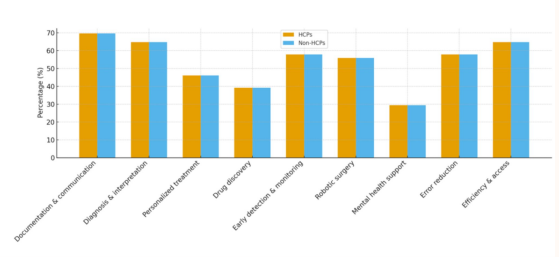
Background Information

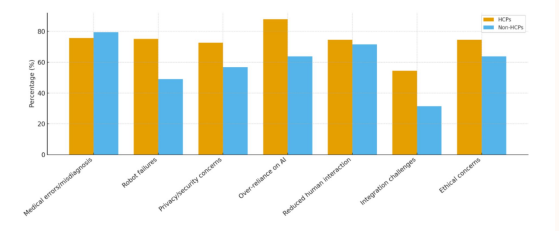
Access gaps are driven by workforce shortages, facility closures, and prolonged wait times; rural and marginalized communities are disproportionately affected.

Stakeholder survey (n=297) showed broad optimism about AI, but substantial ethical and bias concerns.

Trust in AI to make fair and ethical healthcare decisions was low (≈65% of HCPs and 60% of non-HCPs reported little to no trust).

Design goal: evidence-grounded responses that are multilingual, empathic, and transparent.





Problem and Predictions

From trust gaps to human-centered AI design targets

Problem

- Most health chatbots are rule-based, minimally personalized, limited in multilingual support, and often lack empathy and transparent reasoning.
- Trust-related gaps (empathy, explainability, language quality) can limit real-world adoption.

Research Question

- Can a multilingual, agentic, retrieval-augmented digital CliniKiosk deliver accurate, empathetic, explainable, and linguistically rich health recommendations?

Predictions

- Stakeholder-informed, agentic + RAG design will raise trust-related performance (higher empathy and explainability) while maintaining competitive accuracy on safety-relevant tasks (triage, differentials).

Methodology

Study Design (2 phases)

Phase 1: Build a modular system architecture with a React/TypeScript frontend and a Supabase/PostgreSQL backend with row-level security

Phase 2: Build CliniKiosk (Agentic AI workflow + multilingual UI + RAG evidence pipeline).

Audio input was implemented through Whisper-1, and multilingual output was supported via i18next.

Md. Abdur Rouf, Vecteezy

Agentic Architecture

Methodology

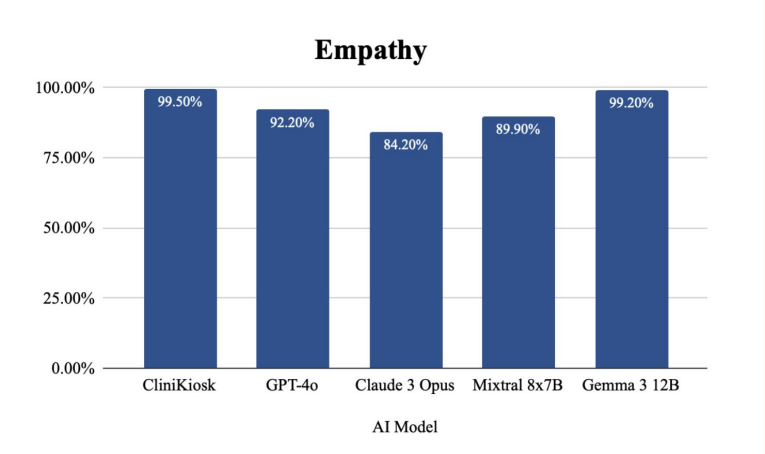
RAG sources: PubMed, CrossRef, ClinicalTrials.gov, plus AHA/ACP/NICE guidance (citation-linked responses).

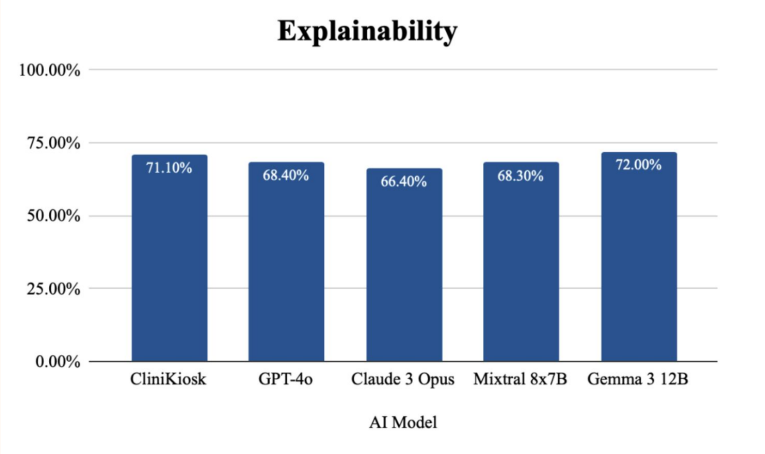
Clinician-Verified Dataset: 500 simulated cases spanning demographics, acuity, organ systems; gold standards set by a board-certified physician. CliniKiosk was benchmarked against GPT-4o, Claude 3 Opus, Mixtral 8×7B, and Gemma 3 12B.

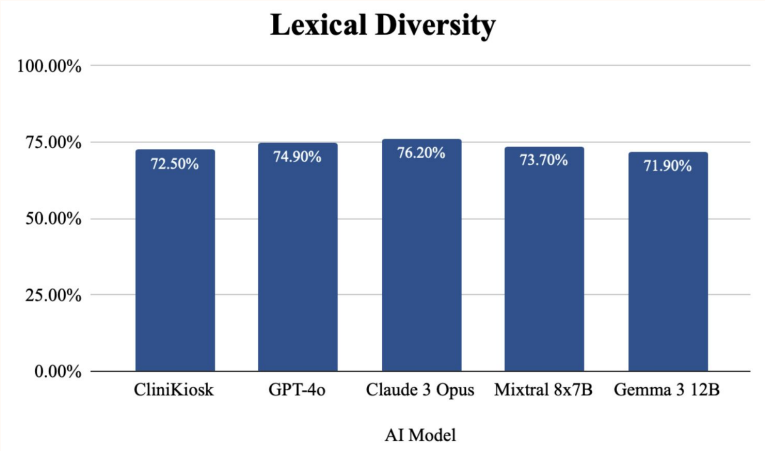
Metrics: triage, differential diagnosis, final diagnosis, treatment accuracy; empathy (ECCS), explainability (SCS), lexical diversity (MTLD).

Statistics: χ^2 + pairwise z-tests (Holm); ANOVA/Welch t-tests (Holm).

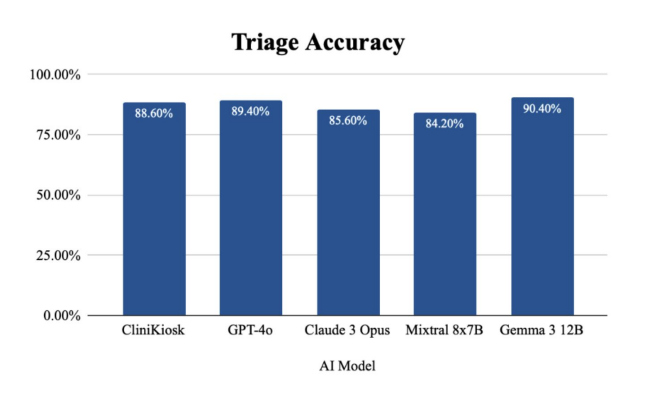
Results

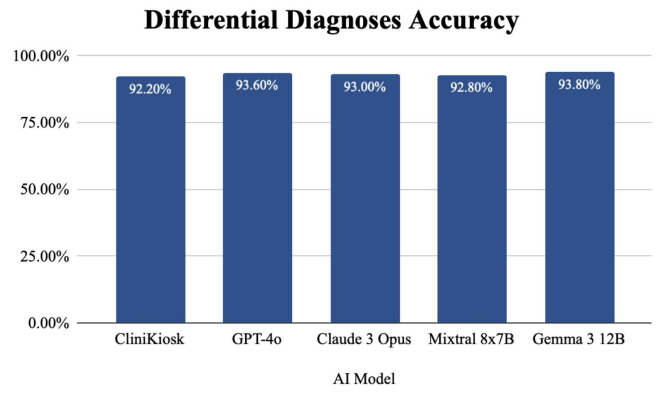

Empathy
100.00%
75.00%
50.00%
25.00%
0.00%
99.50%
92.20%
84.20%
89.90%
99.20%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model

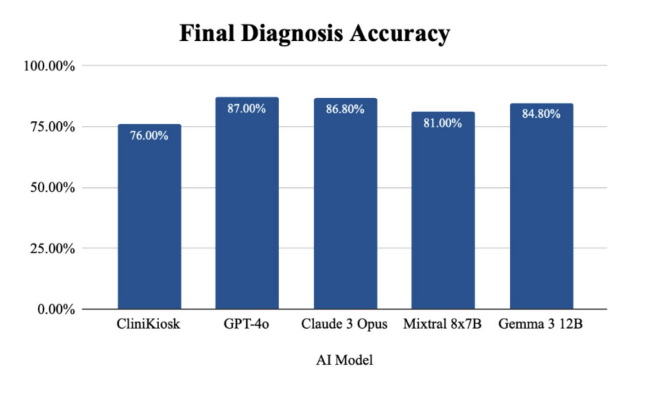

Explainability
100.00%
75.00%
50.00%
25.00%
0.00%
71.10%
68.40%
66.40%
68.30%
72.00%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model

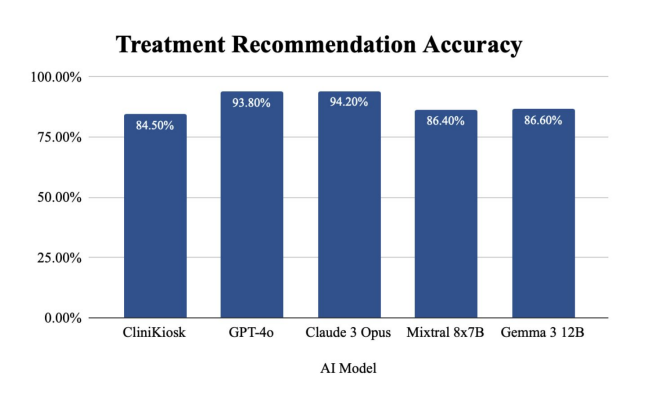

Lexical Diversity
100.00%
75.00%
50.00%
25.00%
0.00%
72.50%
74.90%
76.20%
73.70%
71.90%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model

Results


Triage Accuracy
100.00%
75.00%
50.00%
25.00%
0.00%
88.60%
89.40%
85.60%
84.20%
90.40%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model


Differential Diagnoses Accuracy
100.00%
75.00%
50.00%
25.00%
0.00%
92.20%
93.60%
93.00%
92.80%
93.80%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model


Final Diagnosis Accuracy
100.00%
75.00%
50.00%
25.00%
0.00%
76.00%
87.00%
86.80%
81.00%
84.80%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model


Treatment Recommendation Accuracy
100.00%
75.00%
50.00%
25.00%
0.00%
84.50%
93.80%
94.20%
86.40%
86.60%
CliniKiosk
GPT-4o
Claude 3 Opus
Mixtral 8x7B
Gemma 3 12B
AI Model

Discussion and Interpretation

Differential Diagnosis accuracy was uniformly high (~92–94%); CliniKiosk led empathy and explainability, while final-diagnosis accuracy was lower than general-purpose LLMs.

CliniKiosk approached general-purpose LLMs on triage and treatment recommendations, while achieving the highest empathy scores.

RAG-grounded responses likely supported clearer, more structured explanations (explainability advantage).

Lower final-diagnosis accuracy (76%) may reflect the difficulty of single-label diagnosis from text-only vignettes and conservative safety thresholds.

For kiosks, safe triage pathways and strong differentials may be more clinically meaningful than definitive diagnosis.

Slightly lower lexical diversity may improve comprehension for users with lower health literacy.

Implications and Ideas for Future Research

Implications

- Patient-centered kiosks could provide on-demand, culturally aligned guidance in multiple languages in community settings (e.g., pharmacies, libraries, workplaces, schools).
- High empathy + explainability may enhance trust, usability, and adherence – key for underserved populations.

Limitations

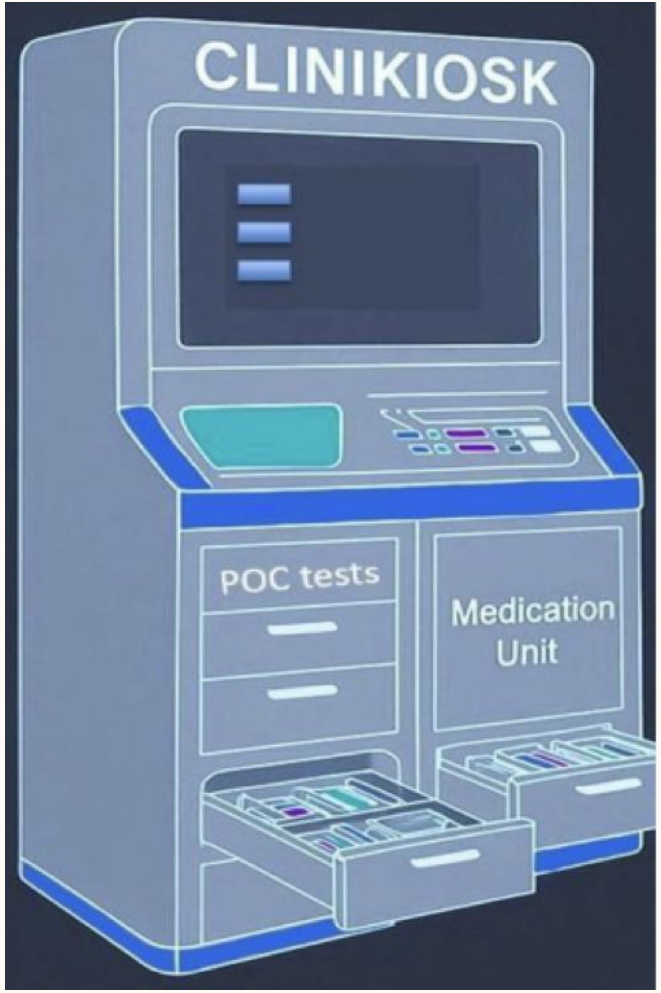
- Simulated cases may not capture the full complexity of real patient presentations.
- Controlled testing may not reflect noisy inputs, clinical workflow constraints, or real-world bias patterns.

Next Steps

- Real-world implementation studies with monitoring for safety and bias.
- Integrate point-of-care diagnostics and (where appropriate) medication dispensing to reduce uncertainty.
- Connect to telehealth/EHR workflows; evaluate user trust, interpretability, and clinician oversight models.

Conclusions

Our findings demonstrate that a multilingual, RAG‑based, agentic CliniKiosk can deliver empathetic, evidence‑based health recommendations and triage decisions with accuracy comparable to leading LLMs. CliniKiosk demonstrated superior empathy and explainability while using more accessible language, aligning human-centered communication with trust and transparency.


CLINIKIOSK
POC tests
Medication Unit

References

1. World Health Organization; World Bank. Tracking Universal Health Coverage: 2023 Global Monitoring Report. Geneva, Switzerland: World Health Organization; 2023.
2. Akinyemi JO. Healthcare access disparities among marginalized communities. Glob Public Health Manag Netw. 2024;2(1):44-59.
3. Wen H, Sadeghi N. Racial disparities in access to health care infrastructure across US counties. Front Public Health. 2023;11:897007.
4. Rural Health Redesign Center. 20 States With the Most Rural Hospital Closures. Pennsylvania Office of Rural Health; 2024. Accessed November 29, 2025. https://www.porh.psu.edu/20-states-with-the-most-rural-hospital-closures/
5. 705 hospitals at risk of closure, state by state. Becker's Hospital CFO Report. Accessed November 29, 2025. https://www.beckershospitalreview.com/finance/705-hospitals-at-risk-of-closure-state-by-state-november-2024.html.
6. United States Department of Agriculture Economic Research Service. 146 Rural Hospitals Closed or Stopped Providing Inpatient Services Between 2005 and 2023. USDA ERS; 2025. Accessed November 29, 2025. http://ers.usda.gov/data-products/charts-of-note/chart-detail?chartId=110927
7. Chartis. 2025 Rural Health – State of the State: Rural Hospital Closures & Care-Access Crisis. Chartis; 2025. Accessed November 29, 2025. https://www.chartis.com/insights/2025-rural-health-state-state
8. Garcia MC, Faul M, Massetti G, et al. Premature and preventable deaths in rural versus urban counties in the United States. MMWR Surveill Summ. 2024;73(SS-x):1-12.
9. Centers for Disease Control and Prevention. Rural Americans Are at Higher Risk of Early Death Than Urbanites. CDC; 2024. Accessed November 29, 2025. https://abcnews.go.com/Health/rural-americans-higher-risk-early-death-urbanites-cdc/story?id=109742216
10. Rural Health Information Hub. Rural Health Disparities. Rural Health Information Hub; 2025. Accessed November 29, 2025. https://www.ruralhealthinfo.org/topics/rural-health-disparities
11. AMN Healthcare. 2025 Survey of Physician Appointment Wait Times. AMN Healthcare; 2025. Accessed November 29, 2025.
12. Association of American Medical Colleges. Walk-in clinics tackle urgent mental health needs. AAMC News. August 21, 2024. Accessed November 29, 2025.
13. Derlet RW, Richards JR. Overcrowding in the nation's emergency departments: complex causes and disturbing effects. Ann Emerg Med. 2000;35(1):63-68.
14. World Health Organization. Health Workforce: Overview and Key Facts. Geneva, Switzerland: World Health Organization; 2020. Accessed November 29, 2025. https://www.who.int/health-topics/health-workforce
15. Mercer. 2021 U.S. Healthcare Labor Market: The Future of Healthcare Labor. New York, NY: Mercer; 2021. Accessed November 29, 2025.
16. American Hospital Association. Data Brief: Health Care Workforce Challenges Threaten Hospitals' Ability to Care for Patients. American Hospital Association; October 2021. Accessed November 29, 2025. https://www.aha.org/fact-sheets/2021-11-01-data-brief-health-care-workforce-challenges-threaten-hospitals-ability-care
17. Popa V, Geissler J, Vermeulen R, Priest E, Capperella K, Susuzlu G, Terry SF, Brooke N. Delivering Digital Health Solutions that Patients Need: A Call to Action. Ther Innov Regul Sci. 2024 Mar;58(2):236-241.
18. Sun G, Zhou X. AI in healthcare: navigating opportunities and challenges with chatbots for digital health. Front Public Health. 2023;11:1307636.
19. Rebelo F, et al. Chatbots in healthcare: A systematic literature review and directions for future research. Comput Biol Med. 2022;150:106081.
20. Abd-Alrazaq AA, et al. Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review. J Med Internet Res. 2020;22(12):e19127.
21. Laymouna M, et al. Roles, Users, Benefits, and Limitations of Chatbots in Healthcare: A Rapid Review. J Med Internet Res. 2024;26(7):e484931.
22. Jin E, Ryoo Y, Kim WJ, Song YG. Bridging the health literacy gap through AI chatbot design: the impact of design cues on trust and use intention for low health literacy users. Internet Res. 2024;35(3):1299-1329.
23. Koenig CJ, Young AJ, Stephenson J, et al. Empathic communication in dignity therapy: feasibility of the Empathic Communication Coding System (ECCS). Patient Educ Couns. 2022;105(10):2783-2791.
24. Holzinger A, Müller H, Kieseberg P, et al. The System Causability Scale (SCS): assessing the quality of machine explanations. Artif Intell Med. 2020;103:101792
25. McCarthy PM, Jarvis S. MTLD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. Behav Res Methods. 2010;42(2):381-392.
26. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. Bull Med Libr Assoc. 2011;89(3):291-298.
27. Topol EJ, et al. A systematic review and meta-analysis of empathy in patient care: human clinicians vs generative AI chatbots. Br Med Bull. 2025;156(1):ldaf017