DC4 · TUAI MSCA DOCTORAL NETWORK

Personalized Health Recommendation Systems Powered by LLMs

Summer School SUT Poland

Camilo España
espanajuan@uniovi.es

University of Oviedo
Supervisors: Edward Núñez
July 2026



Universidad de Oviedo

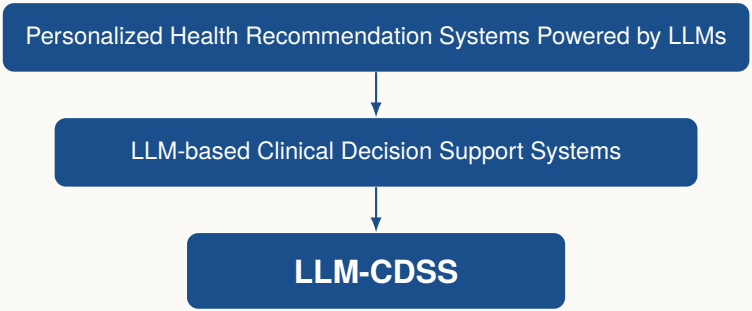


Funded by
the European Union

ROADMAP

Agenda

1. The Research Scope
2. State of the Field
3. Plan to the Gaps
4. First Results
5. Next Steps
6. References


Personalized Health Recommendation Systems Powered by LLMs
LLM-based Clinical Decision Support Systems
LLM-CDSS

1. The Research Scope

1 Where this project sits in TUAI



Medical data are inherently time series, which poses most of the field's theoretical challenges, particularly the need for trustworthiness and explainability

1 The Research Scope

TRANSPARENT

Clinical decisions must be explainable to clinicians and regulatory bodies.

ACCURATE

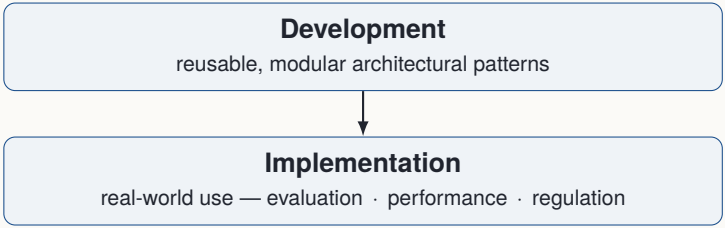
Grounded in medical literature, patient records, and health trends.

AI-ASSISTED

LLMs as the engine for personalised decision support.

Investigate the key techniques, algorithms, and tools needed to **develop** and **implement** highly personalised health recommendation systems that processes time-series patient trajectories.

1 Zooming in on the scope



2. State of the Field

2 A fast-moving field: research

Academia is already building this [1]

- General-purpose: ChatGPT-4, Gemini
- Domain pre-trained: Med-PaLM 2, BioGPT
- RAG & agents: HandRAG, PharmacyGPT
- Specialist agents: OncoGPT, BC-SLM

Systems inventoried

42

LLM-based CDSS [1]

2 A fast-moving field: industry

Big tech is already building this

- Alphabet: AMIE matches physicians [2]
- Microsoft: clinical copilots inside Epic [3]
- OpenAI: HealthBench, 262 physicians [4]
- Amazon: Health AI agent for Prime members [5]

Investment scale

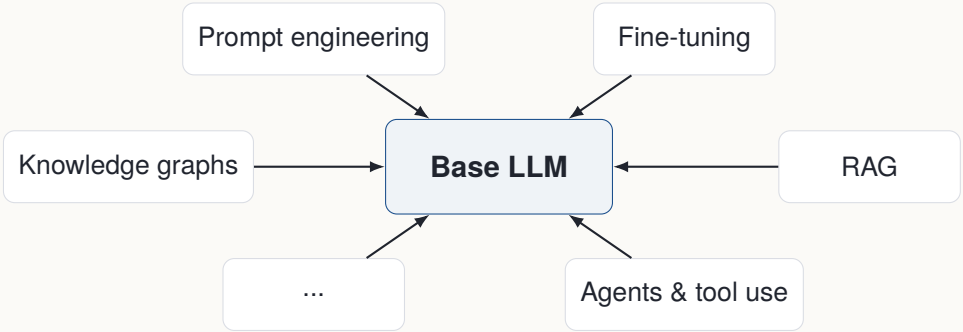
$3.95B

AI-health startups, 2025 [6]

Why is such a highly competitive field still a good research topic?

2 Development

- Patient data as time-series trajectories remains underexplored [1]
- No single system can cover the full universe of clinical decisions [1]
- Each new component extends what a system can do [7]



2 Implementation

United States · FDA

0 / 1,016

AI devices use an LLM, 2024 [8]

European Union · MDR

1

LLM system approved, 2025 [9]

Even if these figures have improved since, implementation will need to keep evolving alongside with development innovations .

2 Why implementation is not happening

Evaluation lags the technology

- Outdated benchmarks [10]
- Not deployment skills [11]

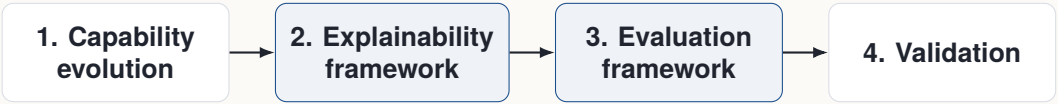
Transparency & explainability

- The unresolved trust barrier
- Required by the **EU AI Act**, effective **Aug 2026** [12]

Trustworthiness is the bottleneck between the tools being built and their safe, real-world use.

3. Plan to the Gaps

3 A four-step research plan

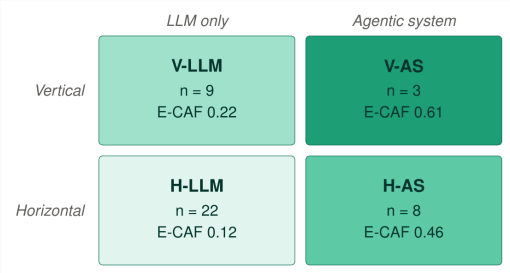


Evaluation and explainability: the two key starting components of trustworthiness.

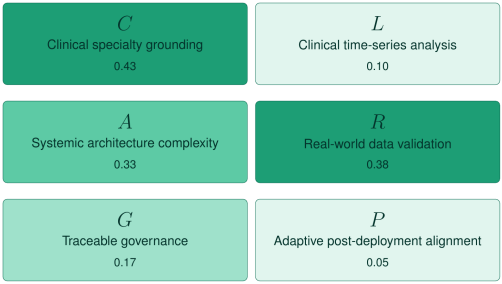
4. First Results

4 Conference paper published (FLICS 2026)

Taxonomy of LLM-CDSS



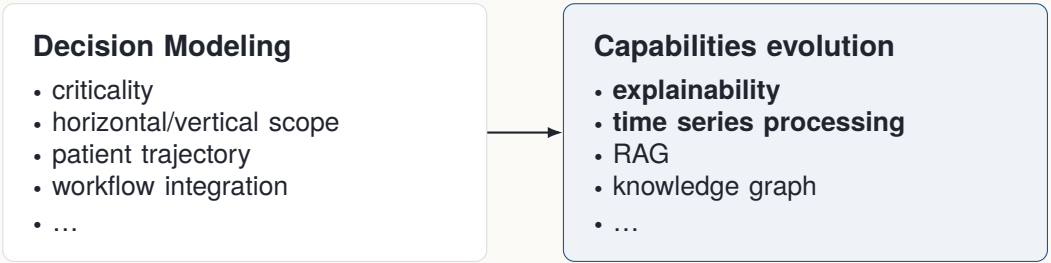
Capability framework



- 42 systems inventoried [1]
- Post-deployment alignment near to zero
- Time series processing second lowest score

4 Scoping review protocol

Registered in Open Science Framework (OSF) website [13].



Mapping how each decision archetype in the literature governs the system built to support it.

5. Next Steps

5 What's next

Scoping review journal paper – in progress

Submission expected after the summer break

Explainability framework

Closing the trust barrier

Evaluation framework

Providing a tool for safe real-world deployment

5 Conclusion

- Development is driven by reusability
- Implementation challenges increase with technology complexity

> *Making **LLM-CDSS** transparent, accurate, and trustworthy enough to actually be used.*
>
> —

6. References

6 References I

[1] Juan C. España-Lopera, Edward R. Núñez-Valdez, and Vicente García-Díaz. “A Technical Taxonomy and Evidence-Based Framework for Assessing Clinical Large Language Model Systems”. In: *IEEE FLICS 2026*. Valencia, Spain, 2026.

[2] Tao Tu et al. “Towards conversational diagnostic artificial intelligence”. In: *Nature* (2025).

[3] Microsoft. *Meet Microsoft Dragon Copilot: Your new AI assistant for clinical workflow*. Microsoft Cloud Blog, March 2025. 2025.

[4] Rahul K. Arora et al. *HealthBench: Evaluating Large Language Models Towards Improved Human Health*. arXiv:2505.08775, OpenAI. 2025.

[5] Amazon. *Amazon launches Health AI agent with free 24/7 access to virtual care for Prime members*. Amazon News, March 2026. 2026.

6 References II

[6] Silicon Valley Bank. *17th Healthcare Investments and Exits Report*. 2025.

[7] Bo Yuan et al. “Agentic Large Language Models for Healthcare: Current Progress and Future Opportunities”. In: *Medicine Advances* (2025).

[8] R. Singh et al. “How AI is used in FDA-authorized medical devices: a taxonomy across 1,016 authorizations”. In: *npj Digital Medicine* 8 (2025), p. 388.

[9] VDE. *Medical co-pilot: First AI-supported medical device to support clinical decisions*. Press release, April 2025. Prof. Valmed, Class IIb CE mark under EU MDR. 2025.

[10] Paul Hager et al. “Evaluation and mitigation of the limitations of large language models in clinical decision-making”. In: *Nature Medicine* 30 (2024), pp. 2613–2622.

6 References III

[11] Shreya Johri et al. “An evaluation framework for clinical use of large language models in patient interaction tasks”. In: *Nature Medicine* (2025).

[12] Panos Vardas et al. “Medicine, healthcare and the AI act: gaps, challenges and future implications”. In: *European Heart Journal – Digital Health* (2025).

[13] Juan C. España-Lopera, Edward R. Núñez-Valdez, and Vicente García-Díaz. *Clinical Decision Characteristics as Governors of LMM-CDSS Architecture: A Scoping Review Protocol*. OSF Registration, DOI: 10.17605/OSF.IO/K7CDY. 2026.

Thank you

Questions welcome



Camilo España

espanajuan@uniovi.es