Delphyr

Engineering Reliability for Clinical Environments

World Summit AI 2025

Michel Abdel Malek MD

Founder/CEO Delphyr

Delphyr

an agentic platform

built for clinical workflows

to reduce time, regain effiency and

improve well being of staff and patients

TO CONTIGO, EN LA AMISTAD

Clinical Administrative Tasks

We started here

Ambient Listening

With Patient

[10%]

[Finding data, Generating Insight Writing notes]

90%

Delphyr

Requirements/Constraints

- Fit in legacy electronic health record systems
- Fit in legacy data models (or lack thereof)
- Fit in legacy hospital infrastructure

- Fit in highly regulated markets
- Fit in data and patient privacy

- Robust, predictable, repeatable outcomes
- Safe and ethically built
- Environment-friendly
- etc

Delphyr

How do you build reliable GenAI systems in given the legacy constraints?

Robust Retrieval & Reduction of Hallucinations

Challenges in working with real world clinical data

Lack of proper metadata

Structured vs Unstructured

Information is context dependent

Contextual terminology

Missing notes

Delphyr

How difficult ~~can it be~~ is it?

SOTA Semantic Retrieval Recall @ 5: 55-65% (EHR Benchmark)

Delphyr

Breaking down the retrieval pipeline

Evals

Expansion - Enrichment - Classification - Indexing

Chunking - Embeddings - Graphs

Query Understanding - Query Routing - Reranking

Delphyr

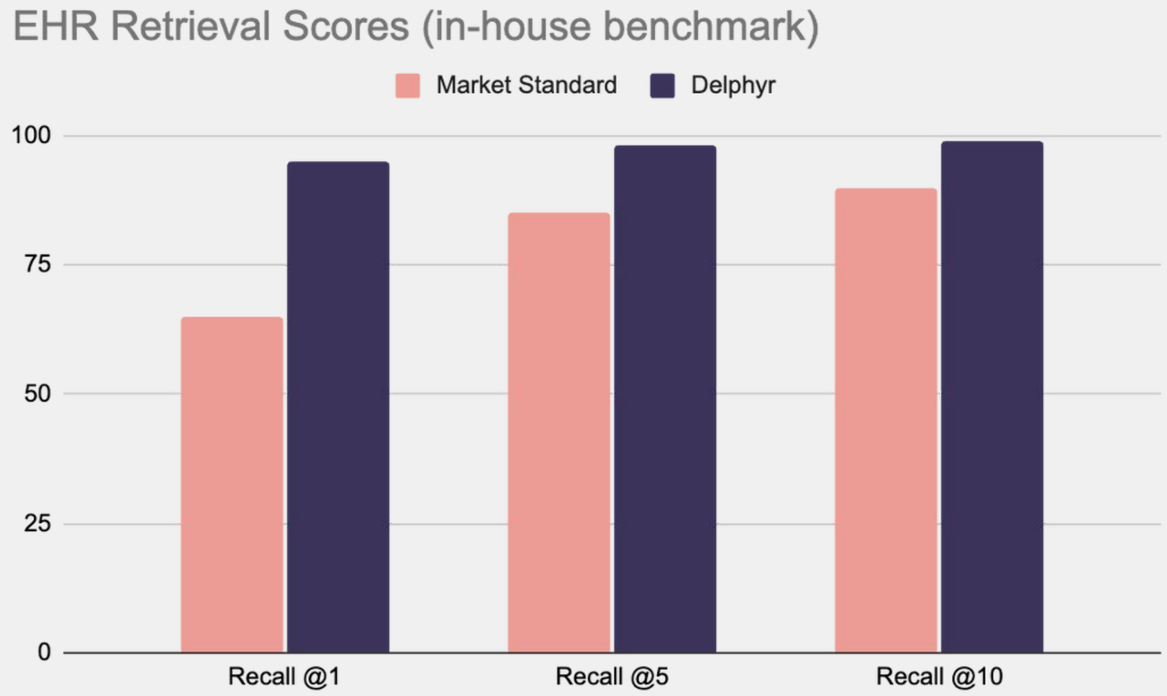
Delphyr's Retrieval Performance

Delphyr's clinical retrieval platform is till date still the best performing retrieval platform for clinical data.

(Recall@1 95%, Recall@5 98%)

Did we overengineer it? Probably.

Is it robust? Yes.



Reducing hallucinations

Hallucinations in Clinical AI

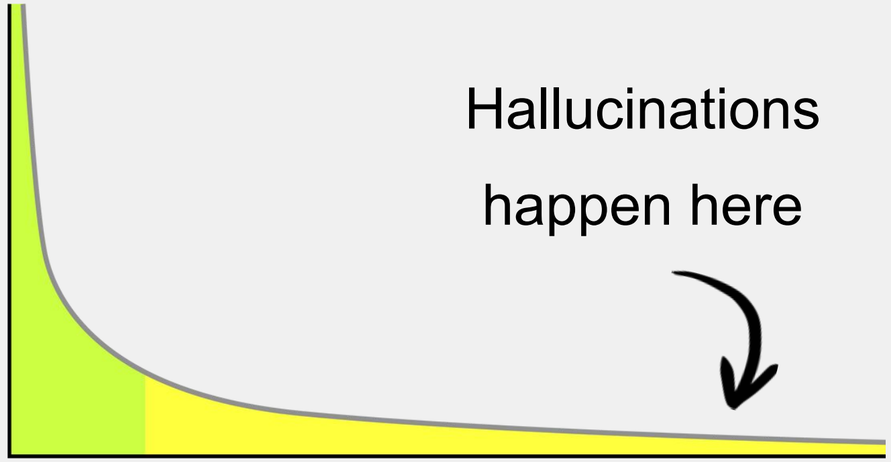
Hallucinations are a fundamental feature/bug of LLMs.

Long-tail occurrences

Compression Failures

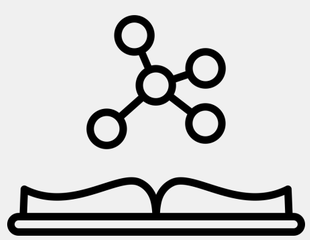
Lack ‘metacognition’

“*Just put the temperature on 0”*



Delphyr

Delphyr mitigates hallucinations at system level



Detection rate 3-6%

Hallucination rate < 1 %.

Thoughts on clinical reasoning

Clinical Reasoning

The cognitive process used to assess and manage a patient’s medical condition with the universal goal to improve outcome for the patient.

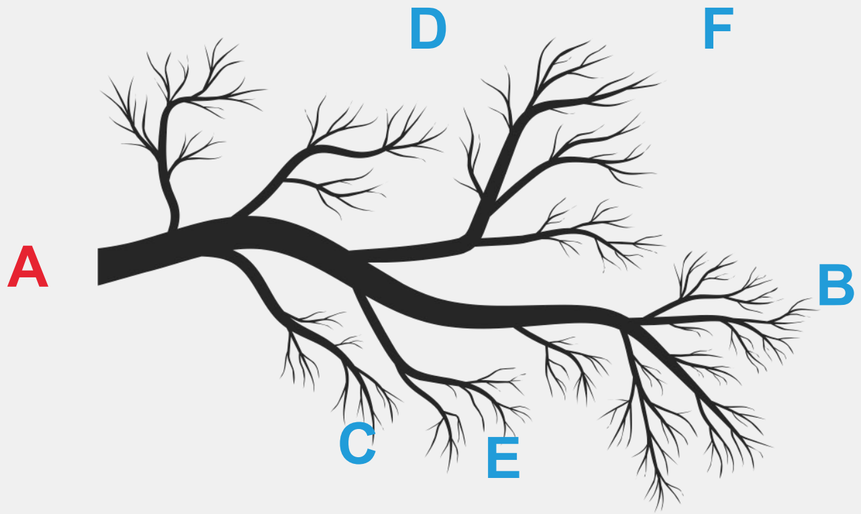
Universal Reward-modeling is not straight forward, so subdivide into Diagnostic, Therapeutic, Procedural

Delphyr

Clinical Reasoning According To:

Engineers

Doctors



Clinicians switch effortlessy between different strategies:

Heuristics, probabilistic, formal, and pathophysiological reasoning.

Delphyr

Does Clinical Reasoning resemble retrieval-augmented generation?

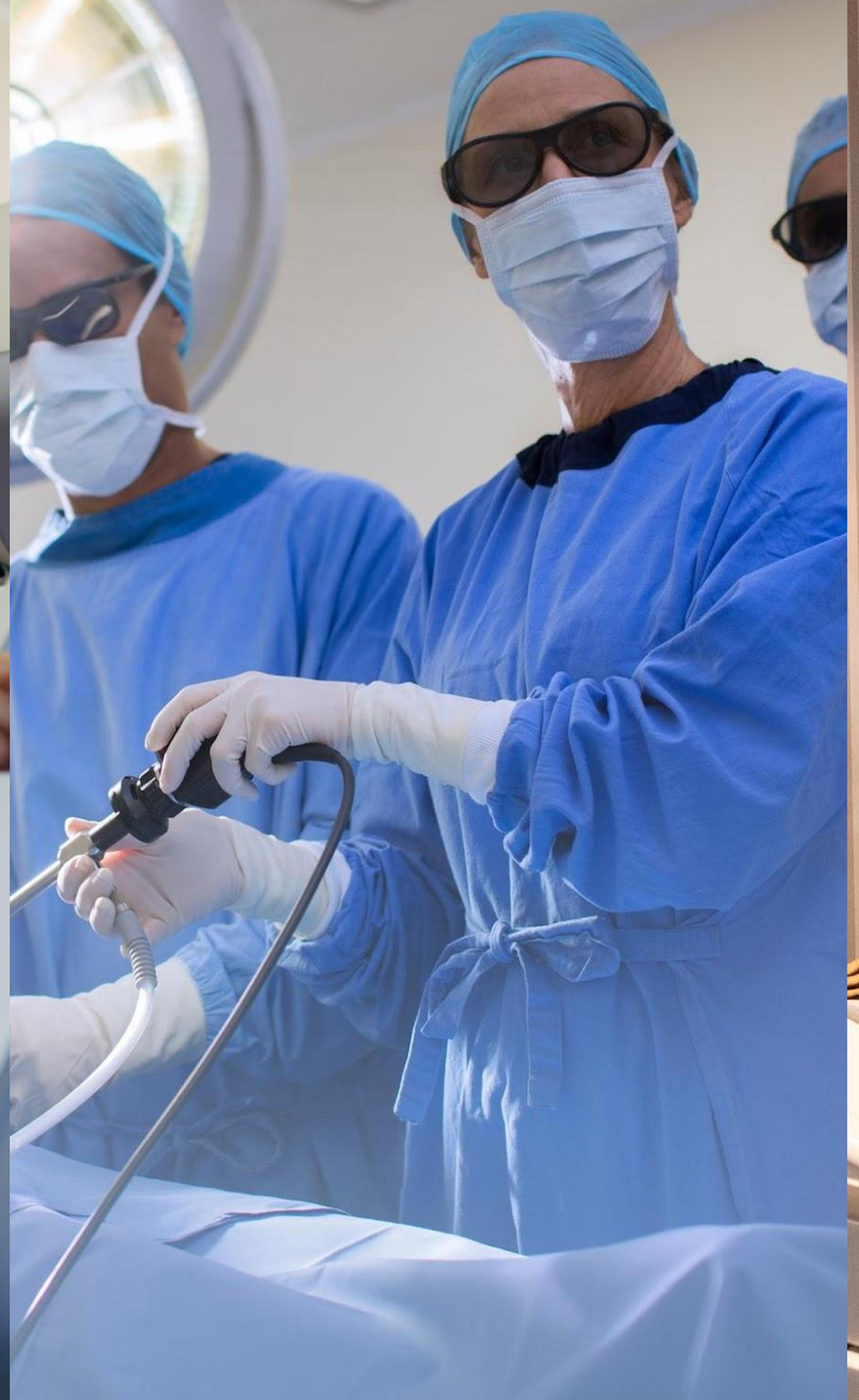
*We now understand much of clinical reasoning as **information processing** and **retrieval,** given the way that expert knowledge is stored in scripts and schemas and activated when given clinical problems.*

Prof. Eric Topol, The Lancet, 2024

Delphyr

Should we train algorithms how to correctly predict a diagnosis,

or how to make decisions?



Delphyr

Q&A

World Summit AI 2025

Michel Abdel Malek MD

Founder/CEO Delphyr

michel@delphyr.ai