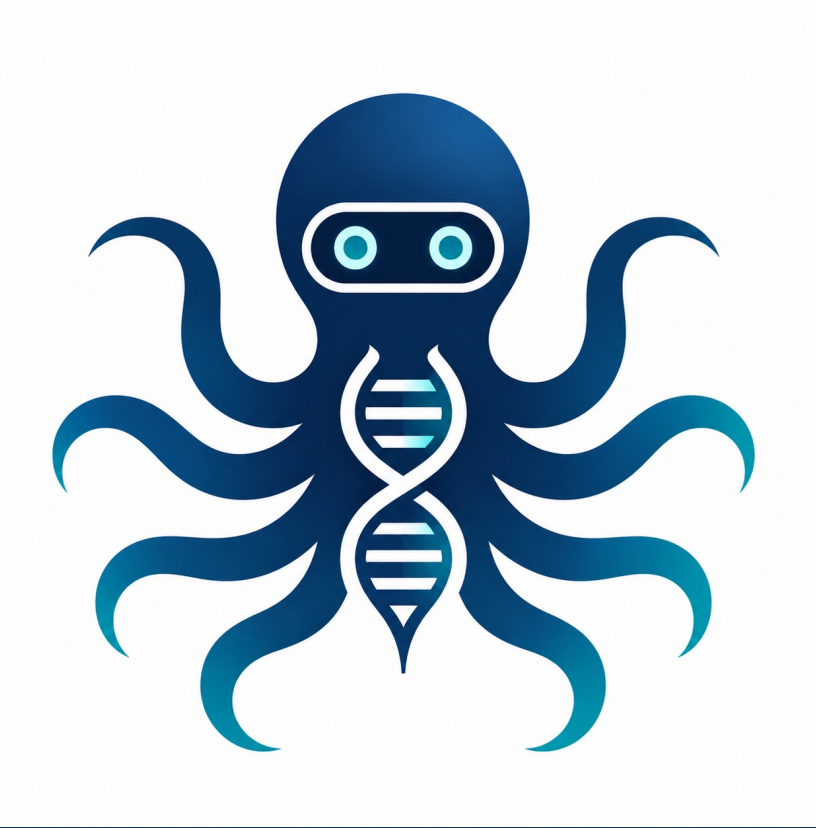


OncoBot – A Safe Large Language Model Guideline Assistant for the Oncology Multidisciplinary Team



Safe, traceable oncology guideline extraction using constrained LLM workflows



Nabeel Ashfaque Sheikh¹, Felix Jozsa², Julia Ive²

¹ Somerset NHS Foundation Trust | ²UCL Institute of Health Informatics

Introduction

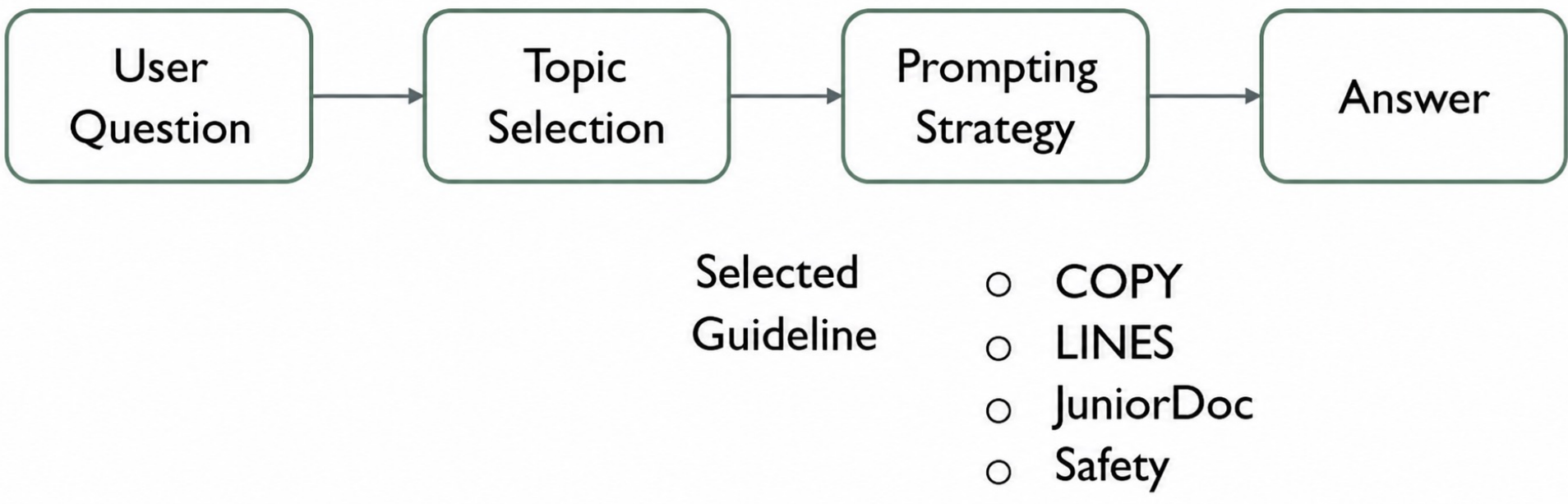
- Oncology teams frequently navigate multiple complex guidelines and protocols.
- LLM systems may hallucinate, omit or reword clinically important guidance.
- We developed OncoBot, a guideline-grounded oncology assistant focused on safe evidence extraction.
- Aim: improve workflow efficiency while preserving fidelity to local NHS guidance.

Methodology

- Safe-LLM extraction framework adapted to local oncology guidelines.
- Six local oncology guidelines from an NHS district general hospital evaluated.
- System designed to return concise verbatim guideline excerpts.
- Multiple prompting strategies tested to minimise hallucination and unsafe rewriting.
- Clinician-led testing performed using realistic oncology queries.

Technical Flow

- Topic-selection identifies the relevant oncology guideline.
- Evaluation used full-guideline and retrieval-constrained contexts.
- LLM extracted evidence rather than generating unrestricted answers.
- Prompting strategies aimed to improve traceability and reduce hallucination risk.



Prompting Strategies

- **COPY** – direct verbatim sentence extraction
 - **LINES** – line-number selection with deterministic reconstruction
 - **JuniorDoc** – draft-answer workflow refined using guideline evidence
 - **Safety** – minimal clinically relevant sentence extraction
- Human clinician evaluation assessed relevance, completeness, and safety.

Full Guideline Context

Method	Recall	NER Recall	Answer Length
COPY	0.53	0.46	41
LINES	0.57	0.48	147
JuniorDoc	0.64	0.54	223
Safety	0.49	0.45	35

Scores range from 0-1; higher values indicate better recovery of relevant guideline information.

JuniorDoc achieved the highest information recovery in long full-context oncology guidelines but produced substantially longer responses.

Retrieval-Constrained Context

Method	Recall	NER Recall	Answer Length
COPY	0.52	0.46	43
LINES	0.65	0.59	117
JuniorDoc	0.65	0.58	142
Safety	0.46	0.41	39

After retrieval-constrained section selection, LINES achieved comparable recovery with stronger traceability and moderate response length.

Recall = relevant guideline content recovered
NER Recall = clinically important entities recovered
Answer Length = average response length (words)

Results

- Performance differed according to structure and retrieval setting.
- In full-context guidelines, JuniorDoc achieved the highest information recovery, suggesting draft-first workflows may benefit long context-heavy documents.
- Following retrieval-constrained section selection, LINES achieved comparable recovery with stronger traceability.
- Findings suggest safe guideline assistants should adapt extraction strategy to document type: JuniorDoc for long full-context guidance; LINES for focused retrieved sections.
- Clinician evaluation of LINES demonstrated high clinical relevance, with no major omissions and no hallucinated guidance identified outside source documents.

Future Direction

- Prospective simulated clinical workflow testing
- Quantitative time-saving evaluation
- Broader oncology guideline integration
- Exploration of NHS deployment pathways

Conclusion

OncoBot demonstrates the potential of safe, guideline-grounded LLM workflows for oncology practice. JuniorDoc performed best in long full-context guidance, while LINES supported traceable extraction after retrieval-constrained section selection. These findings support constrained extraction as a safer design direction for AI-assisted oncology guideline navigation.