

Evaluating and deploying AI to aid outpatient clinic preparation

Helena Langford¹, Nick Kennedy^{2,3,4}, Jane Masoli^{3,4}, Craig Willis⁴ and Royal Devon University Healthcare NHS Foundation Trust outpatient evaluation team³
1. NHS Fellowship in Clinical AI 2. Exeter Healthtech Research Centre 3. Royal Devon University Healthcare NHS Foundation Trust 4. University of Exeter

The problem

It is standard practice for clinicians to review historical notes to familiarise themselves with a patient's background ahead of outpatient appointments. This process can be time-consuming and adds to administrative burden, impacting burnout rates and career satisfaction (1).

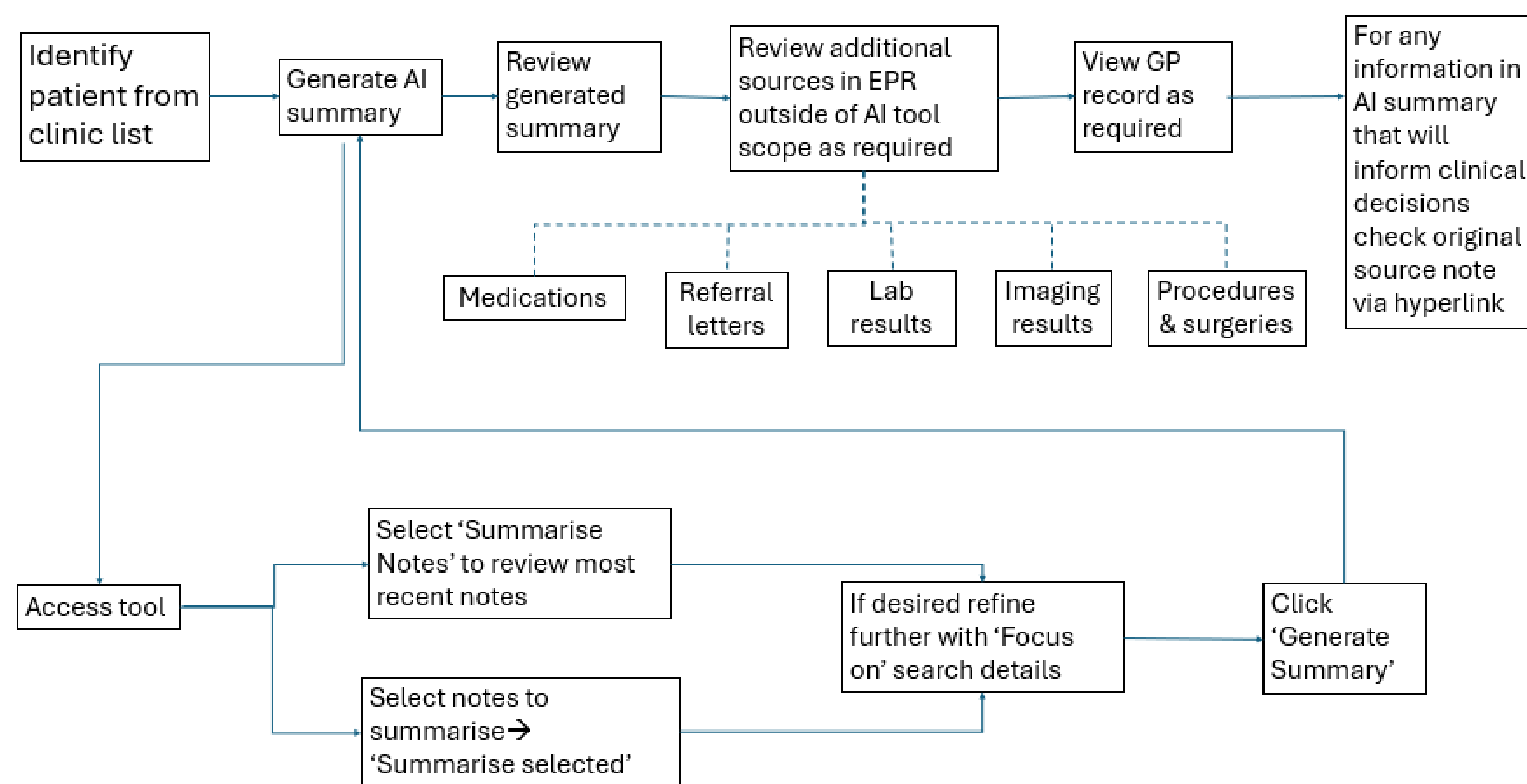
The Project

This project aimed to evaluate whether an AI tool (Epic Outpatient Insights Summarisation tool) could generate patient summaries of comparable quality to clinician-written summaries.

If comparable, a pilot study would be undertaken to assess whether implementing the tool could reduce clinic preparation time and improve clinician wellbeing.

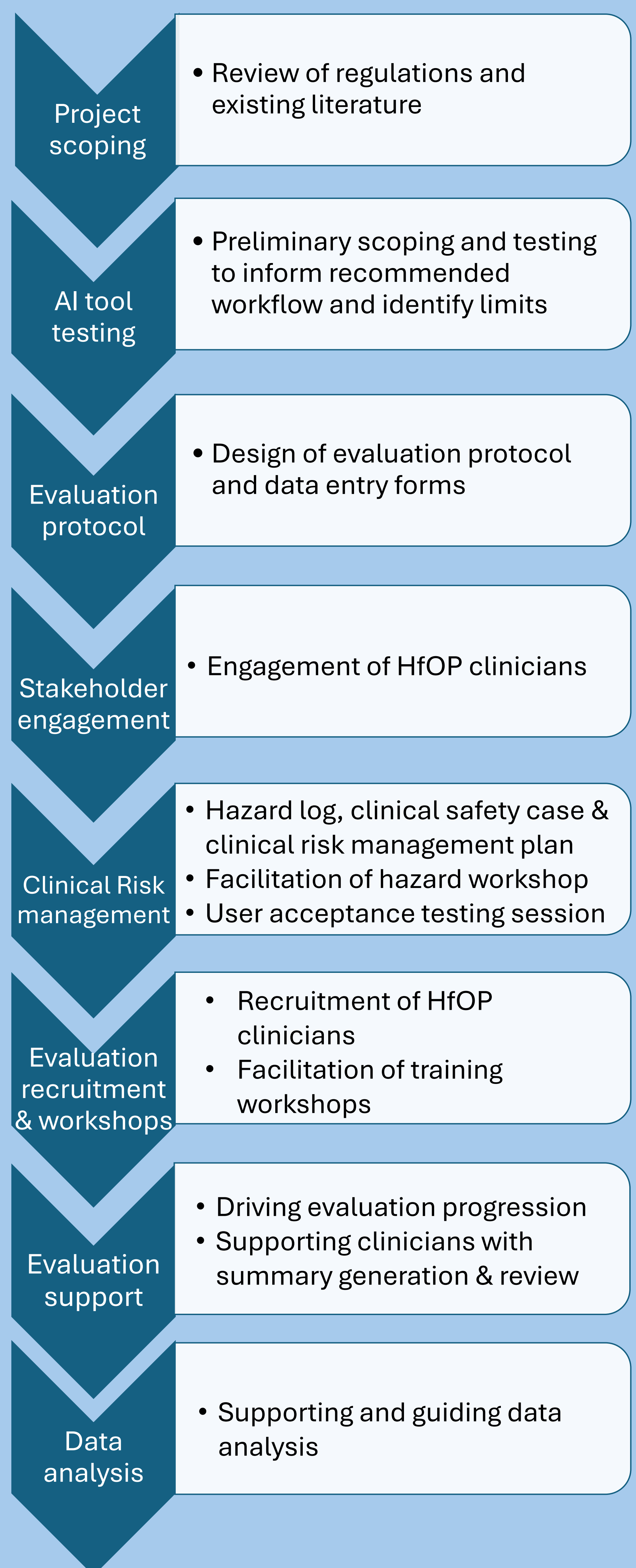
The AI solution

The Epic Outpatient Note Summarisation tool utilises the large language model (LLM) GPT-4.1 to generate a summary of a patient's previous notes within their Epic electronic patient record (EPR). By reducing need for manual note review, the tool has the potential to enhance efficiency and reduce administrative burden, potentially reducing burnout rates.



Milestones

My input as a fellow was supported by the host AI team, the local clinical safety officer, Epic team, senior research data scientist at the University of Exeter and the faculty.

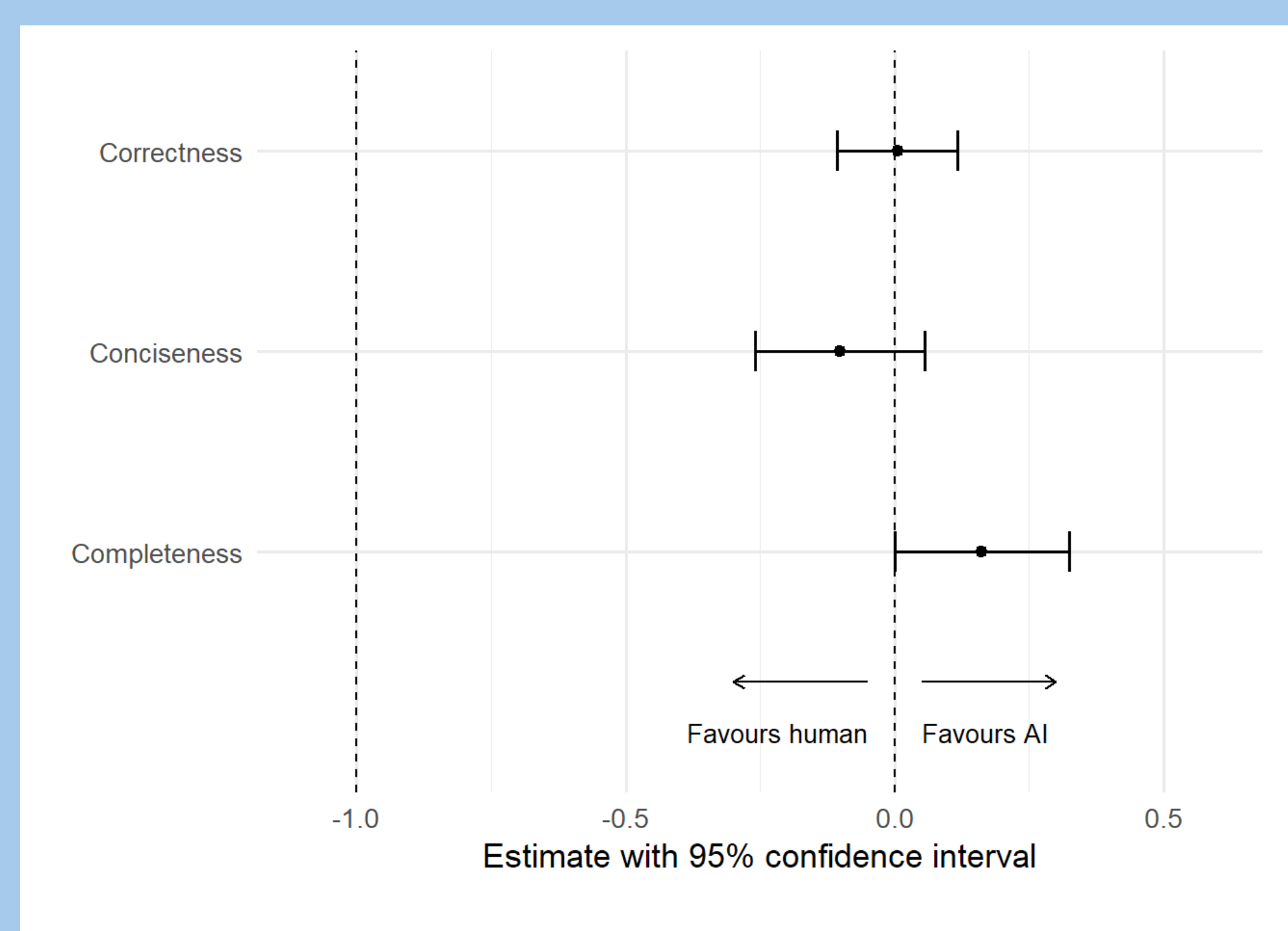
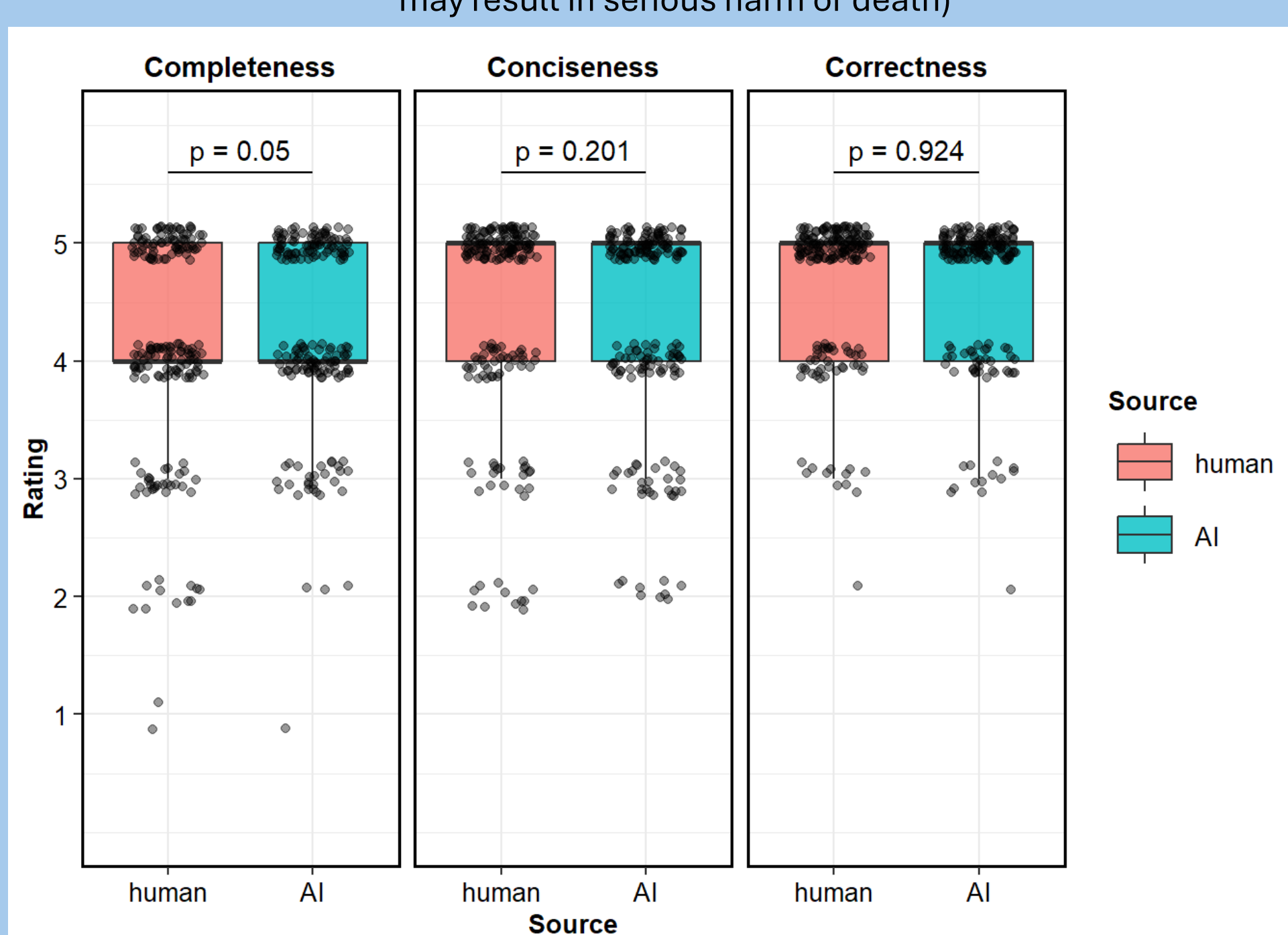


Clinical evaluation

- Healthcare for Older People (HfOP) and Gastroenterology specialties
- 10 clinicians recruited
- Generation of clinician written and LLM summaries for 100 patients
- Assessed for hallucinations and omissions
- Blinded clinician evaluators compared pairs of summaries (LLM and clinician written) and rated them 5-point Likert scale for:
 - Completeness
 - Correctness
 - Conciseness

	Frequency	Clinical Severity
Relevant information omitted in clinician summary but included in AI summary	49 (49%)	3 [2-3]
Relevant information omitted in AI summary	38 (38%)	3 [3-3]
False information (hallucinations) included in AI summary	36(36%)	3 [3-4]

Clinical severity of hallucinations and omissions scored on 5-point likert scale (1=no risk, 2= negligible, 3= some risk e.g. may alter diagnosis/ treatment considerations, 4= high risk e.g. may have a crucial impact on treatment decisions or delay diagnosis, 5= Major e.g. may result in serious harm or death)



Next steps...

- Review of results & completion of safety case to consider progression to pilot phase
- Training
- Plans for monitoring
- 12-week pilot to assess & monitor:
 - Efficiency improvements
 - Summary quality
 - Clinician wellbeing through surveys and focus groups