Evaluating Agentic Al as a Research Assistant in Biomedical Informatics

Yaa Kumah-Crystal, MD, MPH, MS Vanderbilt University Peter Embí, MD, MS School of Medicine

Jacob Beckstead Masters of Applied Clinical Informatics

How easily can you tell an AI to collect information? Not from the web but from a person?

- Al can be helpful but also intrusive – we need balance!
- How do out-of-the-box LLMs handle info-gathering?
- How do *people* react?

METHODS

- Amplify LLM Host Platform
- □ N = 488 text interviews
- Gave nearly-identical prompts to interview subject and 4 targeted topics

ADDITIONAL INVESTIGATIONS

- Does your name or title matter to the Al's approach?
- Can it accurately detect its own bias in questions?
- Will it hallucinate answers to questions it forgot to ask?

Want to participate? Get notified when interviews resume in November!



Every model has its quirks, but performance depends on your research question.

	Model 60-83 each	Que	stions GS	Cros	Rote	rdge olo
	3 Haiku	5.5	1.1	26.5	59	
	3.5 Haiku	8.4	1.3	4.6	75	
	3.5 Sonnet V2	9.3	1.3	0.0	71	
	3.7 Sonnet	11.3	1.5	19.3	79	
	4 Sonnet	11.6	1.8	0.0	74	
	40	8.1	1.4	0.0	58	
	4o-mini	8.7	1.5	0.0	73	
	50	28.2	2.4	1.6	95	
	Advanced Voice*(40)	3.8	0.79	0.0	80	

^{*} Assessment involved human volunteers (n=5)

ADVANCED VOICE INTERVIEWEE COMMENTS:

- Follow-up summaries seen as helpful
- Asked surface-level questions, limited follow-up or depth
- Deemed useful for initial trial screenings and interviews
- Added benefit of 24/7 availability and scheduling
- ← Lend your voice to this study! Sign up to be notified for the next round of interviews!

Open-ended, leading, and

OTHER METRICS UNDER ASSESSMENT:

- clarifying questions
- Active listening markers
- Positive reinforcement rate
- ☐ Topic order adherence
- Summary completeness
 - Topic deviation

APPLYING AI IN RESEARCH?

- Evaluate risks AND rewards
- ☐ Pre-define tolerable failure and appropriate safeguards
- Voice-trained LLMs quicken conveyance, broaden equity
- □ Expert opinion is not enough design for real-world culture

Have thoughts on AI in health, medicine, and research? We'd love to hear from you!



Presenter:

Jacob Beckstead



IRB #250086

jacob.beckstead@vumc.org

