Evaluating AI Voice Agents for Patient Education and Follow-up in Ophthalmology

Jaison Jain, BS[1]; James Kaan, MD[2]; Suraj Jain, BS[3]; Camilo Martinez, MD[4]; Apostolos Tassiopoulos, MD[5]

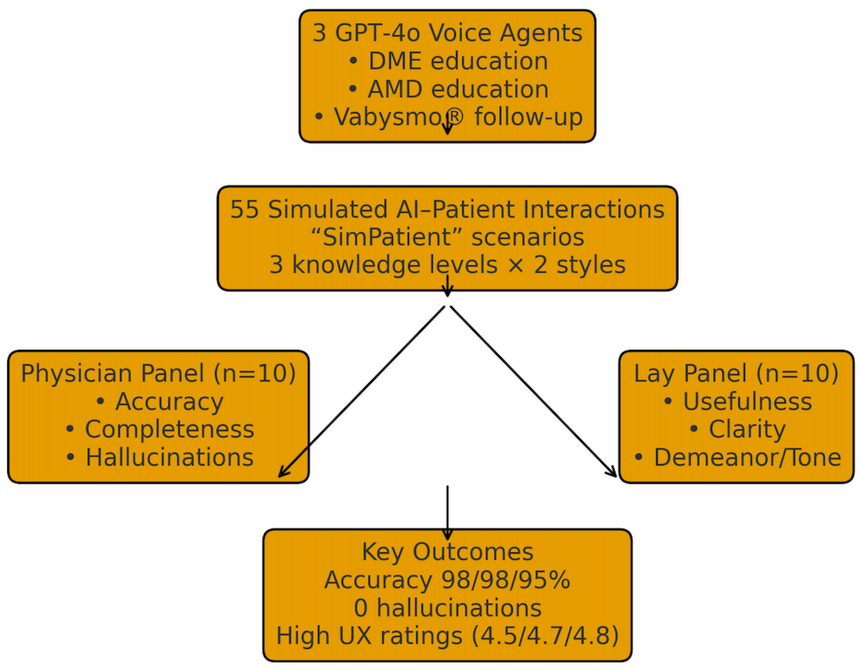
[1] Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, USA
[2] Department of Emergency Medicine, Penn State Health, Hershey, PA, USA
[3] New York Institute of Technology College of Osteopathic Medicine, Old Westbury, NY, USA
[4] Division of Vascular Surgery, Stony Brook University Hospital, Stony Brook, NY, USA
[5] Department of Surgery, Stony Brook University Hospital, Stony Brook, NY, USA

Introduction

- Chronic ophthalmic conditions like **Diabetic Macular Edema (DME)** and **Age-related Macular Degeneration (AMD)** require repeated education and reliable follow-up to sustain adherence and understanding.
- **AI voice agents** can scale expert, guideline-concordant communication for both education and treatment follow-up.
- We piloted three GPT-4o–based voice agents for: (1) DME education, (2) AMD education, and (3) **Vabysmo®** follow-up for AMD, and evaluated their quality in simulated patient encounters.

Methodology

- **Agent design:** GPT-4o prompt-engineered with (1) a curated/summarized corpus of **>300 ophthalmology articles and AAO guidance**, and (2) **guardrail prompts** to suppress unsupported content.
- **Scenarios:** 55 simulated AI-patient interactions ("**SimPatient**" framework) spanning 3 patient knowledge levels (naïve, moderate, advanced) × 2 communication styles (concise vs. verbose).
- **Evaluation panels:**
  - **Physicians (n=10):** Rated **accuracy** (valid clinical statements), **completeness** (required counseling elements delivered), and **hallucinations** (fabricated statements when the correct answer was negative).
  - **Non-physicians** (n=10; medical students & premed undergraduates): 5-point Likert ratings of usefulness, clarity, and professional demeanor/tone.



Results

- **Accuracy (physician-rated):**
  - DME education: **98%**
  - AMD education: **98%**
  - Vabysmo follow-up: **95%**
- **Safety: 0 / 55** interactions contained hallucinations.
- **User experience (Likert means across agents):**
  - Usefulness **4.5 / 5**
  - Clarity **4.7 / 5**
  - Demeanor/Tone **4.8 / 5**

Takeaway: Agents consistently delivered correct, complete, and professionally toned information across diverse patient profiles.

Agent	Accuracy (%)
DME Education	98
AMD Education	98
Vabysmo® Follow-up	95

Agent	Usefulness (1–5)	Clarity (1–5)	Demeanor/Tone (1–5)
DME Education	4.5	4.7	4.8
AMD Education	4.5	4.7	4.8
Vabysmo® Follow-up	4.5	4.7	4.8

Conclusion

- AI voice agents show strong potential to support **ophthalmic patient education and follow-up**, achieving **high accuracy**, **no observed hallucinations** in simulation, and **very positive user ratings**.
- **Limitations:** Simulated (not live-patient) calls, small panels, short-term outcomes, and a single-vendor LLM/corpus.
- **Next steps:** Prospective trials with real patients and clinical endpoints (adherence, visit/therapy completion, vision outcomes); comparisons with standard nurse/tech counseling or IVR; broader indications, languages, and accessibility features; continuous monitoring for model drift.

Acknowledgements

- We thank the **10 physician reviewers** and **10 non-physician participants** for their time and feedback.
- Affiliations: Renaissance School of Medicine at Stony Brook University; Penn State Health; New York Institute of Technology College of Osteopathic Medicine; Stony Brook University Hospital (Divisions of Vascular Surgery and Surgery).
- Disclosures/Funding: No financial conflicts reported.