

## **Optimization of a RAG Model Chatbot for Medical Education**

**Kenzie Knight**<sup>1</sup>, Kelsey Pape<sup>2</sup>, Esther Kim<sup>2</sup>, Caroline Rouse<sup>2</sup>, David Hanson<sup>3</sup>, Anthony Shanks<sup>2</sup>

<sup>1</sup>Indiana University School of Medicine; <sup>2</sup>Indiana University School of Medicine, Department of Obstetrics and Gynecology; <sup>3</sup>Indiana University School of Medicine, Department of Information Services & Technology

### **Background/Objective:**

In fast-paced clinical settings like labor and delivery, quick access to accurate, evidence-based guidance is critical. Traditional methods of sharing protocols can be inefficient, especially for trainees managing complex scenarios. Advances in artificial intelligence (AI) and large language models (LLMs) offer the potential to transform how critical clinical decision-making tools are disseminated, making them more accessible, responsive, and integrated into real-time learning. Our objective was to develop our own chatbot to facilitate quick retrieval and enhanced learning.

### **Methods:**

The AI chatbot was developed by IUSM researchers using Microsoft Copilot Studio and made accessible via Microsoft Teams. The AI chatbot was built on a retrieval-augmented generation (RAG) framework, leveraging a fine-tuned LLM. The base LLM was augmented with domain-specific documents, including IUSM protocols – developed by the Maternal-Fetal Medicine division - which incorporate evidence-based practices and provide a consistent management strategy for OB conditions. The RAG architecture was configured to retrieve contextually relevant documents from the collection of protocols in response to user queries.

### **Results:**

The chatbot retrieves information from the protocols based on a user's query and generates a relevant answer with reference to the sources. Upon testing the chatbot, we learned that the chatbot was unable to retrieve information from figures, flowcharts, and diagrams. Therefore, figures were manually annotated and narrated, enabling the model to process and retrieve relevant visual information during inference. With other adjustments, the chatbot is able to read all aspects of the protocols and synthesize accurate answers to clinical questions.

### **Conclusion and Potential Impact:**

We optimized an RAG model AI chatbot for clinical education, demonstrating that existing educational material can be used to activate a chatbot to serve clinical needs with sufficient oversight to ensure clinical accuracy. Future directions include evaluating its impact on clinical decision making and student perception of the learning environment once the chatbot is more uniformly adopted.