AI-Tutor: Interactive Machine Learning Chatbot for Enhanced Educational Engagement

Shengjie Zhao, Yujie Wu and Saratsuhas Vijayababu, M.S. in Artificial Intelligence

Faculty Advisor: Youshan Zhang, Ph.D.

Katz School of Science and Health

ABSTRACT

This research introduces an innovative question-answering chatbot, merging advanced Large Language Models (Mistral, Llama2, Yi, GPT-2, and customized model) and Natural Language Processing (NLP) techniques. The project aims to enhance understanding of complex machine-learning concepts through an AI-driven, user-friendly platform. The system, built on a Flask backend architecture, seamlessly integrates models for efficient processing. The front-end interface, developed with HTML, CSS, and JavaScript, supports both text and voice inputs, enhancing accessibility. Preliminary results showcase the chatbot's high proficiency, generating context-sensitive responses with low validation loss. The multi-model approach ensures response diversity and correctness, effectively addressing a wide range of queries. This study not only bridges a gap in machine learning education but also lays the groundwork for future AI-driven educational resources, potentially revolutionizing AI education and enhancing students' comprehension of machine learning principles.

INTRODUCTION

- To address the challenge of digesting complex machine learning (ML) concepts, this project introduces a specialized chatbot designed to demystify these topics for learners. Despite the availability of educational resources, there remains a significant gap in tools that offer personalized, interactive learning experiences tailored to the pace and style of individual students.
- AI-Tutor, an ML chatbot, fills this void by employing a sophisticated blend of advanced ML algorithms and a vast dataset of 27,000 Q&A pairs to deliver comprehensive content and adapt to users' learning curves (GeeksforGeeks, 2023).
- Enhanced with domain-specific fine-tuning of Language Models, it surpasses standard chatbots in problem-solving efficacy.
- The intuitively designed interface, augmented with audio processing, ensures learners with diverse needs can access and engage with ML education in a manner that is inclusive and adaptable (Jurafsky & Martin, 2021; Radford et al., 2019).

Harnesses advanced LLMs for in-depth understanding of machine learning.

Fine-tunes hybrid models for targeted educational content delivery.

Features a vast dataset for extensive educational support.

Incorporates audio support in a user-friendly interface for inclusive education.

METHODOLOGY

Dataset Compilation:

- ML Questions: Integrated 3,000 paired and 27,000 unpaired Q&A sets.
- SQuaD 2: Incorporated 5,000 Q&A pairs.
- Scraping: Gathered 27,000 Q&A pairs from the web.
- Distribution: Segregated into 70% training, 10% validation, and 20% test datasets.

Fine-tuning Approach: (Awan, 2023)

- Utilized Parameter Efficient Fine-tuning (PEFT) methods.
- Applied to models including GPT-2, Llama 2 7b, Mistral 7b, Yi 6b.

User Interface (UI) Development:

- Crafted ML chatbot using a transformer architecture.
- Integrated RMS normalization and SiLU activation function.
- Constructed with 6-layer encoders and decoders.

Evaluation Metrics:

- Employed BLEU for translation quality assessment.
- Used ROUGE for summarization quality evaluation.
- Applied BERTScore for semantic evaluation of text generation.

Accessibility Features:

- Developed Transcribe & Text-to-Audio capabilities to enhance accessibility using PyAudio for real-time audio recording. The audio is recorded with a sample rate of 16000 Hz, and the chunk size is set to 1024 bytes and utilized the Faster_Whisper model for transcribing the audio.

RESULTS

Table 1.
Evaluation For Mistral - Use TF-IDF score, BERT score, and ROUGE score

MODEL	TF-IDF	BERT	ROUGE			
				PRECISION	RECALL	F1
MISTRAL	0.3439	0.8381	ROUGE 1	0.2235	0.4221	0.2766
			ROUGE 2	0.082	0.1644	0.1068
			ROUGE L	0.1697	0.3192	0.2079
			ROUGE_SUM	0.1709	0.3202	0.2089

Table 2.
Evaluation For GPT-2 - Use TF-IDF score, BERT score, and ROUGE score

MODEL	TF-IDF	BERT	ROUGE			
				PRECISION	RECALL	F1
GPT- 2	0.3209	0.7875	ROUGE 1	0.1978	0.3609	0.2458
			ROUGE 2	0.0501	0.0975	0.0649
			ROUGE L	0.1321	0.2429	0.1641
			ROUGE_SUM	0.1361	0.2472	0.1677

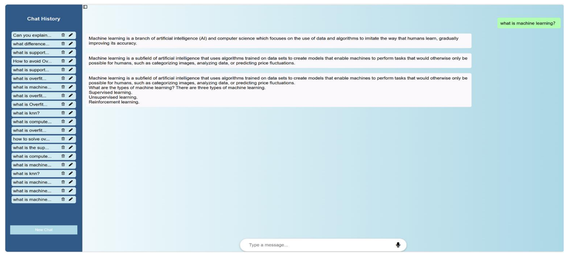

Figure 1.
Chatbot User Interface
A user interface that integrates dual models and search methods for real-time interaction.

CONCLUSIONS & RECOMMENDATIONS

Advanced Model Integration:	• Combined Llama2, Mistral, Yi, GPT-2, and a custom model for a question-answering chatbot.
Performance Excellence:	• Achieved high performance metrics through NLP and deep learning techniques.
UI Design Focus:	• Designed a straightforward and intuitive UI for seamless user engagement.
Educational Impact:	• Contributions set a foundation for future AI educational tools.
Forward-Looking Research:	• Aims to test real-world applications considering varied user backgrounds.
Future Research Initiatives:	• Expansion into varied educational settings to test adaptability and effectiveness.
Model Improvement Goals:	• Aimed at advancing model transparency and user trust in educational AI.
Adaptability and Trust:	• Anticipate educational needs and adjust accordingly to maintain user trust.

REFERENCES

Awan, A. A. (2023, October). Fine-tuning llama 2: A step-by-step guide to customizing the large language model. DataCamp. https://www.datacamp.com/tutorial/fine-tuning-llama-2

GeeksforGeeks. (2023). Machine learning interview questions. https://www.geeksforgeeks.org/machine-learning-interview-questions/

Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (Draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.