KatzBot-Intelligent ChatBot using Large Language Models

Deepa Paikar, Haider Ali, and Kiran Vutukuri,
M.S. in Artificial Intelligence

Faculty Advisor: Youshan Zhang, Ph.D.

Katz School of Science and Health

ABSTRACT

Today, university websites contain information that is diversified across various pages. Previously, a chatbot was develop using RASA which lacked the ability to answer complex questions. This project introduces KatzBot, an innovative chatbot system to enhance communication within and beyond the Yeshiva University community. By utilizing the RASA framework as a foundation, our development centered on the enhancement and rigorous evaluation of various Large Language Models (LLM) to improve the efficiency and accuracy of information delivery, as well as to foster greater user interaction. KatzBot is tailored to address the specific needs of Yeshiva University, acting as an interactive portal for prospective students, current students, alumni, and anyone in search of detailed university information. The primary objective of KatzBot is to streamline the information gathering process, significantly reducing the reliance on conventional methods. It aims to provide quick and easy access to data on academic programs, admissions procedures, campus life, and more, thereby simplifying the overall experience for its users.

INTRODUCTION

This research project explores the design, implementation, and evaluation of an advanced chatbot system, aiming to enhance user experience and functionality. As the demand for intelligent conversational agents grows, understanding chatbot development intricacies and user engagement nuances becomes increasingly crucial.

- **Problem Statement:** To develop a chatbot leveraging Large Language Models (LLM) with the goal of extracting and presenting information about university, courses, admissions, etc., directly to the end user.
- **Research Focus**: Our research focus is on the design, implementation, and evaluation of an advanced Large Language Model (Baglivo et al., 2023) aimed at enhancing user experience and functionality, particularly by feeding quality data and fine tuning the model.
- **Approach:** The development process of KatzBot follows a robust Data Science pipeline (Li et al., 2019). Initially, we conducted extensive data scraping, data cleaning, both within and outside Yeshiva University's website. We gathered 6,280 sentence pairs and 7,334 question & answer pairs. Subsequently, to enhance the model's comprehension and response capabilities, we undertook double fine-tuning exercises using LLMs.

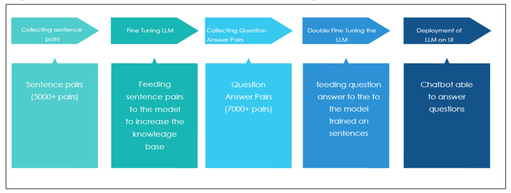
METHODOLOGY

Our approach combines data creation, model training, and performance evaluation:

- **Data Preparation**: In the data preparation phase, we embarked on an extensive data collection, sourcing relevant information from a variety of web pages both within and external to Yeshiva University's website.

Data Type	*Description*	*Quantity Generated*
Sentence Completion Pairs	Pairs created for model training	5,600
Question-Answer Pairs	QA pairs for detailed understanding	7,600
Test QA Pairs (Separate)	Pairs for model testing consistency	1000

- **Model Training**: We utilized LLMs like GPT-2, Llama2, Mistral Instruct, Phi (Li et al., 2023) to fine-tune in phases, first on sentence completion for contextual understanding, then on a question-answer dataset to improve response accuracy (Vaswani et al., 2017). Additionally, we worked on creating a custom model to meet specific needs in university data.



- **Evaluation**: The model's accuracy, precision, recall, and F1 score were measured to ensure robust detection capabilities against the test dataset (Xiang et al., n.d.).

RESULTS

GPT-2 excels in all the rouge scores, likely owing to its pre-training approach:

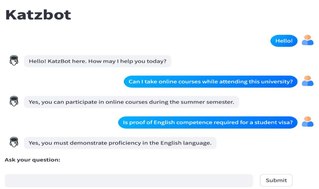
Metric	Microsoft Phi	GPT 2	Llama 7B	Mistral Instruct
rouge-1	0.403886	0.667471	0.342769	0.44855
rouge-2	0.256521	0.549339	0.212769	0.331086
rouge-l	0.379973	0.659397	0.334161	0.440821

Predictions Comparison of tested LLMs

- Question: what is a CRN?
- Ground truth: A CRN is a course registration number, which is unique to each course.

Model Name	Predicted Response
Llama 2	A CRN is a course that meets regularly, often in the evening, and is often offered in the summer……
GPT2	A CRN is a course registration number, which is unique to each course
Microsoft Phi	A CRN is a course registration number, which is unique to each course.
Mistral Instruct	A CRN, or Course Registration Number, is a unique identifier for each course offered at Yeshiva University……

The UI has been created and hosted accessible for testing:



Proposed deliverables for ongoing development and integration, aimed at improving outcomes:

Creating LLM from Scratch	Analyzing the fine-tuned LLM results, we're crafting our own LLM using GPT as the foundation, given its superior performance on our dataset.
Creating UI for chatbot	We've developed and hosted a user-friendly chatbot UI, enabling users to ask questions and receive instant answers.
Comparing Scratch-built vs. Pre Trained LLMs	We're comparing the accuracy and ROUGE score of the freshly created LLM with the pre trained LLM to determine the best model for our chatbot.
LLM Deployment in UI	Deploying the best performing model as the backend for the chatbot to ensure optimal performance and user experience.

CONCLUSIONS & RECOMMENDATIONS

Evaluation indicates performance discrepancies based on data; currently, GPT performs the best, but a customized model leveraging GPT as a foundation could yield superior results.

Challenges:

- Converting raw data to high-quality datasets.
- Storage memory and training cost for LLMs.
- Implementing contextual hate speech detection.

Future Research:

- Leveraging advanced RAG techniques to generate better results.
- Evaluating LLM with an LLM judge**.**

ACKNOWLEDGEMENTS

We thank professor Youshan Zhang for his invaluable guidance throughout this project. Thanks also to Atreish Ramlakhan, Jatin Kayasth, Lakshmikar Reddy Polamreddy, Ruslan Gokhman, and Manish Kumar Thota for their contributions in initiating this project and laying the foundation.

REFERENCES

Baglivo, F., Arzilli, G., Privitera, G. P., Ferragina, P., Tozzi, A. E., Rizzo, C., & De Angelis, L. (2023). HatGPT and the rise of large language models: The new AI-driven infodemic threat in public health.

Li, D., Zhou, L., Gao, J., & Shum, H.-Y. (2019). The design and implementation of Xiaoice: An empathetic social chatbot.

Li, Y., Bubeck, S., Eldan, R., Del Giorno, A., Gunasekar, S., & Lee, Y. T. (2023). Textbooks are all you need II: Phi-1.5 technical report.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

Xiang, Y., Dai, D., Zang, S., Wu, Y., & Wu, Y. (n.d.) Multi-task learning with multilingual transformers.