

ABSTRACT

Today, university websites contain information that is diversified across various pages. Previously, a chatbot was developed 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.
- **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