Text

Description automatically generated with low confidence

AI-Powered Coffee Shop Chatbot

**EAI 6010 – Applications of Artificial Intelligence**

Final Project Draft

**Prof. Sergiy Shevchenko**

Group: Husky-2

Rohit Kosamkar

Sapna Chavan

Leena Doultani

Harshal Joshi

# Abstract

In the evolving landscape of artificial intelligence, chatbots have become integral to enhancing customer experiences across various industries. This project focuses on developing an AI-powered chatbot tailored for a coffee shop's customer service operations. Leveraging Large Language Models (LLMs), Python, RunPod (Cloud Computing), Hugging Face, React Native, Pinecone, Firebase, and Retrieval-Augmented Generation (RAG). This project entails the development of a chatbot specifically designed for a coffee shop's customer service operations. The chatbot is equipped to take orders, provide detailed menu information, filter out irrelevant conversations, and make intelligent product recommendations based on customer interactions. To achieve these functionalities, we have integrated several advanced technologies:

* **Hugging Face:** Offers access to pre-trained LLMs and embedding models, enhancing the chatbot's intelligence.
* **Llama 3.1-70B-Instruct:** Pre-trained LLM offers advanced natural language understanding and response generation capabilities.
* **Python:** Provides a robust ecosystem for AI development, facilitating seamless integration of machine learning and chatbot functionalities.
* **Pinecone:** Serves as a vector database that indexes and stores vector embeddings for fast retrieval and similarity search, enhancing the chatbot's ability to provide relevant responses.
* **Retrieval-Augmented Generation (RAG):** Combines the strengths of traditional information retrieval systems with generative AI, allowing the chatbot to access external knowledge bases to generate more accurate and contextually relevant responses.
* **RunPod:** Enables cost-effective, scalable serverless deployment for AI models, reducing infrastructure complexity.
* **Firebase:** Acts as a real-time database and backend service, managing coffee shop products details, authentication, and synchronization across devices.
* **React Native:** Ensures seamless, cross-platform mobile application experience for customers.

The implementation of this chatbot offers several benefits, including enhanced customer experience through prompt responses and personalized recommendations, scalability and cost efficiency via serverless deployment, and automation that reduces the workload on human employees. However, we also anticipate challenges such as potential latency in response generation, difficulties in understanding highly ambiguous or sarcastic customer inputs, and the need for substantial computational resources during initial deployment and model fine-tuning.

In conclusion, this project demonstrates the potential of AI-powered chatbots in the food and beverage industry, making customer interactions smarter, more efficient, and engaging. As AI continues to advance, such technology will pave the way for more intuitive, context-aware, and personalized customer experiences in the retail and hospitality sectors.

# References

 Python Software Foundation. (n.d.). *Python Language Reference* (Version 3.8.10) [Software]. Retrieved from <https://www.python.org>

 RunPod. (n.d.). *RunPod Documentation* [Software]. Retrieved from <https://docs.runpod.io/overview>

 Hugging Face, Inc. (n.d.). *Transformers* (Version 4.0) [Software]. Retrieved from <https://huggingface.co/docs/transformers>

 Meta AI. (2023). *Llama 3.1-70B-Instruct* [Large language model]. Retrieved from <https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct>

 Pinecone Systems, Inc. (n.d.). *Pinecone Vector Database* [Software]. Retrieved from <https://www.pinecone.io>

 Google LLC. (n.d.). *Firebase* [Software]. Retrieved from <https://firebase.google.com>

 Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Riedel, S. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. Retrieved from <https://arxiv.org/abs/2005.11401>

 Facebook, Inc. (n.d.). *React Native* [Software]. Retrieved from <https://reactnative.dev>