

Research Proposal

Name: Vipin Mishra

Course: Masters in Machine Learning and AI (LJMU/Upgrad)

Research Title:

Revolutionizing Organizational Information Retrieval through Advanced Generative AI:

A Multi-Modal Approach through conversational Chatbot and Intelligent Question Answering for assisting customers in Air Travel Domain

Summary of the proposed research:

Airline organizations are overloaded with information like travel policy, pet travel policies, checkout policies, seats policies, travel promotions, source and destination news and a lot of other information, which makes it harder and harder for customer and travel agent to get required information in faster manner and time. The goal of this research project is to create a conversation chatbot and question-answering system driven by generative artificial intelligence (Gen AI). This system will enable airline customers and travel agents to instantly obtain the knowledge they need, expediting decision-making and raising overall productivity and customer satisfactions.

This research will primarily concentrate on utilizing Large Language Models capabilities to develop an advanced artificial intelligence system. Using a vast body of information, this AI system will be trained to comprehend and react to natural language inquiries, offering succinct, context-sensitive responses.

The study will examine Large Language Models from well-known providers, including Google, AWS, LangChain and Open AI, and evaluate their efficacy in enhancing organizational information retrieval through advanced indexing, Artificial Intelligence (AI), and Machine Learning (ML) techniques.

Research questions:

- What is the pet travel policy for [Airline Name] within the USA?
- Can I bring my pet on board as a carry-on passenger, or do they have to go in cargo?
- What are the size and weight restrictions for pets traveling in-cabin?
- Is there a fee for traveling with a pet, and how much does it cost?
- Do I need to make a reservation for my pet in advance?
- What types of pets are allowed on board (e.g., dogs, cats, birds)?
- Are there breed restrictions for traveling with dogs?
- Do I need a specific pet carrier, and if so, what are the requirements?
- What health and vaccination requirements are there for pets traveling domestically?
- Can my pet travel in-cabin with me if they are an emotional support animal or service animal?
- Are there any additional rules or fees for international pet travel?
- What is the process for checking in a pet at the airport?
- Can I feed or water my pet during the flight?
- What happens if my pet becomes anxious or agitated during the flight?
- Is there a limit to the number of pets allowed on a single flight?
- How can I track my pet's journey when traveling in cargo?

Research Aim & key objectives:

Large Language Model (LLM) Integration:

Investigate and compare various LLMs, such as Open AI GPT-4, Google Palm2 and AWS Bedrock to create a conversation chatbot or intelligent question answering system.

Knowledge Base Analysis:

Analyze various airlines knowledge base and data to source input data to chatbot or question answering system for creating natural language response for any query.

Indexing and Retrieval Techniques:

Implement advanced indexing and retrieval techniques to gather required data from airlines knowledge base and index it for feeding data to large language models.

Contextual Understanding:

Develop a deep contextual understanding mechanism to provide nuanced responses that consider the context of the customer's or travel agent's query for any travel policy requirements.

Customization and Adaptation:

Enable the system to be easily customized and adapted to the unique needs of different airlines, also ensuring versatility and scalability based on customer's needs.

Evaluation and Performance Metrics:

Establish comprehensive evaluation metrics to compare the chatbot's accuracy, efficiency, and customers satisfaction.

Ethical Considerations:

Address ethical concerns related to airline customer's data privacy, data security and propose strategies for creating responsible and secure chatbot system.

Machine Learning for Personalization:

This research will provide guidance to create customize models to personalize responses based on user preferences, airline origin country, and real-time data.

Comparison and Benchmarking:

Conduct a rigorous comparative analysis of the LLM-powered chatbots between AWS Bedrock, Google Plam2 and Open AI GPT models, highlighting their strengths, weaknesses, security and suitability for organizational information retrieval.

Recommendations and Guidelines:

This research will provide recommendations and guidelines for airlines considering the adoption of conversational chatbots, offering insights into model selection, customization, and secure deployment strategies.

This research aims to provide airlines with a powerful chatbot tool that can transform their customers how they access and leverage their knowledge resources. Ultimately, the proposed chatbot and question-answering system will contribute to more informed decision-making processes, enhanced competitiveness and improved productivity for airlines customers.

This research will represent a significant step in data retrieval using generative AI-driven knowledge to create conversational chatbot and question answering system.

Dataset:

In this research, I will utilize publicly available information obtained from multiple airline websites, with a specific emphasis on gathering information pertaining to airline policies regarding pet's travel. The information serves as an input resource for the development and implementation of a secure conversational chatbot or intelligent question-answering system. The primary objective of this chatbot system is to effectively and promptly provide essential information to both airline consumers and travel agents.

Data set links:

- <https://www.delta.com/us/en/pet-travel/overview>
- <https://www.aa.com/i18n/travel-info/special-assistance/pets.jsp>
- <https://www.united.com/en/us/fly/travel/traveling-with-pets.html>
- <https://www.southwest.com/help/booking/pet-policy>
- <https://www.alaskaair.com/content/travel-info/pets>
- <https://www.aircanada.com/us/en/aco/home/plan/special-assistance/pets.html>
- <https://www.airfrance.us/information/passagers/voyager-avec-son-animal-chien-chat>

Methods Used for Analysis:

A generative AI chatbot is a generative artificial intelligence based system with the help of natural language processing and deep learning technologies to generate natural human like text responses for asked question.

I am using below methods/technologies to develop conversational Chatbot and Intelligent Question Answering system.

Vertex AI:

Vertex AI is a Google Cloud based Machine learning platform which provides platform to train and deploy machine learning and large language models and generative AI based applications. Vertex AI also provides open-source models and google palm 2 models like chat bison and text bison in its model garden interface

LangChain:

LangChain is a framework built around large language models and is used for developing applications powered by language models. We can use it for chatbots, generative question-answering systems and much more. LangChain can index content from various format of data like text, html, pdf and many more.

PaLM API:

PaLM API is part of Google's Pathways Language Model 2 (PaLM 2), which is a large language model (LLM) that can generate text or chat in response to natural language prompts. It is a language model developed by Google AI Language. Palm 2 full form is Pattern- and Landmark-based Model 2. It uses a combination of pattern-based and landmark-based methods to generate more accurate and natural language responses. I'll use text bison and chat bison models from PaLM API for this research.

Amazon Bedrock:

Amazon Bedrock is a generative AI based fully managed artificial intelligent service from Amazon that makes Amazon native large language models or any third party large language models providers accessible through rest API , it also offers a choice of high performing foundation models from artificial intelligence based companies like Anthropic, Meta, Claude and Amazon. It also provides a broad set of capabilities that we need to build generative AI applications like conversational chatbot and intelligent question-answering systems

Open AI GPT:

Open AI GPT is a large language model that's uses image and text inputs and provides text human like responses or outputs. It is also a generative AI powered large language model that enables us to have human-like conversations and so much more with a chatbot by using organizations date. We can set a prompt based on our requirement and when we ask any question using this model, it uses source indexed data and provides human-like response based on given prompt.

Outcome:

Main outcome of this research is to build a generative artificial intelligent based conversation chatbot using AWS, Google Cloud, and Open-AI technologies. I expect this chatbot to be really good at remembering details and giving responses that fulfil the need of any airline customer and travel agent. It's not just about the technical side; I hope the results of this research will help improve how we use generative artificial intelligent in everyday situations in airline domain. I will analyze and compare AWS, Google Cloud, and Open AI to make sure this chatbot is as smart and interactive as possible. The goal is to bring a more practical and useful approach to using this kind of AI in real-life situations like checking pet travel policy for domestic or international travel for any airline.

In short, my research focuses on crafting a chatbot tailored for the airline industry, with a special focus on pet travel policies. The main aim is to assist travel agents and customers in swiftly accessing the necessary information. So, anticipated outcome is a user-friendly chatbot that excels in providing quick and relevant details about pet travel policies. The goal is to simplify the process and make pet travel hassle-free, creating a win-win for everyone involved.

Risks or Contingency Plan:

Risks:

Technical Challenges:

The integration of AWS, Google Cloud, and OpenAI technologies may present unforeseen technical challenges, such as compatibility issues or limitations in the APIs. This could potentially impede the seamless collaboration between these platforms.

Data Privacy and Security Concerns:

Handling large volumes of data, especially sensitive information for training the chatbot, could pose risks related to data privacy and security. Unauthorized access or data breaches may compromise the confidentiality of user interactions.

Ethical Considerations:

The use of generative AI in chatbots raises ethical concerns, such as biased responses or unintended consequences. Ensuring fairness and mitigating biases in the chatbot's language generation is crucial.

Resource Constraints:

The research may face challenges related to resource constraints, including computational power and financial limitations, especially when working with extensive datasets or computationally intensive models.

Contingency Plan:

Technical Documentation:

Read and understand all the required documentation for AWS Bedrock, Google Cloud Palm 2 LLM, and OpenAI GPT LLM models and foundations models

Data Privacy Measures:

Implement robust data privacy measures, including encryption on source data like using KMS or CMK based encryptions and access controls, to secure user data. Follow data protection regulations and ensure compliance throughout the research process.

Ethical Oversight:

Implement regular audits and adjustments to the training data and algorithms to minimize ethical risks.

Resource Management:

Develop a detailed resource management plan to allocate computational resources efficiently. Explore cost-effective and secure services and models.

Timetables and Milestones using Gantt chart:

Time Table Details:

Week 1:

Research Proposal Planning and Setup

- Define research objectives and scope.
- Establish a project timeline.

Week 2:

Literature Review

- Literature review on generative AI, chatbots, and AWS Bedrock, GCP Palm2, Open AI GPT, LangChain and relevant technologies.

Week 3:

Platform Familiarization

- Familiarize the team with AWS, Google Cloud, and Open AI platforms.
- Identify existing methodologies and best practices.

Week 4:

Data Collection and Data Preprocessing

- Identify relevant information and data for training.
- Ensure data compliance with ethical and privacy standards.
- Pre-process and clean the collected data to enhance quality.
- Address any missing or inconsistent data.

Weeks 5-7:

Model Development and Training

- Design the architecture for the generative AI-based chatbot.
- Define parameters and features.
- Develop and implement the chatbot model.
- Train the model using the preprocessed data.
- Conduct initial tests to identify issues.

Weeks 8-9:

Evaluation and Iteration & Mid-Thesis Preparation

- Evaluate the chatbot's performance against predefined metrics.
- Identify areas for improvement and potential issues.
- Iterate on the model based on the evaluation results.
- Fine-tune parameters for optimal performance.

Weeks 10-11:

Comparison and Optimization of Models & Mid-Thesis Preparation

- Compare models and results
- Optimize models

Week 12:

Documentation & Mid Thesis Report Submission

Weeks 12-20:

Finalization and Documentation & Final Thesis Report Preparation

- Conduct final testing to ensure the model's stability.
- Address any remaining issues and debug as needed.
- Document the entire research process, methodologies, and challenges.
- Prepare a final report summarizing key findings, insights, and the model's performance.
- Create a presentation for stakeholders to share the results.

Weeks 21:

Final Thesis & Video Presentation Submission

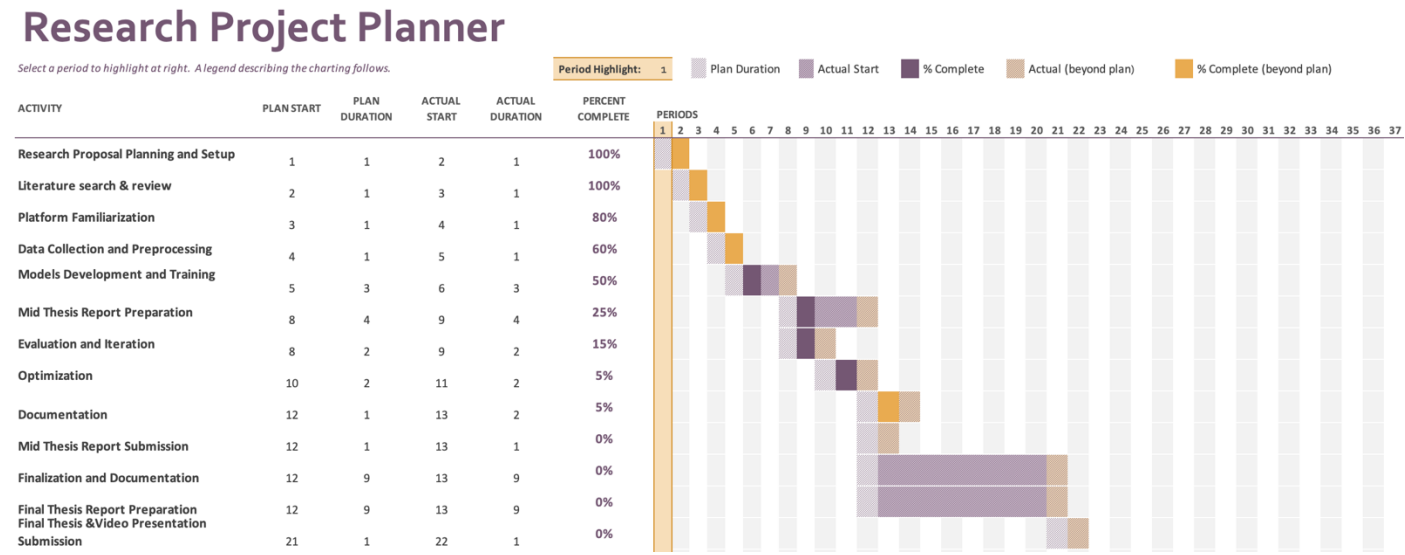
Time Table (Dates):

Task	Start Date	End Date
Research Proposal Planning and Setup	Oct 16, 2023	Oct 20, 2023
Literature Review	Oct 23, 2023	Oct 27, 2023
Platform Familiarization	Oct 30, 2023	Nov 03, 2023
Data Collection and Preprocessing	Nov 06, 2023	Nov 10, 2023
Models Development and Training & Mid Thesis Report Preparation	Nov 13, 2023	Dec 01, 2023
Evaluation and Iteration & Mid Thesis Report Preparation	Dec 04, 2023	Dec 15, 2023
Optimization & Mid Thesis Report Preparation	Dec 18, 2023	Dec 29, 2024

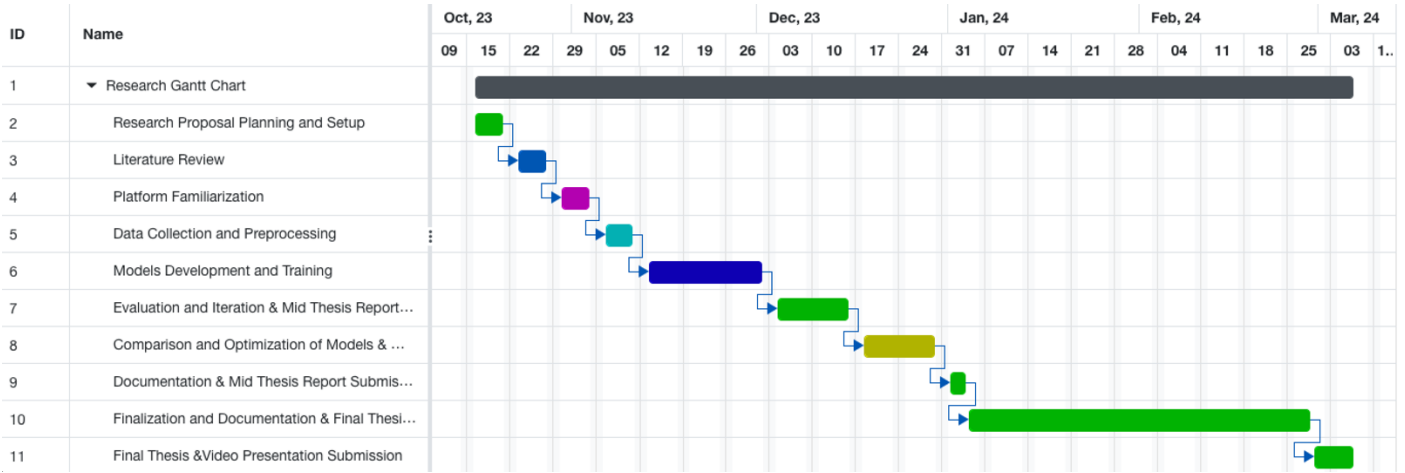
Documentation & Mid Thesis Report Submission	Jan 01, 2024	Jan 3, 2024
Finalization and Documentation & Final Thesis Report Preparation	Jan 04, 2024	Feb 28, 2024
Final Thesis &Video Presentation Submission	Feb 29, 2024	March 06, 2024

Gantt Charts:

Gantt Chart (with number of weeks):



Gantt Chart (With Dates):



References:

Title	Link to the Paper	Understanding of the Dataset	Understanding the Methodology Used
Retrieval-Augmented Generative Pre-Training for Question Answering	https://arxiv.org/abs/2005.11401	This given research paper proposes a retrieval augmented generative pre-training model for chatbot and question answering. This model combines a generative AI based pre-training model with a retrieval module. This module retrieves relevant data from a knowledge base, and the generative model generates answers based on the given documents and the question.	Retrieval-Augmented Generative AI
Comparative study of retrieval based and generative AI based chatbots using Machine Learning & Deep Learning	https://www.sciencedirect.com/science/article/pii/S2772442523000655	This given paper provides literature analysis of the generative based chatbot techniques, it also explains the implementation, and the dataset, along with the explanation of the experimental results.	Conversational chatbot (Retrieval and Generative)
Enriching Conversation Context in Retrieval-based Chatbots	https://arxiv.org/abs/1911.02290	This given paper provides details about work on information retrieval based AI chatbots, which is almost like sequence pair matching task with the help of cross encoders which is used for word matching over the pair also for bi-encoders that encode the different pairs separately.	Cross encoders, Bi-encoders, a sequence pair matching task,
Conversational question answering: a survey	https://link.springer.com/article/10.1007/s10115-022-01744-y	This survey paper provides an overview of the techniques and approaches used in question-answering systems for conversational AI.	Conversational question answering & Sequential KB-QA systems
Future directions for AI based chatbot research	https://link.springer.com/article/10.1007/s00607-021-01016-7	This given paper proposed, future research directions for research work done as part of the conversational work, it also provides a perspective of influence of chatbots at the individual age groups.	NLP (Natural Language Processing), Embodied conversational agents
Generative Artificial Intelligence based Conversational Chatbots integration methodologies and challenges	https://www.mdpi.com/2071-1050/15/5/4012	This paper provides details about conversational chatbots integration methodologies and challenges	Natural Language Processing and Artificial Neural Networks
Pre training of bi directional Transformers for unlabeled text	https://arxiv.org/abs/1810.04805 https://paperswithcode.com/paper/bert-pre-training-of-deep-bidirectional	This given paper represents Bidirectional Encoder Representations from Transformers, which is a pre-trained model that has been widely used for various NLP tasks.	Bidirectional Encoder Representations - Transformer
A web UI based chatbot tool to represent text based communication between humans and with machines	https://arxiv.org/abs/1911.00665	This research papers described web UI based chatbot tool to represent text based communication between humans and with machines.	Dialogue Systems in Natural Language Processing
Large Scale Generative Pre-training for Conversational Response Generation using Dialog GPT	https://aclanthology.org/2020.acl-demos.30/	This paper presents DialoGPT, a model specifically designed for generating human-like responses in conversations.	Dialogue Generative Pre-Trained Transformer