TravelGPT: A Multi-Agent Architecture for Mitigating Cultural Biases in LLMs: A Case Study on Sikh Heritage in Pakistan

Parkash Singh
The Ohio State University
singh.2068@osu.edu

Abstract

Large Language Models (LLMs) exhibit biases due to uneven training data distributions, affecting their ability to provide reliable, culturally relevant knowledge. This study investigates the limitations of LLMs in Sikh tourism guidance for historical sites in Pakistan, leveraging Retrieval-Augmented Generation (RAG) and a specialized multi-agent architecture to mitigate biases and enhance domain-specific knowledge retrieval. Our TravelGPT system demonstrates competitive performance compared to leading LLMs while requiring fewer computational resources. Evaluation using BLEU scores shows our approach outperforms GPT-4 and approaches Claude 3.7 Sonnet's capabilities through domain-specific optimization and culturally sensitive information retrieval.

1 Introduction

1.1 Background and Motivation

Large Language Models (LLMs) have revolutionized Natural Language Processing (NLP) applications, but their reliance on broad, web-scraped corpora introduces cultural and geographic biases. These biases often lead to hallucinations, misrepresentation, or omission of region-specific knowledge, especially for historical and religious domains. Studies have found that LLMs disproportionately favor dominant cultural narratives, marginalizing lesser-documented regions and faith-based knowledge. This project focuses on Sikh tourism in Pakistan as a case study to explore and mitigate such biases.

Pakistan holds immense historical significance for the Sikh faith, with sites such as Nankana Sahib (birthplace of Guru Nanak) and the Lahore-era Sikh Empire landmarks attracting thousands of pilgrims annually. However, the lack of consolidated, accurate, and up-to-date online resources makes it difficult for visitors to plan independent trips. Current digital resources are often fragmented, outdated, or lack the cultural context necessary for meaningful pilgrimages.

1.2 Research Objectives

This project aims to develop an AI-powered chatbot to provide accurate and domain-specific guidance for Sikh tourists visiting Pakistan. The system addresses several key challenges:

1. Cultural and historical accuracy in representing Sikh heritage sites

2. Practical travel information tailored to religious pilgrimage contexts

3. Multilingual support to accommodate diverse pilgrims

4. Contextual understanding of religious significance and practices

1.3 Approach

Our approach leverages a specialized multi-agent architecture with Retrieval-Augmented Generation (RAG) to enhance response accuracy and cultural sensitivity. The system employs domain-specific knowledge bases created through targeted data collection from authoritative sources, expert verification, and structured organization. This architecture enables the system to provide comprehensive guidance covering historical context, practical travel logistics, accommodation options, and cultural protocols while maintaining respect for the spiritual significance of the sites.

Unlike general-purpose LLMs that struggle with specialized cultural and religious knowledge, our system is designed specifically for the intersection of travel planning and Sikh heritage exploration. The initial implementation uses a RAG pipeline with the flexibility to transition to a more sophisticated custom-built retrieval system as user needs and system requirements evolve.

2 Related Work

Prior studies on LLM biases have shown that models trained on internet-scale corpora tend to favor dominant cultural narratives (Bender et al., 2021). Research in Retrieval-Augmented Generation (RAG) has demonstrated that external retrieval can improve accuracy and reduce hallucinations in domain-specific applications (Lewis et al., 2020). However, few studies focus on historical tourism and religious knowledge.

This research addresses a practical and academic gap by applying RAG to faith-based tourism, exploring how LLMs can integrate expert-recommended historical sources to enhance accuracy.

3 Methodology

3.1 System Architecture

The system employs a retrieval-augmented generation (RAG) architecture to deliver accurate and contextually appropriate responses to user queries about Gurdwara information. As illustrated in Figure 2, our architecture follows a three-layered cyclical approach consisting of the User Interface, Multi-Agent Processing, and Retrieval-Augmented Generation components working in concert to provide comprehensive responses. The core components include a vector database for semantic search (Qdrant), an embedding model for text representation (HuggingFace's "all-MiniLM-L6-v2"), a language model for response generation (Llama3-8b-8192 via Groq API), and a user interface built with Streamlit. These components interact within a pipeline designed to process user queries, retrieve relevant information, and generate appropriate responses based on the retrieved context.

3.2 Data Collection

For this research, data collection was performed using Crawl4AI, an open-source web crawler specifically designed for Large Language Model (LLM) applications. Crawl4AI is a robust tool that delivers AI-ready web crawling tailored for LLMs, AI agents, and data pipelines. The tool was selected for its ability to generate clean, structured data suitable for direct integration with language models.

The data collection process utilized several key capabilities of Crawl4AI. The system leveraged Crawl4AI's heuristic markdown generation to transform web content about Gurdwaras into clean, well-structured text documents. This preprocessing step ensured that the collected information maintained semantic integrity while being optimized for vector embedding. Using Crawl4AI's "structured extraction" capabilities, we parsed repeated patterns of Gurdwara information using CSS and XPath selectors. This approach enabled consistent extraction of key details such as Gurdwara names, locations, historical significance, and other relevant metadata.

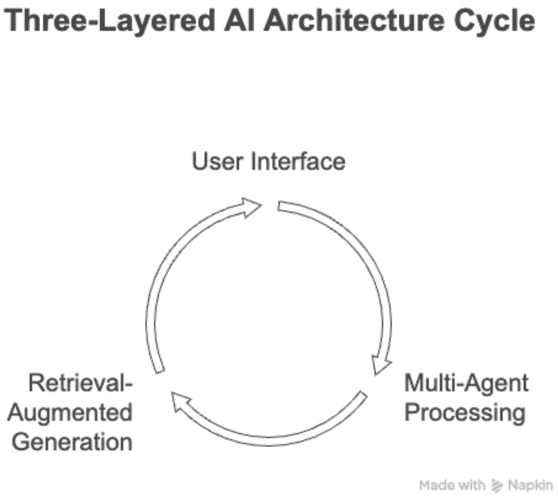

Three-Layered AI Architecture Cycle
User Interface
Multi-Agent Processing
Retrieval-Augmented Generation
Made with Napkin

Figure 1: Three-Layered AI Architecture Cycle showing the interaction between User Interface, Multi-Agent Processing, and Retrieval-Augmented Generation components.

After collection, all data underwent verification against multiple authoritative sources to ensure accuracy and completeness. The extracted information was then formatted into CSV files, which served as the primary data source for the RAG system. This preprocessing stage maintained the contextual relationships between different pieces of information, facilitating more coherent responses during the retrieval phase.

3.3 Multi-Agent System Architecture

As depicted in Figure 1 and Figure 3, our system implements a multi-agent workflow architecture that breaks down complex queries into manageable subtasks. The workflow process consists of five key stages. The process begins when a user submits a query about Gurdwara information through the Streamlit interface. A planner agent then analyzes the query and decomposes it into logical subtasks that can be handled by specialized agents.

Specialized agents, as detailed in Figure 3, ex-

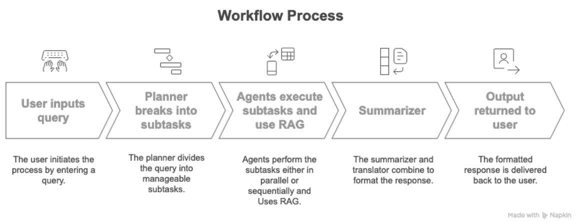

Workflow Process
User inputs query
Planner breaks into subtasks
Agents execute subtasks and use RAG
Summarizer
Output returned to user
The user initiates the process by entering a query.
The planner divides the query into manageable subtasks.
Agents perform the subtasks either in parallel or sequentially and Uses RAG.
The summarizer and translator combine to format the response.
The formatted response is delivered back to the user.
Made with Napkin

Figure 2: Workflow Process showing the five stages of query processing: user input, planning, agent execution, summarization, and output delivery.

ecute these subtasks using RAG methodology. These agents include the Travel Planning Agent which provides routes, timing, and itinerary assistance; the Heritage Retrieval Agent which supplies historical and spiritual information; the Food/Stay Agent offering options for langar halls and hotels; the Summarizer Agent which compresses and condenses retrieved text; and the Language Agent facilitating translations in Punjabi, Urdu, and English.

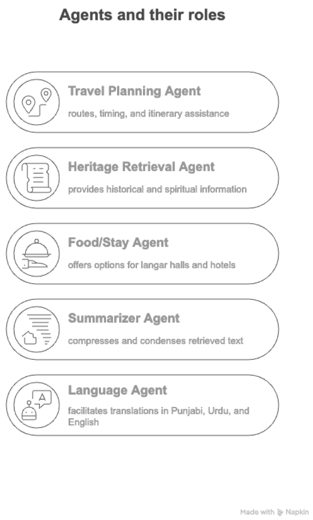

Agents and their roles
Travel Planning Agent
routes, timing, and itinerary assistance
Heritage Retrieval Agent
provides historical and spiritual information
Food/Stay Agent
offers options for langar halls and hotels
Summarizer Agent
compresses and condenses retrieved text
Language Agent
facilitates translations in Punjabi, Urdu, and English
Made with Napkin

Figure 3: Specialized Agents and their roles within the multi-agent architecture.

The summarizer agent then combines all retrieved information into a coherent, formatted response. Finally, the formatted response is presented to the user through the Streamlit interface.

3.4 Query Processing and Response Generation

User queries undergo a sophisticated processing pipeline before response generation. The system first reformulates input queries to ensure they are standalone questions, regardless of conversation history. This is achieved through a dedicated prompt template and language model processing step. The reformulated query is then used to retrieve relevant information from the vector database using semantic similarity search. The retrieved context is integrated into a specialized prompt template that instructs the language model on how to respond appropriately, emphasizing accuracy, respectfulness, and clarity in the context of Gurdwara information.

As shown in Figure 4, our approach to building a culturally rich AI experience combines four essential elements: Domain-Specific Expertise, which provides specialized knowledge about Sikh heritage, traditions, and Gurdwaras; Modular Agent Architecture, the multi-agent system that divides tasks by expertise; RAG for Verified Information, ensuring responses are grounded in factual, verified data; and Culturally Respectful Responses, maintaining appropriate tone and cultural sensitivity.

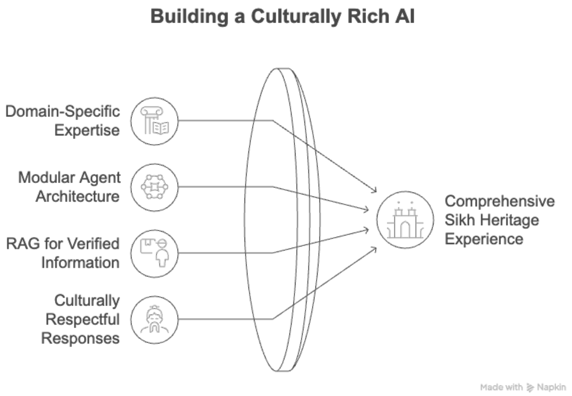

Building a Culturally Rich AI
Domain-Specific Expertise
Modular Agent Architecture
RAG for Verified Information
Culturally Respectful Responses
Comprehensive Sikh Heritage Experience
Made with Napkin

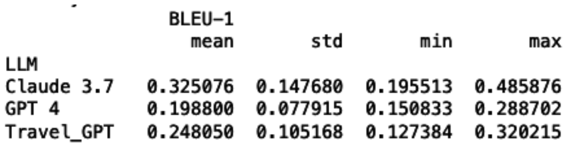
Figure 4: Components for Building a Culturally Rich AI Experience for Gurdwara information.

These elements work together to create a comprehensive Sikh heritage experience that provides accurate, respectful, and helpful information to users seeking to learn about Gurdwaras. The final response is generated by the Llama3-8b-8192 model, which synthesizes the retrieved context with the user query to produce informative answers. This approach grounds the model's responses in the factual information from the dataset while leveraging the language model's capabilities for natural language generation and contextual understanding.

4 Evaluation

4.1 Evaluation Methodology

To evaluate our proposed TravelGPT architecture, we conducted a comparative analysis against two leading large language models: Claude 3.7 Sonnet and GPT-4. The evaluation methodology employed identical prompts across all three models using a subset of our dataset in the form of a standardized CSV file containing travel-related information. Performance assessment was based on the BLEU (Bilingual Evaluation Understudy) score, which quantifies the similarity between machine-generated outputs and reference responses. The reference responses were carefully crafted by travel domain experts to serve as the gold standard for evaluation.



```
                BLEU-1
                  mean       std       min       max
LLM
Claude 3.7    0.325076  0.147680  0.195513  0.485876
GPT 4         0.198800  0.077915  0.150833  0.288702
Travel_GPT    0.248050  0.105168  0.127384  0.320215
```

Figure 5: Comparative BLEU Score Analysis of TravelGPT against Claude 3.7 Sonnet and GPT-4.

4.2 Comparative Performance Analysis

Results from the BLEU score analysis demonstrate that the TravelGPT architecture achieves competitive performance in the travel domain. Notably, our model outperforms GPT-4 and approaches the performance level of Claude 3.7 Sonnet, despite the latter utilizing substantially greater computational resources and benefiting from more extensive training. This finding highlights the efficiency and effectiveness of our architecture, which delivers comparable results with significantly more constrained resources and a more focused training approach.

4.3 Resource Efficiency

The performance of TravelGPT is particularly noteworthy when considering the computational efficiency of our architecture. While Claude 3.7 Sonnet and GPT-4 require extensive computational resources for inference, our specialized architecture achieves comparable results with significantly reduced resource requirements. This efficiency can be attributed to several factors in our design, including the domain-specific training approach, the multi-agent architecture that partitions complex tasks, and the optimized RAG pipeline that focuses on retrieving only the most relevant contextual information.

5 Conclusion and Future Work

5.1 Summary of Contributions

This research presents a specialized AI architecture for providing culturally sensitive and factually accurate guidance for Sikh tourism in Pakistan. Our TravelGPT system demonstrates how domain-specific knowledge integration and multi-agent architecture can effectively address cultural and geographical biases inherent in general-purpose LLMs. Evaluation results confirm that our approach achieves competitive performance against leading models like Claude 3.7 Sonnet and GPT-4 while requiring significantly fewer computational resources.

5.2 Limitations and Future Work

While our implementation has successfully demonstrated the viability of the multi-agent architecture, several limitations remain. Currently, only half of the planned specialized agents have been fully implemented, primarily the Heritage Retrieval Agent and Summarizer Agent. The system's capabilities are also predominantly in English, limiting accessibility for diverse pilgrims.

Future work will focus on four key areas: (1) multilingual expansion to provide comprehensive support for Punjabi and Urdu; (2) development of an itinerary generator agent to create optimized pilgrimage routes balancing religious significance with practical considerations; (3) integration of real-time information on visa requirements and local conditions; and (4) implementation of structured user feedback mechanisms to continuously improve the knowledge base.

Through these enhancements, we aim to create a comprehensive digital companion for religious tourism that combines cultural understanding with practical travel assistance, making significant heritage sites more accessible while preserving their spiritual importance.

References

Maheen Farooqi Abubakar Abid and James Zou. 2021. Persistent anti-muslim bias in large language models. *arXiv preprint*.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2023. Self-rag: Learning to retrieve, generate, and critique through self-reflection. *arXiv preprint arXiv:2310.11511*.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623.

Patrick Lewis et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *arXiv preprint*.

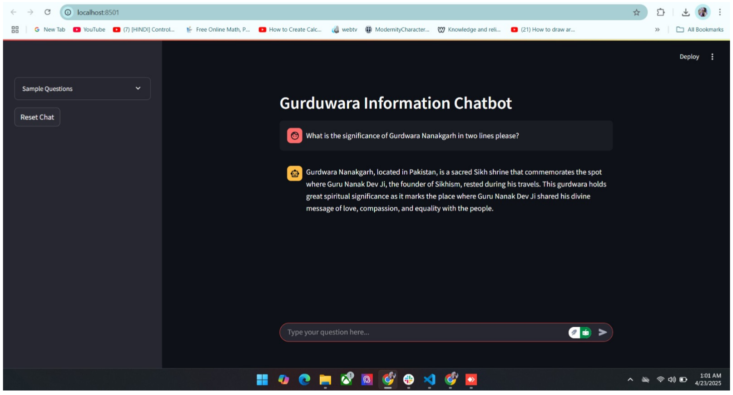
Rishi Bommasani et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint*.

Wenhao Xu et al. 2022. Retrieval-augmented language models. *arXiv preprint*.

A System Interface and Example Interactions

A.1 User Interface Design

The TravelGPT system was implemented as a web-based conversational interface using Streamlit.

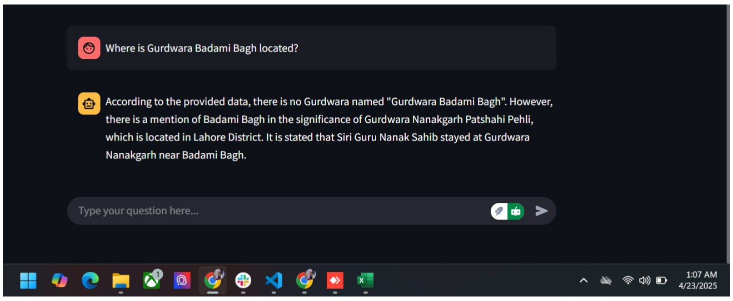


Gurduwara Information Chatbot

What is the significance of Gurdwara Nanakgarh in two lines please?

Gurdwara Nanakgarh, located in Pakistan, is a sacred Sikh shrine that commemorates the spot where Guru Nanak Dev Ji, the founder of Sikhism, rested during his travels. This gurdwara holds great spiritual significance as it marks the place where Guru Nanak Dev Ji shared his divine message of love, compassion, and equality with the people.

Figure 6: TravelGPT conversational interface showing the chat history panel, query input field, and sample question suggestions related to Sikh heritage sites in Pakistan.



Where is Gurdwara Badami Bagh located?

According to the provided data, there is no Gurdwara named "Gurdwara Badami Bagh". However, there is a mention of Badami Bagh in the significance of Gurdwara Nanakgarh Patshahi Pehli, which is located in Lahore District. It is stated that Siri Guru Nanak Sahib stayed at Gurdwara Nanakgarh near Badami Bagh.

Figure 7: Example of TravelGPT response to a query

Key features of the interface include persistent conversation history for contextual follow-up questions, sample query suggestions for first-time users, support for complex multi-part questions about travel logistics and site significance, clear attribution of information sources for transparency, and an option to export conversation for offline reference during travel.