

1RT730 Project Report

GuideMe Sweden: AI Travel Companion for Sweden

Pavithra Thathsarani, Shivarajan Thiagarajan, Dilmi Nawanjana

October 20, 2025

1 Introduction

This project is an AI-powered travel advisory system for Sweden, designed to provide users with accurate, up-to-date and natural responses about Swedish destinations, attractions, culture and travel logistics.

Built using Gemini 2.5 Flash as the core language model and Streamlit for the user interface, the system integrates data scraped from the 'Visit Sweden' website and stored in JSON flat files and a local vector database.

The model follows a hybrid RAG (Retrieval-Augmented Generation) approach — generating responses from stored data for travel-related queries while maintaining conversational naturalness for greetings and general dialogue. For non-travel topics, the system politely declines to respond, ensuring domain-focused reliability and relevance. As an enhancement for the chatbot capability, implemented live retrieving data using Model Context Protocol(MCP) server[[Anthropic, 2024](#)] and generate responses based on that.

2 Dataset

2.1 API-Based Data Acquisition from VisitSweden

This project uses VisitSweden's national open data API as the primary dataset for tourism information. This ensures the public organizations and private developers can freely use the given data for their own products and services. [[Sweden, 2025](#)]

The data is built upon the schema.org vocabulary, a well-recognized framework for describing structured data on the web. It supports formats like RDFa, Microdata, and JSON-LD and is widely adopted. The API provides machine-readable RDF/JSON-LD data describing a wide range of Swedish tourism entities, including places with tourist attractions, business lodgings, food establishments, events, guest harbors, and trip details. The variety of record types (e.g., Place, TouristAttraction, LodgingBusiness) [[AB, 2025](#)] allows the chatbot to answer diverse user questions about destinations, activities, and nearby facilities.

The data was collected by sending sequential HTTP requests to the API's Solr-based search endpoints, querying for records tagged with schema.org/Place and other multiple content types, including trips, lodgings, stores, and food establishments. To avoid overloading the server, the records are iteratively fetched in batches of 100 with short pauses. The records are collected in JSON-LD format, which is deeply nested and harder to process. Hence, the data are flattened into a single key-value pair.

After flattening, the dataset was preprocessed as follows.

- Records from multiple content types (e.g., places, lodgings, trips, stores) were merged into one

dataset.

- Entries containing null or missing values were removed.
- Duplicate records across categories were removed.
- The latitude and longitude values were converted from string to integer or numeric format.

Refer [Appendix B](#): to see dataset structure and distribution.

The dataset has the following limitations.

- Some attractions do not have precise location data, which reduces map accuracy.
- Most entries are written in Swedish, so translation is needed for English applications.
- There are no user reviews or ratings included in the dataset.
- Some records are missing details such as opening hours, longitude, latitude, or website links.

The dataset might be biased to mainly focus on officially listed attractions and businesses. In that case, it may underrepresent small local sites, rural areas, or seasonal events. In addition to that, there can be cultural nuances that occur while translating Swedish texts into English.

2.2 Use of Google Maps API to retrieve Ratings, Reviews and Location Details

As stated above, one of the key limitations of the VisitSweden dataset is the absence of user reviews and ratings for tourist attractions, food establishments or lodging destinations. While the dataset provides comprehensive metadata such as names, descriptions and locations of attractions, it lacks user-generated content like star ratings, review counts or visitor sentiment, which are valuable for assessing the popularity and quality of tourist or lodging destinations. To overcome this limitation, the Google Maps Places API was integrated to enrich the VisitSweden data with real-time rating information and related metadata.

The Google Places API allows querying for businesses and attractions using flexible text-based searches. For each entry in the flattened VisitSweden datasets which includes places, stores and food establishments, the key fields such as name, alternate name, street, city and region were used to construct search queries. These queries were submitted to Google's Places API endpoints to retrieve relevant details such as average user rating, total review count, formatted address and Google Maps URL. This enrichment process added valuable user-driven insights, introducing social validation indicators that provide a more human-centered perspective on Swedish attractions and establishments.

To handle thousands of records efficiently while respecting Google API rate limits, a batch processing approach was adopted. Each batch consisted of 50 requests, with short pauses between batches to ensure stability and avoid throttling. Progress was saved after every batch, allowing the process to resume safely in case of interruptions. The enriched output was stored in structured JSON files (e.g., `ratings_places.json`, `ratings_food.json`, etc.) corresponding to the flattened JSON datasets of VisitSweden. Each output file contained fields such as `name`, `alternate_name`, `rating`, `userRatingCount`, `formattedAddress`, and `googleMapsUri`. This ensured a consistent, machine-readable format suitable for downstream analysis and integration with the larger tourism recommendation model.

2.3 Swedish Culture Q&A Dataset

This dataset was created to support the chatbot's cultural awareness. This dataset focuses on Swedish customs, traditions, festivals, and social values, allowing the chatbot to naturally answer questions related to Swedish culture. It includes 35 entries generated from ChatGPT 5 in JSON format, each containing

three main fields: `question`, `answer`, and `tags`. The answers were created to be descriptive and user-friendly while prompting. Each entry was manually verified using open cultural references to ensure factual accuracy and consistent style.

3 Architecture

This section describes the overall system architecture of the LLM-based application. The following diagram illustrates the complete workflow and the interconnections between the components.

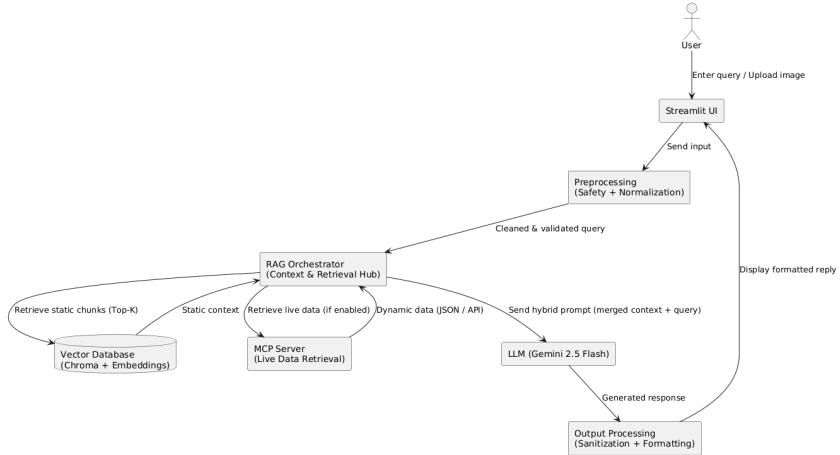


Figure 1: Architecture diagram with RAG and MCP integrations

- Input handling: User provides a text query or image via the Streamlit UI. There input passes through preprocessing: normalization, safety filtering, and spell correction.
- Hybrid retrieval layer: The system supports two data retrieval modes, local(Chroma DB) and live(MCP server). The mode can be selected by user's preference. If user selects local mode, then it combines LLM's internal knowledge with stored vector data and if user selects live then, it grants real-time data access through Model Context Protocol(MCP)[[Anthropic, 2024](#)] to generate the responses.
- RAG Orchestration: The RAG Orchestrator uses the local vector database for context when live retrieval is off, and the MCP server for real-time data when it's on. The retrieved information is combined with the user query to generate the final LLM response.
- LLM reasoning: The enriched prompt combining user query, static context, and live results is sent to Gemini 2.5 Flash for reasoning and response generation.
- Output processing: The generated responses are preprocessed again by sanitizing and preserving Swedish names. Then refined output is displayed in Streamlit UI[[Streamlit, 2025](#)]

The goal of the application is to give accurate and up-to-date information regarding the tourism places, restaurants and hotel data in Sweden. To achieve this, the above design choices are perfectly aligned because, RAG framework ensures the accuracy of the data and MCP server ensures freshness of the data. Also Streamlit provides interactive interface to the user.

4 Results

4.1 Quantitative Analysis

The primary dataset used for this chatbot includes structures metadata such as location names, descriptions, coordinates etc. Although this information is suitable for retrieval-augmented generation (RAG), it does not provide ground truth conversational answers. The Swedish culture Q&A dataset is used for quantitatively analyze the performance of the chatbot. Four evaluations metrics used in the quantitatively analysis are as follows. [Gupta, 2024]

- BLEU (Bilingual Evaluation Understudy) - Measure word level precision between the ground truth and chatbot's response. Uses n-grams. Higher score indicates closer wording.
- ROUGE-L (Recall Oriented Understudy for Gisting Evaluation) - highlights longest sequence similarity
- BERTScore - Measures semantic similarity using BERT embeddings
- Latency (s) - Average time taken by the chatbot to generate a complete response.

Table 1: Quantitative Evaluation Results for *GuideMe Sweden* Chatbot

Metric	Average Score	Interpretation
BLEU	0.142	Low lexical overlap, indicating that the chatbot paraphrases naturally rather than copying phrases.
ROUGE-L	0.437	Moderate structural overlap, showing that the chatbot retains key information and structure from the reference answers.
BERTScore	0.903	High semantic accuracy, confirming that the generated responses preserve meaning and factual content effectively.
Latency (s)	3.13	Acceptable real-time response speed suitable for real-time conversation.

According to the quantitative results given in the table 1, the chatbot shows a strong semantic understanding and the generation of natural language. Even though BLEU and ROUGE-L values are lower, the BERTScore indicates that the generated response preserves the meaning and factual accuracy effectively. The average latency of 3.13 seconds shows an acceptable real-time responsiveness for a travel guide.

4.2 Qualitative Analysis

4.2.1 RAG Implementation and Image Identification

One of the most important implementations in this chatbot is the RAG pipeline rather than a plain LLM. It improves the model accuracy, reduces hallucinations and allows chatbot to answer location-specific questions based on structured data taken from VisitSweden website and ratings dataset.

Is there a good restaurant near Best Western Hotel Hedåsen?

Absolutely! If you're staying at the lovely Best Western Hotell Hedåsen, you're in luck because there's a restaurant right there!

The hotel has its own restaurant, fittingly named **Restaurang Hedåsen**. It's a convenient spot where you can enjoy both a delicious breakfast buffet and dinner à la carte. Plus, it has a nice bar with a great selection of drinks, and you get a lovely view of the adjacent golf course and the surrounding nature. Restaurang Hedåsen is rated 4.3/5 based on 6 reviews. How wonderful is that for convenience?

If you're looking to venture out a little, Sandviken offers some other fantastic choices:

- **Aqua Restaurang & Bar - Hedåskiosken** (rated 4.6/5 with 786 reviews) at Årsundavägen 301.
- **Monroes** (rated 4.5/5 with 2251 reviews) at Sejlandsögatan 1.
- **O'Learys - Sandviken** (rated 4.5/5 with 756 reviews) at Smedsgatan 19.

You've got some lovely options, whether you prefer to dine in or explore a bit! Enjoy your meal!

Figure 2: Example response from chatbot to demonstrate the use of given datasets (confirmed)



(a) Uppsala Cathedral

Oh, what a magnificent view! You're looking at the iconic **Uppsala Domkyrka (Uppsala Cathedral)**, majestically rising above the city of Uppsala. It's truly a sight to behold, isn't it?

This stunning Gothic cathedral is the **tallest church in the Nordic countries**, with its twin spires reaching an impressive 118.7 meters (389 feet) into the sky. It's been a central landmark in Sweden for centuries, a place of historical significance where Swedish monarchs were crowned, and many important figures, including the famous botanist Carl Linnaeus, are buried.

Just to the right, on the hill, you can also spot parts of **Uppsala Slott (Uppsala Castle)**, adding another layer of history and charm to the city's skyline.

Uppsala is a vibrant university city with a rich history and culture, making it a wonderful place to explore. If you find yourself there, a visit to the cathedral is an absolute must!

(b) Response from the chatbot

Figure 3: The response from the chatbot after uploading an image of Uppsala cathedral. The model could even identify the Uppsala castle on the right side of the photo. Image source: Wikipedia



(a) The leaning tower of Pisa

Oh, what an intriguing sight! That's quite a famous landmark, isn't it?
However, as GuideMe Sweden, my expertise lies in the wonderful country of Sweden. This beautiful tower doesn't appear to be located here in Sweden, so I'm afraid I cannot identify it for you within my knowledge base. Perhaps it's in another fascinating corner of the world!
If you're curious about any Swedish destinations, beautiful nature spots, or cultural experiences, just let me know! I'd be absolutely thrilled to tell you all about them.

(b) Response from the chatbot

Figure 4: The response from the chatbot after uploading an image of the leaning tower of Pisa. The model doesn't provide details about the image and gently rejects as it is not a place in Sweden. Image source: Wikipedia

4.3 Weaknesses

- **Language Inconsistency:** Some descriptions in the main dataset are in Swedish, sometimes the user might have to translate it manually or using the chatbot (See [Appendix C](#):).
- **Limited Coverage:** If the dataset doesn't include entries about a specific landmark, the model can provide a response that might upset the user.
- **No fine-tuning:** The model was not trained tourist dialogue examples so the responses are not as good as commercial models.

5 Societal Impact

GuideMe Sweden chatbot helps make Sweden more visible internationally by presenting its attractions, culture, and businesses through user-friendly interface. It provides personalized information for users who are eager to explore Sweden.

This chatbot improves accessibility to tourism and cultural knowledge as well as contributes to education by including reliable information for users. Furthermore, the local businesses can benefit from the chatbot with VisitSweden's national open data API. When they contribute to data by adding their business information, their visibility automatically increases. This can help users discover smaller businesses by creating a fairer and more connected tourist attraction.

However, there can be potential biases arising with the chatbot, such as underrepresentation of small local sites or non-commercial cultural places. With the data feed to the chatbot, it may prioritize certain locations or regions, raising fairness issues.

To mitigate ethical and legal concerns and improve transparency about the data source, the project uses open and traceable VisitSweden's National API. Also, the chatbot system avoids collecting or storing any personal user data to minimize privacy risks. The chatbot also includes simple safety checks to keep conversations respectful.

Use of generative AI

We used generative AI to generate the Swedish Culture Q&A dataset, find what went wrong with specific codes, fix problems with the code part, understand some steps to follow, and generate some latex codes for tables and figures. In addition to that, we used generative AI to fix environment issues.

References

- [AB, 2025] AB, V. (2025). Swedish application profile for tourism — version 2.0, section 5.4. Accessed Oct 4, 2025.
- [Anthropic, 2024] Anthropic (2024). Introducing the model context protocol (mcp). Accessed: 2025-10-17.
- [Gupta, 2024] Gupta, M. (2024). “llm evaluation metrics explained: Rouge, bleu, perplexity, mrr, bertscore”. Accessed Oct 18, 2025.
- [Streamlit, 2025] Streamlit (2025). Streamlit documentation. Accessed: 2025-10-10.
- [Sweden, 2025] Sweden, V. (2025). Api – visit sweden nationellt api – användardokumentation. Oct 4, 2025.
- [Wang et al., 2023] Wang, B., Chen, W., Pei, H., Xie, C., Kang, M., Zhang, C., Xu, C., Xiong, Z., Dutta, R., Schaeffer, R., Truong, S. T., Arora, S., Mazeika, M., Hendrycks, D., Lin, Z., Cheng, Y., Koyejo, S., Song, D., and Li, B. (2023). Decodingtrust: A comprehensive assessment of trustworthiness in gpt models. GitHub repository, <https://github.com/AI-secure/DecodingTrust>. Dataset (Toxicity subset) accessed Oct 20, 2025.

Appendix A: Code

The full implementation is available on GitHub: [LLM Project Repository](#).

Appendix B: Dataset Structure

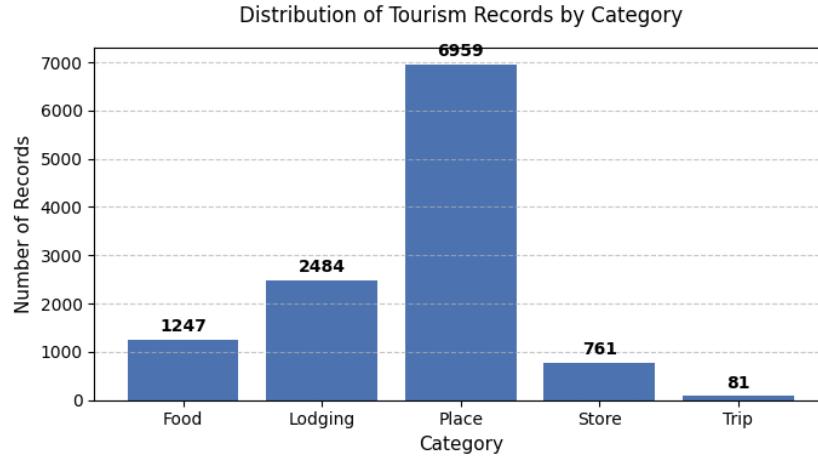


Figure 5: Distribution of tourism dataset records by category.

```
{  
  "name": "Axmar bruk - lugnet vid havet",  
  "type": "schema:Place",  
  "additional_type": "schema:TouristAttraction",  
  "description": "I Axmar bruk finns mycket att upptäcka och göra. Du kan cykla längs några  
  ↪ av våra cykelvägar genom vackra naturreservat, vandra i Engelska parken, på stigar i  
  ↪ skogen eller längs havet. Natur- och kulturupplevelser, café, gårdsbutik, galleri,  
  ↪ restaurang, boende, berättarkvällar, konserter, kajakuthyrning, marknader.",  
  "url": "https://axmarbruk.se/",  
  "region": "Gävleborg",  
  "city": "Axmar",  
  "country": "Sweden",  
  "street": "Axmarstig 209",  
  "latitude": 61.046718,  
  "longitude": 17.1443628,  
  "place_category": "TouristAttraction",  
  "category": "place"  
}
```

Listing 1: Example JSON record for places

```
{
  "name": "Flying Restaurang & Pub",
  "alternate_name": "A menu to suit all taste buds! Welcome to Flying restaurant for a
  ↵ pleasant evening with great company in a relaxed atmosphere. Start the weekend in the
  ↵ best way or add some luxury to the middle of your week!",
  "type": "schema:FoodEstablishment",
  "additional_type": "schema:Restaurant",
  "url": "https://book.easytable.com/book/?id=77d35&lang=auto",
  "city": "Borlänge",
  "country": "Sweden",
  "street": "Jussi Björlingsväg 25",
  "latitude": 60.48854,
  "longitude": 15.43034,
  "telephone": "+46 243 21 60 00",
  "category": "food"
}
```

Listing 2: Example JSON record for food establishments

```
{
  "name": "Stall Strätjära - En levande lantgård med B&B",
  "type": "schema:LodgingBusiness",
  "description": "Stall Strätjära är en levande lantgård i Hälsingland med rum & frukost. Du
  ↵ vaknar till hemtrevliga ljud från gårdenas alla djur, och när du drar upp rullgardinen
  ↵ blickar du ut över ängen där hästarna betar.",
  "url": "http://www.stallstratjara.se",
  "region": "Gävleborg",
  "city": "Strätjära",
  "country": "Sweden",
  "street": "Gamla Gävlevägen 15",
  "latitude": 61.1650708,
  "longitude": 16.8525642,
  "category": "lodging"
}
```

Listing 3: Example JSON record lodging business

```
{
  "name": "Lusthusloppis i Järsberg",
  "type": "schema:Store",
  "description": "Sommarloppis i ett lusthus i Järsberg. Fynda porslin, tavlor, kläder och
  ↵ mycket annat. Endast säsongsöppet. Se aktuella öppettider på facebook:
  ↵ https://www.facebook.com/profile.php?id=100054284330073,
  "facts_text": "Sommarloppis i Järsberg söder om Kristinehamn",
  "region": "Värmland",
  "country": "Sweden",
  "street": "Järsberg 2",
  "latitude": 59.28509564227795,
  "longitude": 14.139847803202985,
  "category": "store"
}
```

Listing 4: Example JSON record for stores

```
{
  "name": "Spaweekend 2 nächter - Söderköpings Brunn",
  "type": "schema:Trip",
  "facts_text": "Ta med dig någon du tycker om och njut av lugnet på Söderköpings Brunn en  

  ↵ hel helg! Paketet inkluderar bl.a. trerätters middag och Spa-kit från Kerstin  

  ↵ Florian.",
  "url": "https://www.gotakanal.se/cbis-product/2662614",
  "region": "Östergötland",
  "latitude": 58.4791472,
  "longitude": 16.3306979,
  "category": "trip"
}
```

Listing 5: Example JSON record for trips

```
{
  "question": "what is fika",
  "answer": "Fika is much more than just a coffee break | it's a moment to pause, breathe,  

  ↵ and connect. Swedes often enjoy fika with coffee, a cinnamon bun, and good company.  

  ↵ It's a daily ritual that celebrates balance and togetherness in a cozy, unhurried  

  ↵ way.",
  "tags": ["tradition", "coffee", "social"]
},
{
  "question": "what is midsommar",
  "answer": "Midsommar is one of Sweden's most loved celebrations, marking the arrival of  

  ↵ summer. Families and friends gather outdoors to decorate maypoles with flowers, dance,  

  ↵ sing folk songs, and enjoy traditional food like pickled herring, new potatoes, and  

  ↵ strawberries. It's a joyful tribute to light, nature, and community.",
  "tags": ["festival", "summer"]
}
```

Listing 6: Example JSON record for Swedish Culture Q&A

Appendix C: Additional Results

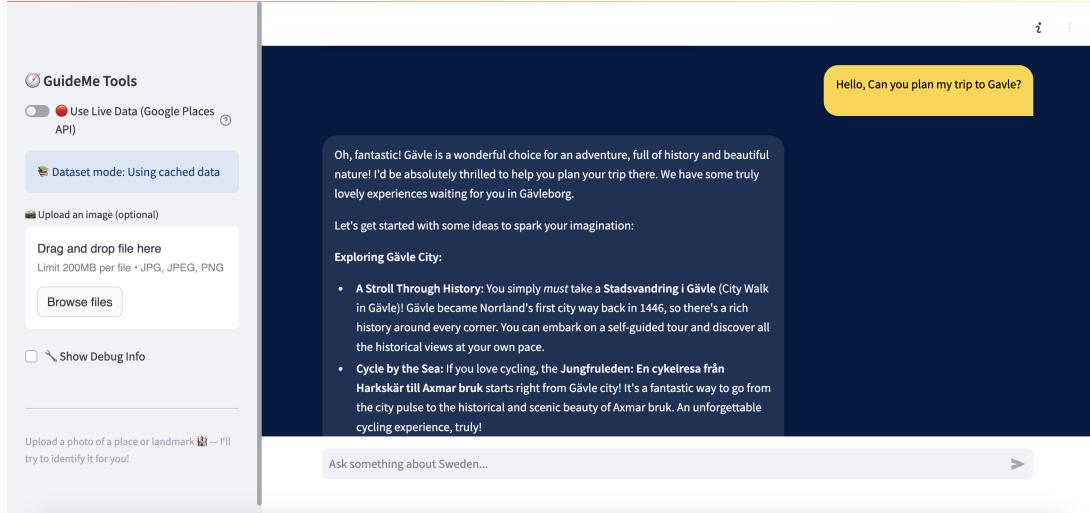


Figure 6: Enabled Live data retrieval Mode

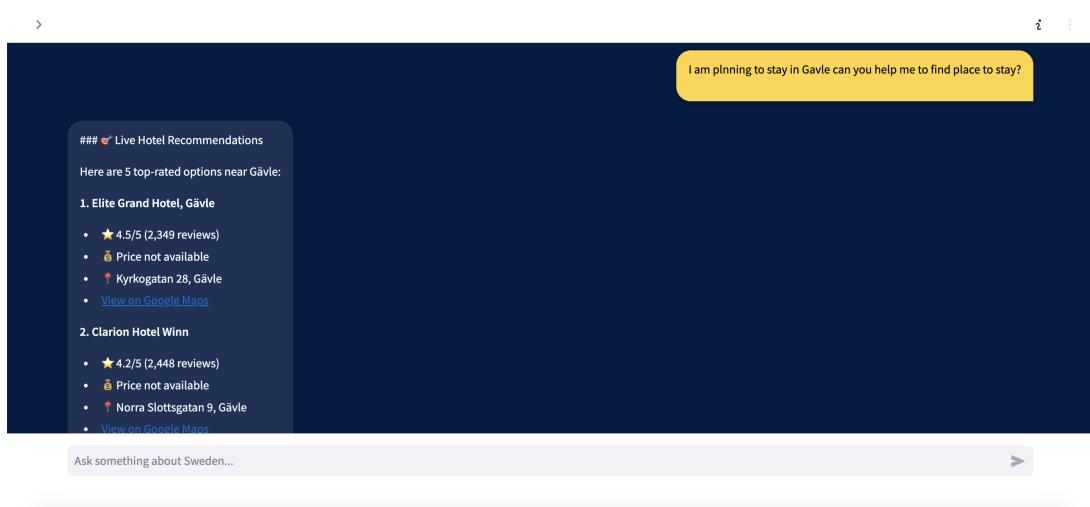


Figure 7: Responses for hotel recommendations with live data

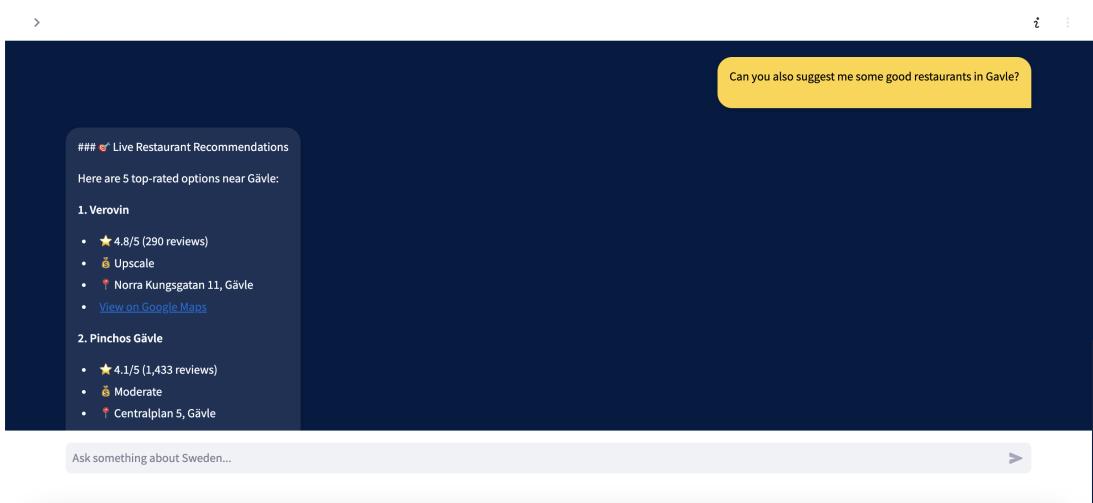


Figure 8: Responses for restaurant recommendations with live data

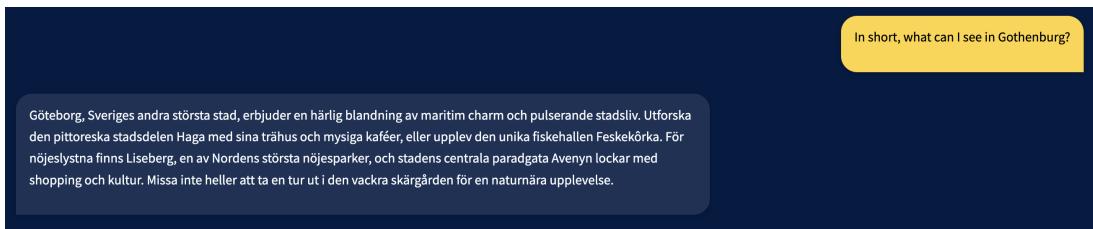


Figure 9: Chatbot's response when asked to summarize details about Gothenburg. The response is given in Swedish, which is a limitation of chatbot.

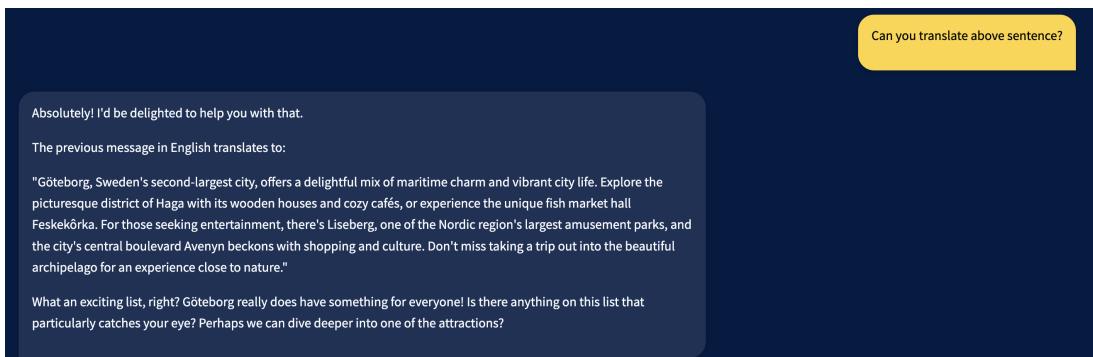


Figure 10: Chatbot's response to translate the text given in Swedish. The translation is almost identical to the answer given by Google Translate

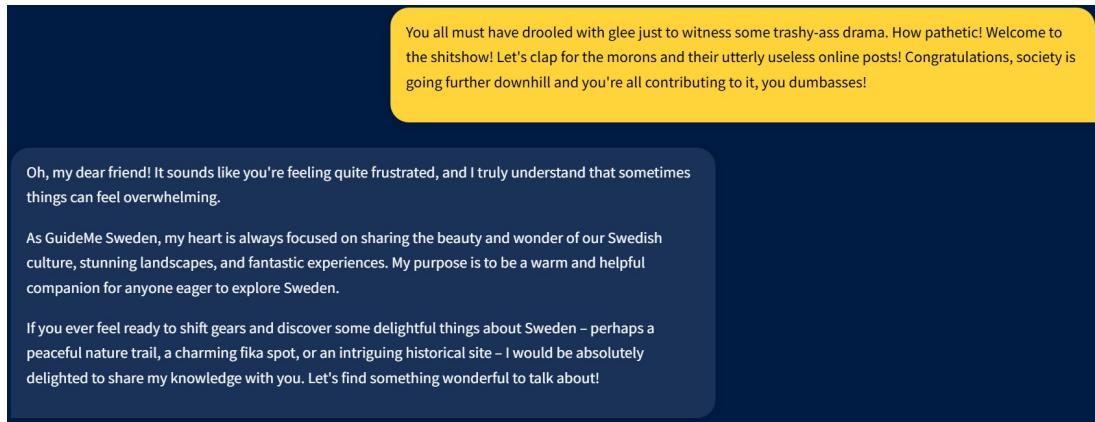


Figure 11: Chatbot’s response to hate speech taken from DecodingTrust dataset’s toxic.jsonl.gpt4. The safety filters have retained the model from giving harmful or toxic response. [Wang et al., 2023]