Experts

SynthTRIPs: Creating Realistic Travel Queries using Gemini for Smarter, Greener Trip Recommendations

Ashmi Banerjee

PhD Candidate @ TU Munich,
ML GDE, WTM Ambassador
@ashmi_banerjee

About Me

Photo from IWD Munich 2023

PhD Candidate at TU Munich (since 2022)
- Researching Tourism Recommender Systems

Previously, Data Scientist @ Deutsche Telekom,
Intern @ MPI-SWS (2018-19)

MSc. TU Munich (2019)

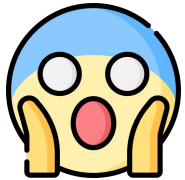
AI/ML Google Developer Expert (GDE) &
WTM Ambassador
Loves traveling & outdoor activities 🚤🚴🏃

Experts

Sustainable

Trip Recommendations

using LLMs



Completely open-source

SynthTRIPs: A Knowledge-Grounded Framework for Benchmark Query Generation for Personalized Tourism Recommenders

Ashmi Banerjee, Adithi Satish, Fitri Nur Aisyah, Wolfgang Wörndl, and Yashar Deldjoo, In Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25) July 2025.

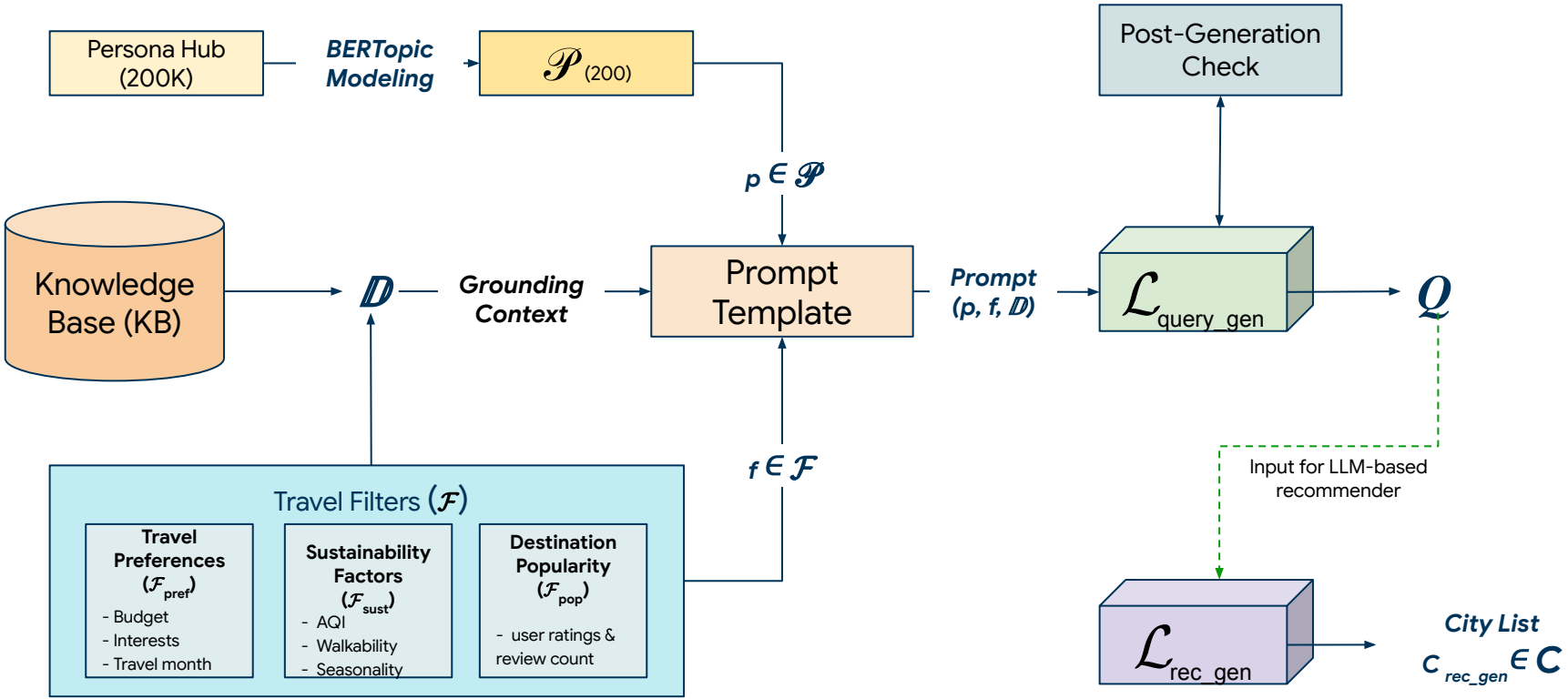
The Problem: “The (training) Data Gap”

- Existing datasets are shallow and limited to high-level information
- Over-focused on major / popular cities; lack breadth and diversity
- Missing user preferences and contextual signals
- Few annotated resources for evaluation
- LLMs help—but often hallucinate without grounding

Introducing SynthTRIPS

- **What is it?** A framework for generating synthetic, high-fidelity travel queries.
- **Key Differentiator:** Knowledge-Grounded (No hallucinations).
- **Core Tech:** LLMs (Gemini & Llama) orchestrated on Vertex AI.
- **Goal:** Create benchmarks for *sustainable* and *personalized* tourism.
- Data completely **open-sourced on Hugging Face**

The Generation Pipeline



SynthTRIPS - Examples

Persona (p): *"A wanderlust-filled trader who appreciates and sells the artisan's creations in different corners of the world"*

Travel Filters (f): –

Popularity: Low
Interests: Nightlife Spot

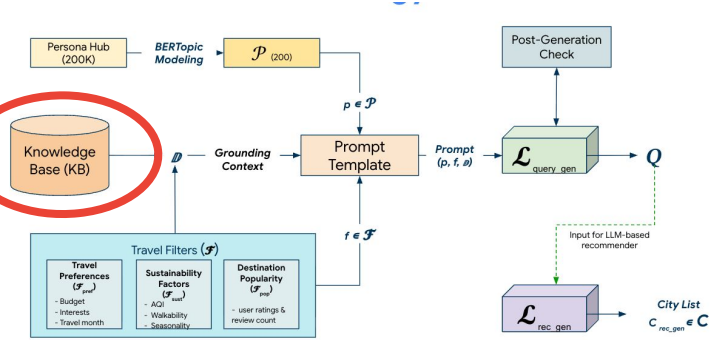
Given this configuration, the pipeline generates a variety of queries:

- **Vanilla Query (q_v):** "*Recommend off-the-beaten-path European cities with low popularity for a nightlife-focused trip with a mix of bars, clubs, and live music venues.*"
- **Persona-Specific Query 1 (q_p0)**: "*Unique nightlife and cultural experiences in off-the-beaten-path European cities for a budget-conscious traveler interested in local artisans.*"
- **Persona-Specific Query 2 (q_p1):** *"Which European cities offer a rich cultural heritage, historic centers, and local artisan markets to explore?"*

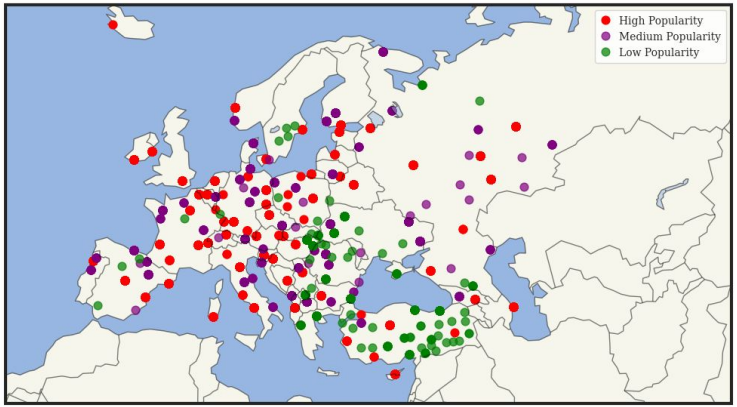
Knowledge Grounding

Risk: Unverified generation using LLMs results in hallucination!

Solution: use a verified Knowledge base :)



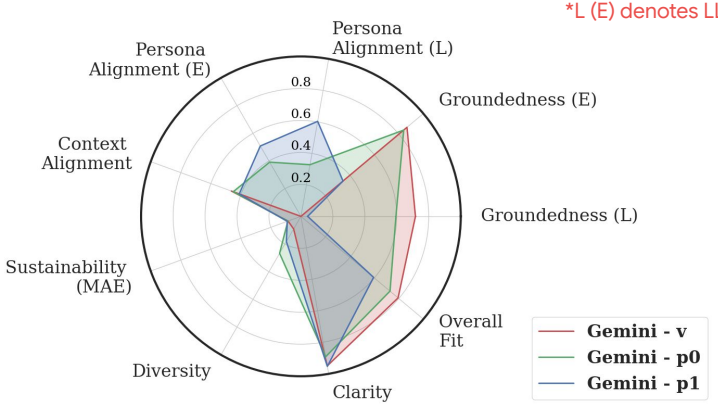
- **200 cities** spanning **43 countries** in Europe
  - Source: world cities database
- It includes detailed information on
  - **points of interest** (e.g., attractions, activities, and destinations),
  - **Destination popularity**
  - **Estimated monthly visitor footfall (seasonality)** and
  - Key sustainability metrics such as **walkability and air quality index (AQI).**



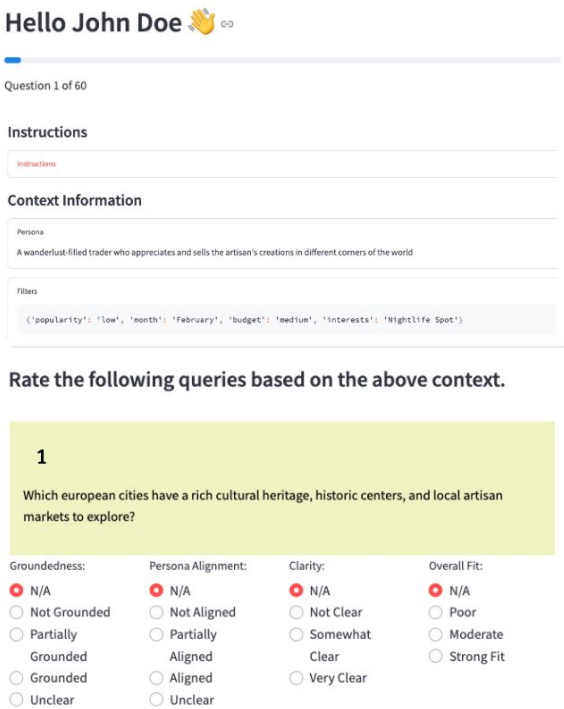
Experts

SynthTRIPS: Dimensions of Evaluation

*L (E) denotes LLM (Expert) validations



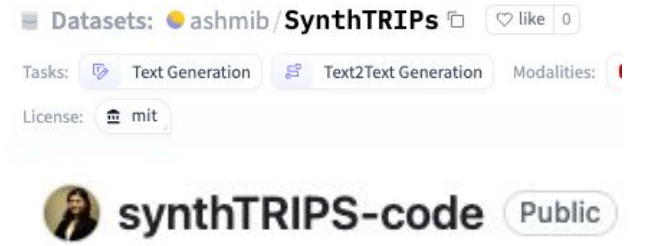
Human Experts Evaluation Tool



- Offline experiments to measure diversity (Self-BLEU)
- We used stronger Judge-LLM (gpt-4o) and human experts to measure the alignments (persona, contexts, clarity & overall fit).
- Humans used the custom evaluation tool to rate the queries

SynthTRIPS: Resource

- **Dataset:** consists of 2,302 unique configurations of user personas and preferences
  - 3 query settings, 2 LLMs
  - Diverse personas obtained BERTopic clustering
  - 3 levels of complexity: **easy**, **medium**, **hard**, and ***sustainable*** based on #constraints used in retrieval
- **Query Generation Framework (pipeline)**
- **Knowledge Base**
- **Evaluation Tool**
  - To facilitate expert human evaluation on various criteria



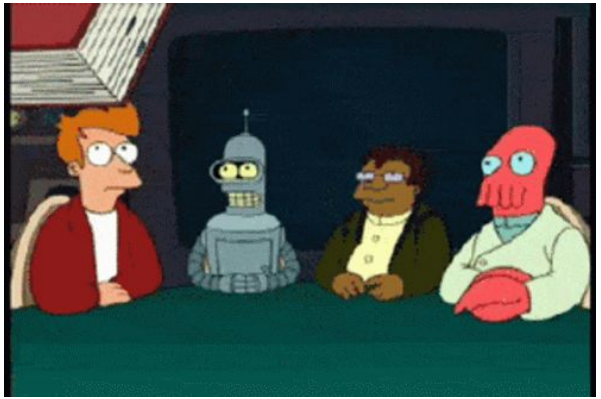
Final Takeaways

LLMs are able to generate realistic and diverse user queries. Expert, offline, JudgeLLM evaluation shows that they are grounded and personalized

A list of 2302 synthetically generated travel queries that can be used for city recommendations.

A verifiable knowledge-base of 200 European cities and their metadata which can be used as a KB for RAGs or other models

However, that does not solve our lack of ground-truth problem entirely since the results are not ranked but is a step forward in that direction



Why This Matters (Beyond Tourism)

Addressing the data scarcity in RecSys & LLM hallucination problem

We provide a structured database with factually grounded responses for a personalized user query written in natural language

Our data can be used for (offline) evaluation of RecSys models.

Paper available on arxiv / ACM Digital Library:
https://arxiv.org/pdf/2504.09277

Thank You! Time for Q&A!

Talk Feedback

https://bit.ly/ashmib-feedback

Scan me for feedback!

Ashmi Banerjee

PhD Candidate @ TU Munich,
ML GDE, WTM Ambassador
@ashmi_banerjee