



Cocodrillo: A Retrieval-Augmented Conversational Agent for Smart Travel Guidance

Davide Belcastro, Constance Monluc, Ondra Pritel

Abstract

Cocodrillo is a multilingual, retrieval-augmented chatbot engineered to assist users in planning safe and informed travel experiences. Unlike traditional chatbots, it combines real-time data retrieval from web sources with curated datasets to answer questions about local safety, weather alerts, cuisine, events, transportation, and sightseeing. The system integrates a compact BERT-based intent classifier, a QA module for entity resolution, and a hybrid retrieval-summarization architecture. Its itinerary planning engine employs graph optimization under time constraints, delivering personalized travel suggestions. The system emphasizes robustness, linguistic adaptability, and user-centric interactions.

Keywords

Conversational AI, Retrieval-Augmented Generation, Travel Assistant, Intent Classification, Route Optimization

Advisors: Slavko Žitnik

Introduction

Modern travelers seek real-time, tailored information to plan their trips safely and enjoyably. While traditional chatbots operate on static knowledge, Cocodrillo introduces a dynamic, retrieval-augmented architecture capable of integrating current data from web sources and APIs. It assists users with travel decisions involving weather, safety, events, restaurants, attractions, and logistics. The system blends BERT-based intent detection, entity resolution, summarization, and optimization techniques to construct relevant, context-aware responses.

Structure

The system recognizes 8 main intents:

- **Safety updates for a location (any city)** – “Is it safe?”
Sources: Bing News + Viaggiare Sicuri
- **Weather alerts and extreme weather conditions (any city)**
Sources: Bing News
- **Recommended places to visit (explained later)**
Data: Dataset downloaded from Lonely Planet, with the help of Python scripts based on Selenium and BeautifulSoup.
Integration: Uses Google Maps to check the availability

of locations (for example, to know if a place is temporarily closed).

- **Information on concerts and events (any city)**
Sources: Bandsintown
- **Best restaurants to eat (any city)**
Sources: Yelp
- **Recommendations on typical dishes or foods (dataset downloaded only across all of Europe)**
Data: Dataset downloaded from TasteAtlas, using Python scripts with Selenium and BeautifulSoup.
- **Future weather forecasts (any city)**
Sources: Il Meteo
- **Information on trains, flights, and buses (any city)**
Sources: TheTrainLine

Methods

Our end-to-end pipeline integrates intent classification, entity extraction, and retrieval of real-time data from heterogeneous sources. The system comprises the following components:

Intent Classification

We employ the `all-MiniLM-L6-v2` BERT encoder fine-tuned for multi-class intent detection across eight categories

(e.g., safety updates, weather alerts, travel logistics). This lightweight model encodes user queries into embeddings, which are subsequently passed through a softmax classifier to infer the most probable intent.

Entity Extraction and Clarification

To identify locations, dates, and other key entities, we use a BERT-based question-answering module (`bert-large-uncased-whole-word-masking-finetuned-squad`). Absent or ambiguous information (e.g., missing travel dates) triggers a clarification subroutine: the system formulates follow-up questions to the user, ensuring complete query parameters. Spelling and formatting errors are corrected via semantic similarity checks against gazetteer and date lexicons.

Check for Date, City, and Error Handling

The system verifies that, depending on the predicted intent, the corresponding information regarding dates and cities is provided. If any required information is missing, the system returns an empty list (specific tests have been conducted for this subtask, yielding excellent results). Additionally, there is a built-in spell-check feature. It is not necessary to input the city or date (either in a standard date format or in a more casual form, such as "today" or "tomorrow") perfectly; small typos are tolerated, and the system will still be able to understand the input.

Data Retrieval and Integration

Depending on the classified intent, the system consults:

- **Bing News and Viaggiare Sicuri** for safety and weather alerts in real time.
- **Lonely Planet and TasteAtlas** (via Selenium & BeautifulSoup) for attractions and local dishes.
- **Bandsintown** for concerts and events.
- **Yelp** for restaurant recommendations.
- **Il Meteo** for extended weather forecasts.
- **TheTrainLine** for train, flight, and bus schedules.
- **Google Maps** to verify real-time availability and status of points of interest.

System Output

Output structure:

- **First Response:**
An introductory sentence generated with a **bi-grams model** based on a simple dataset, explaining that the system is searching for information.

Hello, this is the information about the typical food you asked me for.

- **Second Response:**
A structured output of all the information found.

Insights

News Summarization and Relevance Filtering

For safety-related queries, the system:

1. Retrieves articles in the target language.
2. Summarizes the content using a `facebook/bart-large-cnn` pipeline.
3. Translates the summaries back to English (only if the target locality is not English-speaking).
4. Ranks articles by computing TF-IDF vectors over the corpus (via `sklearn`), retaining only those above a similarity threshold relative to the user query.

Custom Itinerary Optimization

For sightseeing requests, we model points of interest as a weighted graph:

- **Nodes:** attractions with weights for aesthetic score and estimated visit duration.
- **Edges:** travel distances between nodes.

We apply a minimum-cost path algorithm under a time budget constraint (e.g., 8 hours/day number of days) to generate an optimal route. The start location is either user-defined or selected from top-ranked nodes. It is important to mention that the model uses Google Maps to check whether each place is open or closed for business when the user plans to visit. In case of closure, the system will note this in the output.

For example, if the query is:

"I would like to go to Rome for 3 days. Can you recommend the best things to visit? Start from Termini Station."

With the following configuration:

- `importance_time_visit = 0.0`
- `importance_beauty = 0.7`
- `importance_edge = 0.3`

The output is on Fig 1.

The total time for the complete itinerary is approximately **23 hours**, which fits within the 3-day plan (8 hours per day).

We can notice that the **Sistine Chapel** is *closed* — as shown in the image 2 and 1, it is marked as *"temporarily closed"*.

Supported Cities(only for this subtask)

Currently, the route optimization feature supports the following cities:

- Rome
- Ljubljana
- Prague

```
Here's the itinerary for Rome:
⌚ Total estimated time: 23 hours

📍 Places to visit:
✓ Museo Nazionale Romano: Palazzo Massimo alle Terme
✓ Basilica di Santa Maria Maggiore
✓ Colosseum
✓ Roman Forum
✓ Pantheon
✓ Piazza Navona
✓ Villa Farnesina
✓ Castel Sant'Angelo
✓ St Peter's Basilica
✓ Vatican Gardens
✗ Sistine Chapel (temporarily closed)
✓ Vatican Museums
✓ Gianicolo
✓ Museo della Repubblica Romana e della Memoria Gariboldi
✓ Basilica di Santa Maria in Trastevere
✓ Jewish Ghetto
✓ Campo de' Fiori
✓ Trevi Fountain
✓ Galleria Doria Pamphilj
✓ Piazza di Spagna
✓ Pincio Hill Gardens
✓ Museo e Galleria Borghese
```

Figure 1. Places in Rome

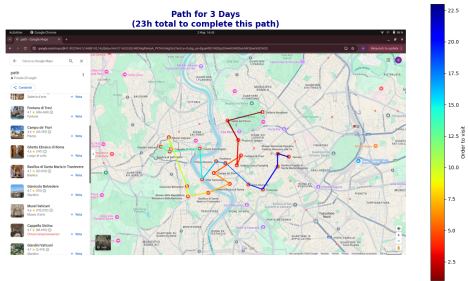


Figure 2. Example Path

- Vienna
- Florence
- Naples
- Maribor
- Paris
- Valencia
- Barcelona
- Madrid

Future Expansion

We plan to extend the list of supported cities in the future. The dataset for additional cities has already been downloaded but requires manual formatting to meet the model’s specifications.

System Execution

Users invoke the system via `run.py`, which orchestrates the above modules, compiles retrieved data, and formats the final response into a coherent, structured summary.

Examples

Query 1

User: *Some concerts in Ljubljana for tomorrow*

Output:

```
Events in Ljubljana
=====
🔍 *Malignant Tumour* - Malignant Tumour
📅 Data e Ora: May 02 - 8
📍 Location: Bazoviška cesta 26Ilirska Bistrica, Slovenia
📄 Descrizione: Mladinski klub Nade Žagar Ilirska Bistrica
=====
🔍 *Yullola* - Yullola
📅 Data e Ora: May 02 - 8
📍 Location: Metelkova ulica 4Ljubljana, Slovenia
📄 Descrizione: Channel Zero
=====
```

Figure 3. Output 1

Query 2

User: *I am going in Berlin, tell me the temperature for friday.*

Output:

📍 **Weather Forecast for Berlin**

📅 **Friday, May 02, 2025**

Time	Temperature	Rain	Wind
05:00	11.4 °C	* absent	🌬 moderate
06:00	11.4 °C	* absent	🌬 moderate
07:00	13.2 °C	* absent	🌬 moderate
08:00	16.1 °C	* absent	🌬 moderate
09:00	19 °C	* absent	🌬 moderate
10:00	21.7 °C	* absent	🌬 moderate
11:00	24 °C	* absent	🌬 moderate
12:00	25.2 °C	* absent	🌬 moderate
13:00	25.8 °C	* absent	🌬 moderate
14:00	26.5 °C	* absent	🌬 moderate
15:00	27.1 °C	* absent	🌬 moderate
16:00	27.2 °C	* absent	🌬 moderate
17:00	26.8 °C	* absent	🌬 moderate
18:00	25 °C	☁ mm	🌬 moderate
19:00	22.5 °C	☁ mm	🌬 moderate
20:00	21.2 °C	☁ mm	🌬 moderate
21:00	19.8 °C	* absent	🌬 weak

Figure 4. Output2

Testing and Performance Evaluation

We evaluated system performance using three complementary test suites:

- **Final-output testing**, to verify that correctly posed queries yield appropriate, informative responses.
- **Intent classification testing**, to measure how often the model assigns the correct intent label.
- **Query-error testing**, to ensure the system gracefully handles under-specified or malformed queries.

In the final-output tests, we submitted well-formed questions and confirmed that the agent both recognized the intent and returned coherent, contextually relevant information.

For intent classification, our test set comprised **118 examples** spanning all eight intent categories.

The model achieved **91.53% accuracy**. Misclassifications are visualized in Fig.5. The most frequent confusion occurred between *Safety updates* and *Severe weather alerts*, likely due to overlapping terminology. Contrary to expectations, *Best restaurants* and *Typical food dishes* exhibited minimal interchange errors. Finally, query-error testing validated the system’s ability to detect missing parameters (e.g., unspecified location or date) and prompt the user for clarification rather than producing nonsensical results.

Limitations

Although our system performs robustly in most scenarios, we identified the following areas for improvement:

- **Rate limiting and blocking**: Automated scraping of transportation schedules (e.g., TheTrainLine) can trigger anti-bot defenses after multiple rapid requests.

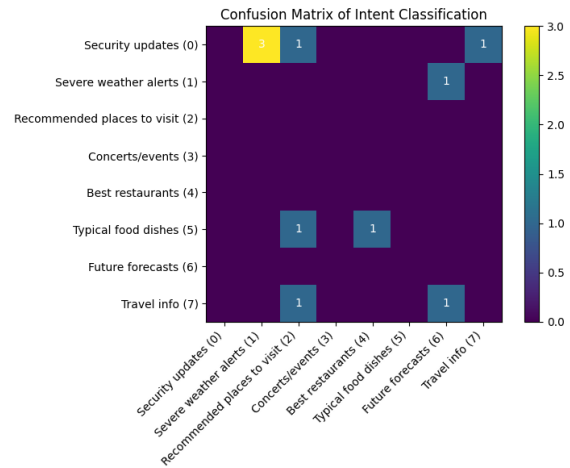


Figure 5. Intent Classification Confusion Matrix. Rows denote the true intent; columns denote the predicted intent.

- **Translation constraints**: To maximize relevance, safety alerts are fetched in the local language and translated back to English; however, our current pipeline only handles short snippets, necessitating sentence-by-sentence translation.

Future Work

Prior to final submission, we will:

- Implement request throttling and caching to mitigate blocking by external sites.
- Integrate fallback data sources to improve robustness when primary website fail.
- Extend the translation module to batch-process longer texts efficiently.

Discussion

The project has achieved satisfactory results: when the system correctly identifies the user’s intent, the output is generally reliable. However, it sometimes struggles to accurately extract the specified time frame or date — an aspect that can certainly be improved.

The most significant limitation lies in the intent analysis module, which assumes that every query must fall into one of the predefined intent categories. A useful extension would be to introduce a confidence threshold, below which the system could explicitly respond that it does not handle the given request. This would also require a more robust model capable of interpreting more complex or less structured queries.

Regarding the point-of-interest search algorithm, a transition from a strictly greedy approach to a beam search strategy could prove beneficial, allowing the system to retain multiple promising candidates during route planning. On the other hand, the implementation of checks for open or closed venues — for instance due to construction or special closures — was

particularly effective and valuable for tourists. The same applies to the analysis of local safety conditions, such as extreme weather, natural disasters, or other threats.

Overall, we believe the project represents solid work, with some limited but meaningful opportunities for refinement.

Acknowledgments

Here you can thank other persons (advisors, colleagues ...) that contributed to the successful completion of your project.

References

For a more detailed explanation, please refer to the README file available on GitHub at the following link.