LLMs in the Travel Domain: An Industrial Experience

Sergio Di Meglio[1,2], Aniello Somma[2], **Luigi Libero Lucio Starace**[1], Fabio Scippacercola[2] , Giancarlo Sperlì[1] , Sergio Di Martino[1]

[1] *Università degli Studi di Napoli Federico II, Naples, Italy*
[2] *Fervento srl, Naples, Italy*

SEKE 2025
37th International Conference on Software Engineering & Knowledge Engineering
September 29-30, 2025 – Pompeii, Italy

MOTIVATION AND CONTEXT

- Online travel platforms rely on aggregated data from multiple providers.
- These data are often **incomplete**, **inconsistent**, or **outdated**.
- Poor-quality descriptions lead to:
  - **User frustration**
  - **Reduced engagement**
  - **Lost bookings** and **revenue**

WHY LARGE LANGUAGE MODELS?

LLMs offer a way to:

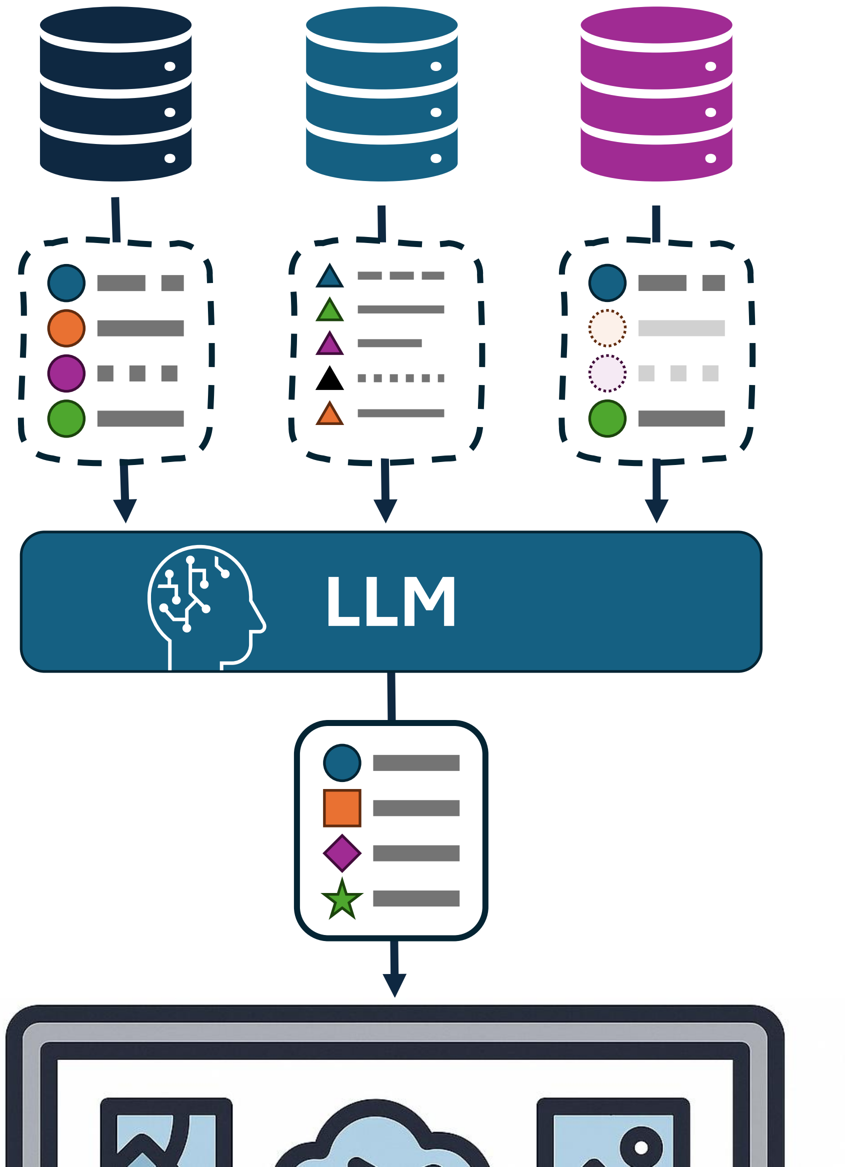
- Generate **coherent**, human-like descriptions
- **Fill gaps** in sparse data
- **Standardize** content across providers

WHY LARGE LANGUAGE MODELS?

LLMs offer a way to:

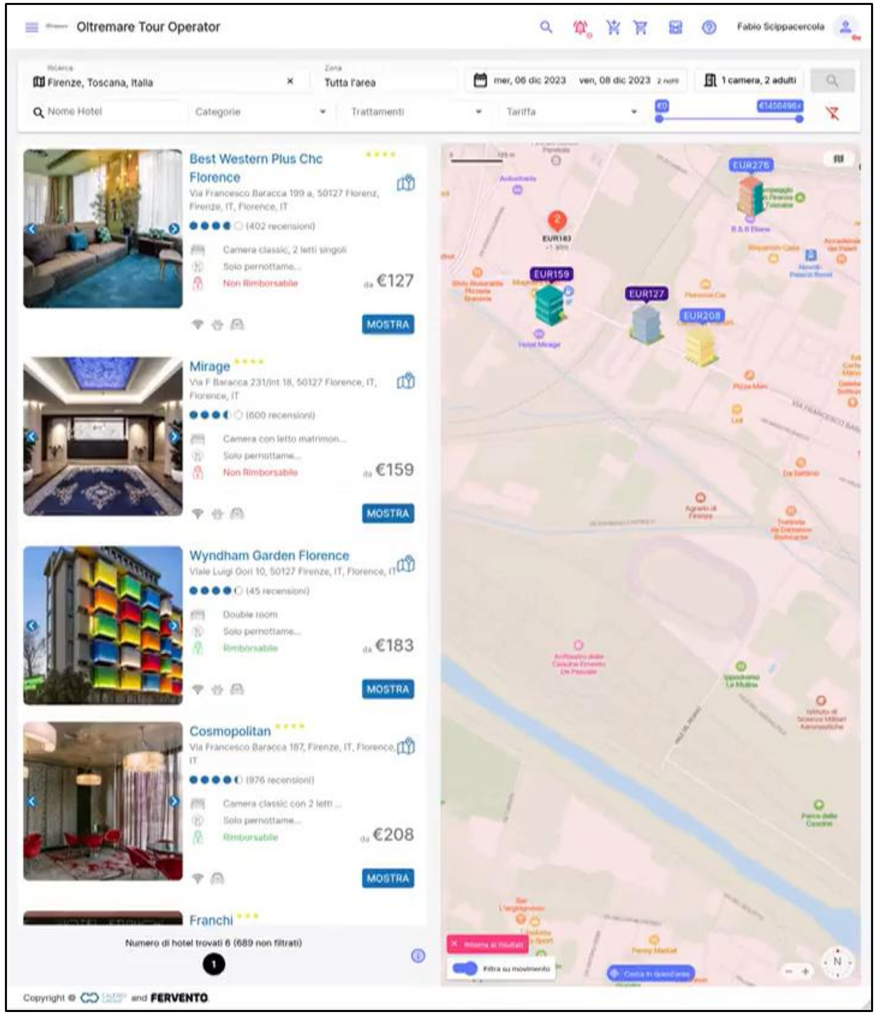
- Generate **coherent**, human-like descriptions
- **Fill gaps** in sparse data
- **Standardize** content across providers

Potential to transform catalog quality and enhance user experience.



INDUSTRIAL CASE STUDY: CALEIDOHOTELS

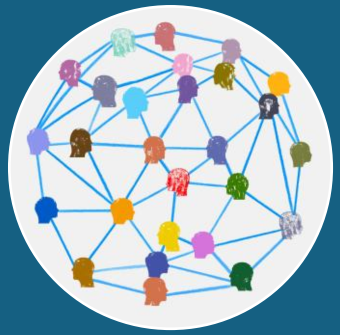
- Developed by FERVENTO
  - High tech startup operating in engineering advanced systems
- Aggregates listings from various providers via an interactive map.
- Faced significant issues with non-uniform facility descriptions
  - Especially when dealing with fallback data sources



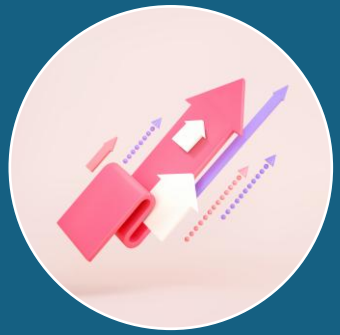
INDUSTRIAL OBJECTIVES

Improve user experience and brand consistency

Enable personalization and seasonal relevance



Simplify integration of new data sources



Differentiate from competitors with proprietary content

THE CALEIDOGEN SYSTEM

- Proposed solution: **CaleidoGen**
- A three-step process was put in place to implement it:

Catalog Analysis	Data Preprocessing	Model Tuning
• Manual inspection of 3 external catalogs • Identy formatting, completeness, and style issues	• Clean raw catalog data • Create curated training and test datasets	• Multiple LLMs / configurations • Select best models • FinalizeCaleidoGen architecture

CATALOG ANALYSIS

- **Goal**: Understand the structure, quality, and variability of input data from **three provider catalogs**.
- **Manual exploration** of catalog content from 3 external providers
  - Assessed **completeness** of facility descriptions.
  - Identified **style differences** and **format inconsistencies**.
  - Evaluated **parser effectiveness** for extracting structured data.
- **Outputs**
  - **Data Report**: Recommendations for structuring facility features. Foundation for preprocessing and model evaluation.

DATA PREPROCESSING

- **Goal**: Prepare high-quality datasets for **fine-tuning** and **testing**
- **Data cleansing**. Remove HTML tags, irregular formatting, catalog-specific quirks using insights from **Catalog Analysis.**
- **Feature extraction**. Parsed attributes into structured categories (Recreation, Services, Dining, Rooms, ...)
- **Dataset Construction**. Create curated structured examples with:
  - **Context**: structured description of facility features
  - **Output**: Manually-curated reference description of the facility

DATASETS

- **Training Dataset** (used for fine-tuning)
  - **100** cleaned and structured examples from 100 different facilities
- **Test Dataset**
  - **20** cleaned and structured examples from 20 different facilities (different from those included in the training dataset).

Context:
Recreation: Rooftop pool and spa; **Amenities**: Free Wi-Fi, business center, laundry services, on-site parking; **Dining**: Rooftop restaurant, lounge, complimentary breakfast; [...]
Output:
Indulge in leisure at the rooftop pool and spa or explore other amenities like complimentary Wi-Fi. Take advantage of services such as the concierge, fitness center, and conference rooms. Savor a meal at the rooftop restaurant or unwind in the lounge. Enjoy a complimentary breakfast in the morning. [...]

Examples from the datasets

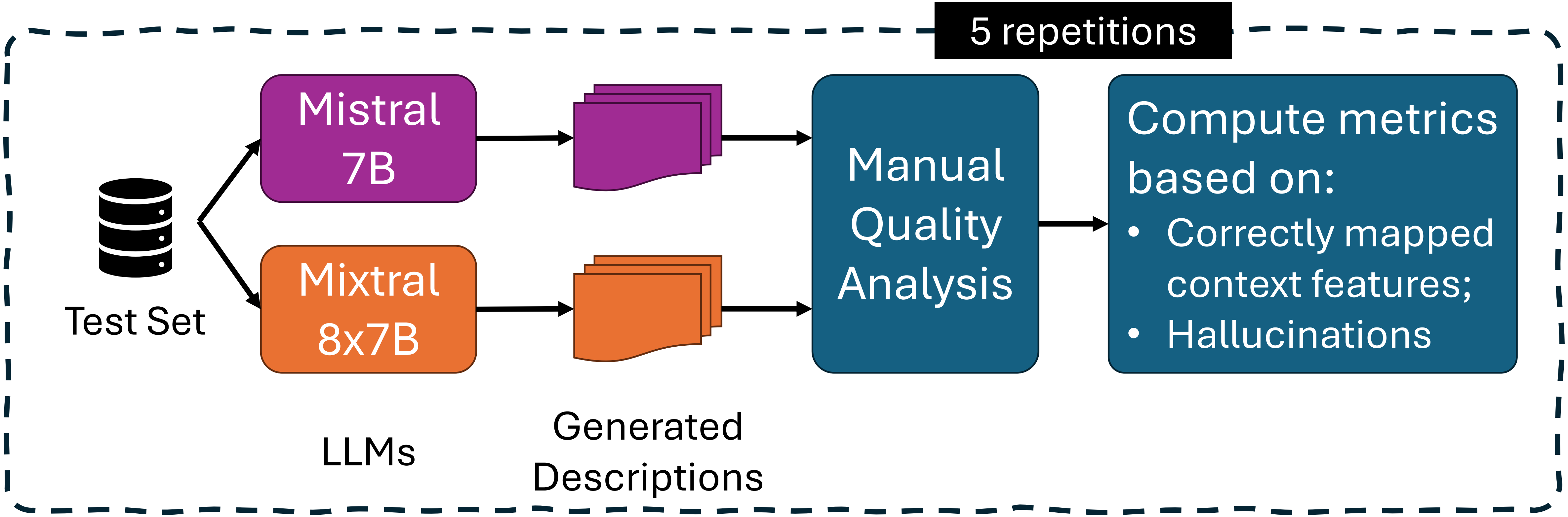
MODEL SELECTION

Two open LLMs by Mistral AI

- **Mistral 7B**:
  - Fine-tuned using QLoRA.
  - 4-bit quantization.
  - Cost-effective: 5GB VRAM, $0.16/hour.
- **Mixtral 8x7B**:
  - No fine-tuning (refined prompt)
  - 8-bit quantization.
  - High-performance: 50GB VRAM, $1.61/hour.

EMPIRICAL EVALUATION

- **Goal**: Compare the performance of the two LLMs
- **Method:**



EMPIRICAL EVALUATION: METRICS

- **Completeness.** How much of the context was included in the description $= \frac{\text{Context Features Added}}{\text{Total Context Features}}$
- **Precision**. Factual correctness of features included in the description $= \frac{\text{Correct Features Added}}{\text{Total Features Added}}$
- **Hallucination Rate**. Quantifies fabricated or incorrect details. $= \frac{\text{Hallucinated Features Added}}{\text{Total Features Added}}$
- **Length of Generation**. Average number of words per description.

RESULTS

Metric	Mistral 7B-FT	Mixtral 8x7B
Completeness	93% ± 8.8%	**99.6% ± 1.4%**
Precision	96% ± 3.2%	**98.8% ± 3.2%**
Hallucination Rate	4% ± 3.8%	**1.2% ± 3.2%**
Length (words)	277 ± 70	**249.2 ± 28**

Key Insights:

- Mixtral 8x7B consistently **outperformed** Mistral 7B-FT
- **Trade-off**: Mixtral requires ~10× more VRAM and costs ~10× more per hour.

CONCLUSIONS AND FUTURE WORKS

- The CaleidoGen system proved effective in transforming fragmented catalog data into a strategic content asset.
- Future Work
  - **Personalized Itineraries**: Generate travel plans based on user preferences and seasonal interests.
  - **Dynamic Content Generation**: Adapt descriptions based on real-time data (e.g., weather, events).
  - **Investigate Multilingual Support**: Extend generation capabilities to multiple languages for broader market reach.

LLMs in the Travel Domain: An Industrial Experience

Luigi Libero Lucio Starace

luigiliberolucio.starace@unina.it

https://luistar.github.io

MOTIVATION AND CONTEXT

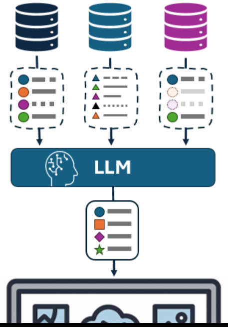
- Online travel platforms rely on aggregated data from multiple providers.
- These data are often **incomplete**, **inconsistent**, or **outdated**.
- Poor-quality descriptions lead to:
  - **User frustration**
  - **Reduced engagement**
  - **Lost bookings** and **revenue**

WHY LARGE LANGUAGE MODELS?

LLMs offer a way to:

- Generate **coherent**, human-like descriptions
- **Fill gaps** in sparse data
- **Standardize** content across providers

Potential to transform catalog quality and enhance user experience.



THE CALEIDOGEN SYSTEM

- Proposed solution: **CaleidoGen**
- A three-step process was put in place to implement it:

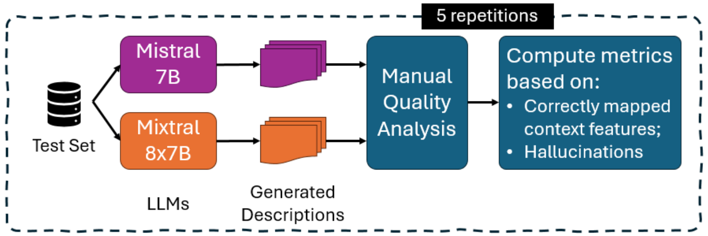
Catalog Analysis
- Manual inspection of 3 external catalogs
- Identy formatting, completeness, and style issues

Data Preprocessing
- Clean raw catalog data
- Create curated training and test datasets

Model Tuning
- Multiple LLMs / configurations
- Select best models
- FinalizeCaleidoGen architecture

EMPIRICAL EVALUATION

- **Goal**: Compare the performance of the two LLMs
- **Method**:



RESULTS

Metric	Mistral 7B-FT	Mixtral 8x7B
Completeness	93% ± 8.8%	**99.6% ± 1.4%**
Precision	96% ± 3.2%	**98.8% ± 3.2%**
Hallucination Rate	4% ± 3.8%	**1.2% ± 3.2%**
Length (words)	277 ± 70	**249.2 ± 28**

Key Insights:

- Mixtral 8x7B consistently **outperformed** Mistral 7B-FT
- **Trade-off**: Mixtral requires ~10× more VRAM and costs ~10× more per hour.