



**WORLD STATISTICS  
CONGRESS  
2025  
THE HAGUE**

**CPS 64**

# Challenges and insights in developing and evaluating a RAG pipeline for Official Statistics

---

Lino Galiana  
Insee, France  
10-06-2025

# Initial example

Is this a good answer ? Hard to tell

What is the GDP per capita in Martinique in 2022?



**qwen3:32b** Today at 16:38

Thought for 20 seconds ▾

The GDP per capita in Martinique in 2022 was approximately **€24,000** (euros). This figure is derived from Martinique's total GDP of around **€9.2 billion** in 2022, divided by its population of approximately **382,000** residents.

- **GDP (2022):** €9.2 billion
- **Population (2022):** ~382,000
- **GDP per capita:** €24,000

This estimate aligns with data from French national statistics (INSEE) and international organizations like the OECD. Martinique, a French overseas territory, uses the euro (€) as its currency. For context, this places Martinique's GDP per capita slightly above the average for the European Union.

# Initial example

- It is not a good answer : invents statistics convincingly
- RAG is a solution :
  - Information retrieval from a knowledge base
  - Text generation comes after with a LLM
- Multiples objectives :
  - Accurate (no hallucination), verifiable (citing sources), up to date
- Can be combined with agents (e.g. to retrieve data from API)

# Challenges

- We have knowledge on search engine optimization (SEO)
- ... far less on generative engine optimization (GEO) / AI readiness
- Many technical and methodological choices behind even simple RAG
  - **Evaluation** needed !

# What do we want ?

- Derive metrics from need instead of using general purpose metrics
  - See [Hamel Husain blog](#) (best resource on the subject)
- We collected expert level annotations
  - 60k pages from insee.fr
  - 62 Q&A written by experts
  - Is RAG citing the expected documents ? Is it providing a good answer ?

# Is it working ?

- From 78.3 % of invented references to 0.3 %
- Help us understand that our website is not adapted to LLM ingestion
  - LLM : good at ingesting standard markdown
  - Our website : not easily structured and parsed into markdown
    - Lots of non standard HTML tags
    - Hard to chunk pages
    - Lots of tables with complex structures

# Perspectives

- RAG quality depends first (and mostly) on how documents are parsed and processed
  - Back to retrieval information problem !
  - See [Barnett et al. \(2024\)](#)

After creating a good knowledge base, we can focus on how to improve relevance :

- e.g. how to prioritize recent content

# THANK YOU.