

Trustworthy Semi-Agentic Retrieval-Augmented Generation

Győző Csóka
/dʒø:zø: tʃo:kɒ/

Historical Archives of the Hungarian State Security (ÁBTL)
HUN-REN Alfréd Rényi Institute of Mathematics

5th November 2025
DLM Forum Members' Meeting Copenhagen



Historical
Archives of
the Hungarian
State Security



Agenda

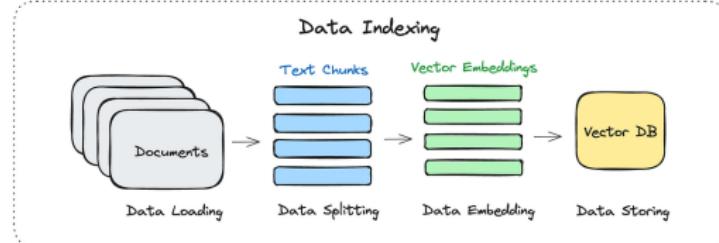
- introduction to Retrieval Augmented Generation (5 minutes)
- technical details of our implementation (15 minutes)
 - backend
 - agentic workflow
 - UI
- ongoing projects (10 minutes)

Problems with plain llm calls

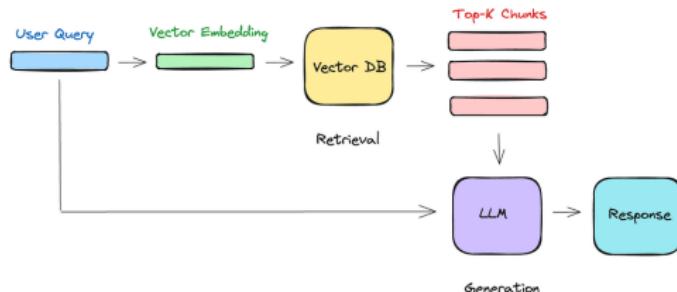
- Lack of factual grounding
- May generate plausible but hallucinated answers.
- Context length & cost limitations
- Can't handle the problem of rapidly changing data

RAG architecture

Basic RAG Pipeline



Data Retrieval & Generation



source: <https://medium.com/@drjulija>

Baseline RAG applications

- NotebookLM
- simple python implementation:

```
from llama_index.core import
    VectorStoreIndex, SimpleDirectoryReader
docs = SimpleDirectoryReader("./docs").load_data()
index = VectorStoreIndex.from_documents(docs)
print(index.as_query_engine()
      .query("Who were the members of the council?"))
```

Our use case

- Council of Ministers' meeting minutes of the first government of Imre Nagy from 1953-1954.
- original scans of typewritten records
- high-quality human transcriptions in PDF format
- tables of agenda items, laws
- long footnotes



Document structure & chunking

- extracted data tables
 - government members and their title/role
 - table of contents
- semantic chunking
 - following agenda items
 - footnotes connected to chunks
- metadata:
 - agenda_item_identifier
 - pdf_path
 - pages
 - date
 - minutes_id

- preprocess user query
 - extract keywords
 - add synonyms
 - find other wordings

- vectorised cosine-similarity

Agentic workflow

- tool calls
 - resolve abbreviations
 - fetch government members
 - fetch name index
 - fetch agenda items

Model choices

- Local/API
- embedding models
 - text-embedding-3-large
- LLM model
 - llama3.1
 - gpt-4o_mini
 - gpt-4o

Evaluation via human feedback

- domain knowledge
- human feedback
- retrieval evaluation
- user behaviour & guardrails

User interface

[click here for the pre-recorded demo](#)

Features

- openwebUI
- streaming
- linking sources
- source highlighting
- chat title generation
- follow-up questions
- admin page

Ongoing projects: Sensitive data detection

Werner Ugyével kapcsolatban ~~nagyon gyakran néztem~~ meg kell jegyezni, hogy az egykori budapesti Gestapo-Ugynökök tevékenysége megelhetősen hosszúban van a hatóságok előtt. Igen sok ilyen Ugynök ma is szabadlábon séti. Igen, hogy sokan ezzel tudnak védekezni, hogy a német bevonulás után ildöntésekkel váltak, de a már említett rendszeráltosás figyelembe-vételével ez a védekeésés elensik. /Ugy láttszik, hogy a német megszállók nem akartak többé miidő megbízottakkal dolgozni./

Rov Krdna-Kinfald nevű zandarmonik törz /Galadras nevű Utean ol-



Werner Ugyével kapcsolatban ~~nagyon gyakran néztem~~ meg kell jegyezni, hogy az egykori budapesti ~~Gestapo-Ugynökök~~ tevékenysége megelhetősen hosszúban van a hatóságok előtt. Igen sok ilyen ~~Ugynök~~ ma is ~~szabadlábon~~ kétál. Igen, hogy sokan ezzel tudnak védekezni, hogy a német bevonulás után ildöntésekkel váltak, de a már említett rendszeráltosás figyelembe-vételével ez a védekeésés elensik. /Ugy láttszik, hogy a német megszállók nem akartak többé miidő megbízottakkal dolgozni./

Rov Krdna-Kinfald nevű zandarmonik törz /Galadras nevű Utean ol-

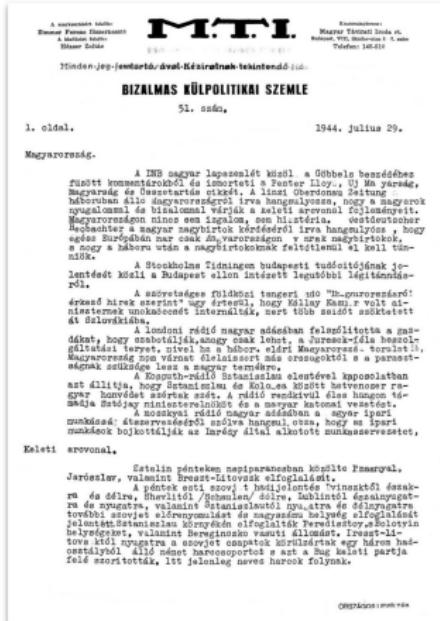
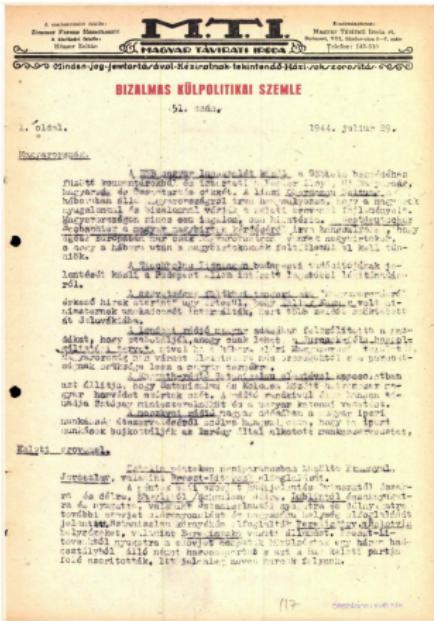
Ongoing projects: Sensitive data detection

- goal is different per use case
- noisy text
- gpt-oss is a game changer for text with OCR errors

Ongoing projects: Summarization



Ongoing projects: Document image enhancement



Ongoing projects: Document image enhancement

Synthetic Data Generator framework: Stages



Text:

- language
- layout
- fonts

Background:

- paper type
- overbleed

Perturbations:

- 30 noise types
- aged document

Geometric:

- rotation
- warp

Ongoing projects: Document image enhancement



https://huggingface.co/spaces/renyi-ai/die_demo

Team and partners

Ai HUN-REN Alfréd Rényi Institute of Mathematics
Artificial Intelligence Group

<https://ai.renyi.hu/>

[in/renyi-ai](#)



Thank you for your attention!
Now it is time for questions.



Historical
Archives of
the Hungarian
State Security

