

# Trustworthy Semi-Agentic Retrieval-Augmented Generation

Győző Csóka  
/dʒø:zø: tʃo:kɒ/

HUN-REN Alfréd Rényi Institute of Mathematics

5th November 2025  
DLM Forum Members' Meeting Copenhagen



Historical  
Archives of  
the Hungarian  
State Security



# Agenda

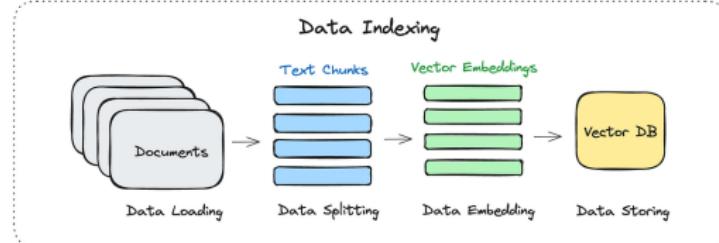
- introduction to Retrieval Augmented Generation (5 minutes)
- technical details of our implementation (10 minutes)
  - backend
  - agentic workflow
  - UI
- ongoing projects (5 minutes)

# Problems with plain llm calls

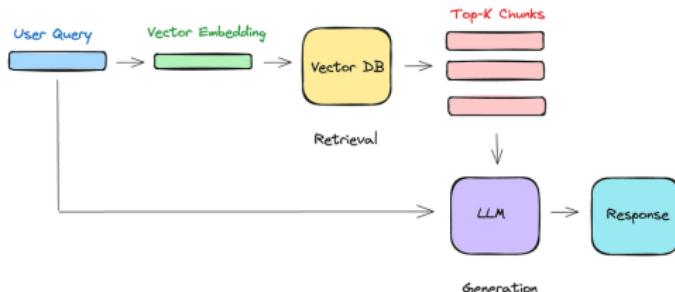
- Lack of factual grounding
- May generate plausible but hallucinated answers.
- Context length & cost limitations
- Can't handle the problem of rapidly changing data

# 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?"))
```

# Imre Nagy



# Imre Nagy

- communist reformist leader
- Prime Minister of Hungary (1953–1955, 1956)
- became the face of the 1956 Hungarian Revolution, supporting neutrality and withdrawal from the Warsaw Pact
- arrested after the Soviet invasion that crushed the revolution, executed in 1958 for treason
- posthumously rehabilitated in 1989, his reburial became a symbol of Hungary's democratic turn



# 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



# Tables, footnotes

2

2.

[500]

Szigorúan titkos!  
Készülő z. példányban  
Kapja: Rákosi Mária elvtárs

Jegyzőkönyv  
a Minisztertanács 1953. évi július hó 17. napján reggel 9 órakor kezdődő üléséről

Jelen voltak:

Nagy Imre, a Minisztertanács elnöke,  
Gerő Ernő, a Minisztertanács elnökönek helyettese,  
Hegedűs András, a Minisztertanács elnökönek helyettese,  
Hidas István nehezítpári miniszter,  
Zsóni Nyicz Mihály köhögéspári miniszter,  
Szalai Béla, az Országos Tervhivatal elnöke,  
Baldovszki János külgazdasági miniszter,  
Óls Károly pénzügyminiszter,  
Endei Ferenc igazságügy-miniszter,  
Kiss Árpád könyvvívápi miniszter,  
Bognár József bel- és külügyminiszter,  
Altomáre Iván elémítésiügyi miniszter,  
Tisza József bevényügyi miniszter,  
Bebrits Lajos közelkedésügyi miniszter,  
Darvas József népművelődési miniszter,  
Endrey-Gróf Tibor oktatási miniszter,  
Nógrádi Sándor, a honvédelmi miniszter elői helyettese,  
Kilitán József, az építésügyi miniszter elői helyettese,  
Simon Lajos, az egyszerűsítésügyi miniszter elői helyettese,  
Házi Árpád, az Állami Ellenőrző Központ elnöke,  
Sütő János, az Országos Tervhivatal elnökönek helyettese és  
Németh Béla, a Minisztertanács Titkárságához Elősök Csoport vezetője.

Egyes napirendi pontokhoz mint meghívottak:

- 1.) napirendi pont: Tausz György elvtárs
- 2.) napirendi pont: Tausz György elvtárs és Antos István elvtárs.
- 3.) napirendi pont: Szász Márton elvtárs, Zentai Ernő elvtárs és Dénes István elvtárs
- 4.) napirendi pont: Tausz György elvtárs, Antos István elvtárs és Szász Márton elvtárs
- 5.) napirendi pont: Szász Márton elvtárs és Pócz Tibor elvtárs

Különféleik: 11.) a.) napirendi pont: Oszetrovszky György elvtárs; b.) napirendi pont:

5

1953. AUGUSTUS 7.

25.)	503/35/1953.	Kormánydelegáció kiiktálása a Lipcsei Várára.
26.)	—	Kötöttségek a vannamap alkalmazásból.
27.)	—	Székelyula János rehívási minisztereljárásról kinevezése.

Napirenden kívül:

Napirendi pont	Hozzájárulásai:	A határozat címe:
1.)	503/nk/1/1953.	A szereplők járvány leküzdéséhez szükséges önkör- és földterhelésből származó árfolyam rendelkezésre hozzájárulás.
2.)	503/nk/2/1953.	Kapitál Pilát kinevezése
3.)	503/nk/3/1953.	Kovács Béla kinevezése
4.)	—	Varga Andris kinevezése.

Napirend előtt:

Nagy Imre elvtárs Bejelenti, hogy Bebrits Lajos elvtárs, közelkedés- és postaiigyi miniszter, Altomáre Iván elémítésiügyi miniszter és Óls Károly pénzügyminiszter ki-kérék rendes elői sahadságukat.

Kéri a Minisztertanács hozzájárulását, hogy az említett miniszter elvtárs megkezdhesek sahadságukat.

[Napirend]

1.) A termelőszövetkezetek és termelőszövetkezeti csoportok hitelezőtársasának rendezése<sup>1</sup>

1) Az 1954. június 30-ig török kiálláson hozzájárulásban figyelembe véve a török és magyar hitelezőkönknél 1874 millió Ft-re, amelyből 1953-ban 496 millió Ft-ot kellene valós visszatéríteni. Ezután 270 millió Ft-ot az új összeg (háromszínű hiba), amelyet az 1952. február 11-én következőleg kellen valós visszatéríteni, de nem volt rá időszak. Az elterjedt szám szerint a török és magyar 1953-ban 1854-ben csökkentett 1953-ban visszatérítendő hitelezőkönknél, részben a török kiállásonként meghosszabbításával, részben a fedezetles hitelezők egy részének elszaporításával – névtelensével kialakultak módjában. 1. Az 1952. december 31-én fennálló hosszú lejáratú hitelek 20 éves időtartamra kell meghosszabbításával, és a többet 1954. június 1-jéig kell elkerülni. 2. Az 1952. december 31-én fennálló kölcsönplázárt beruházási hitelek 1953-ban csökkenések összegük csak 240

82

240

# Document structure & chunking

chunking → user query → retrieval → reranking → tool calls → llm call

- 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

chunking → user query → retrieval → reranking → tool calls → llm call

- extract keywords
- add synonyms
- find other wordings

chunking → user query → retrieval → reranking → tool calls → llm call

- hybrid retrieval
  - semantic (dense)
  - lexical (sparse)

# Reranking

chunking → user query → retrieval → reranking → tool calls → llm call

- vectorised cosine-similarity

# Agentic workflow

chunking → user query → retrieval → reranking → tool calls → llm call

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

# Model choices

chunking → user query → retrieval → reranking → tool calls → llm call

- Local/API
- embedding models
  - bge-m3
  - text-embedding-3-large
- LLM model
  - llama3.1
  - gpt-4o\_mini
  - gpt-4o

# Model choices

chunking → user query → retrieval → reranking → tool calls → llm call

- Local/API
- embedding models
  - bge-m3
  - **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

demo

# Features

- openwebUI
- streaming
- linking sources
- source highlighting
- chat title generation
- follow-up questions
- user feedback
- 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 /Galadoss nevű törz



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 /Galadoss nevű törz

## Ongoing projects: Sensitive data detection

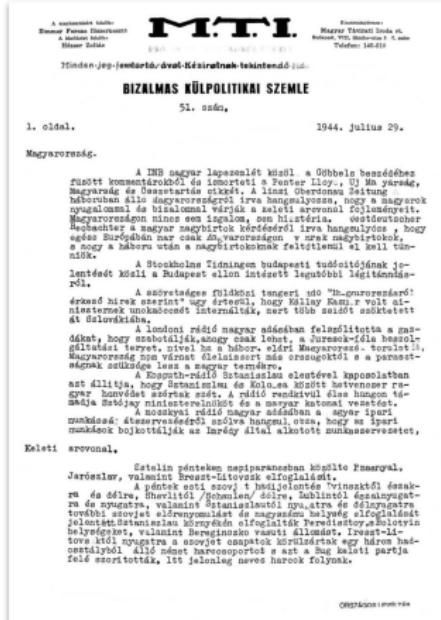
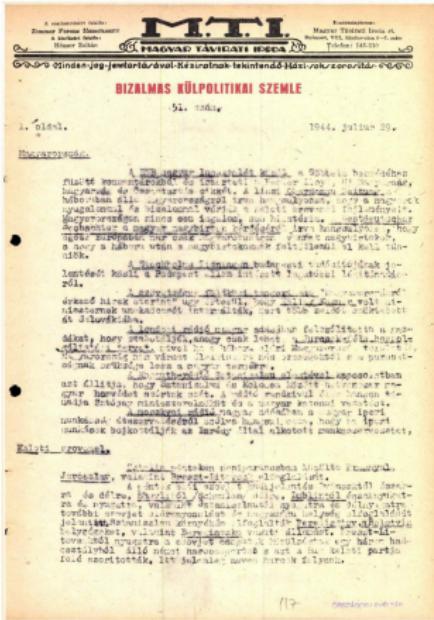
- goal is different per use case
- noisy text
- gpt-oss is big leap for text in Hungarian with OCR errors

# Ongoing projects: Summarization



source: <https://abtl.hu>

# 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](https://huggingface.co/spaces/renyi-ai/die_demo)

# Partners



Zoltán Lux



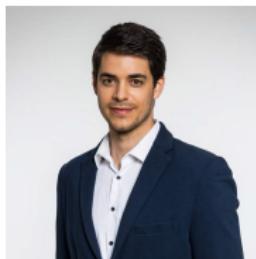
# Team



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

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

[in/renci-ai](#)



Gábor Kovács



Domonkos Czifra



Péter Körösi-Szabó



Győző Csóka

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

