

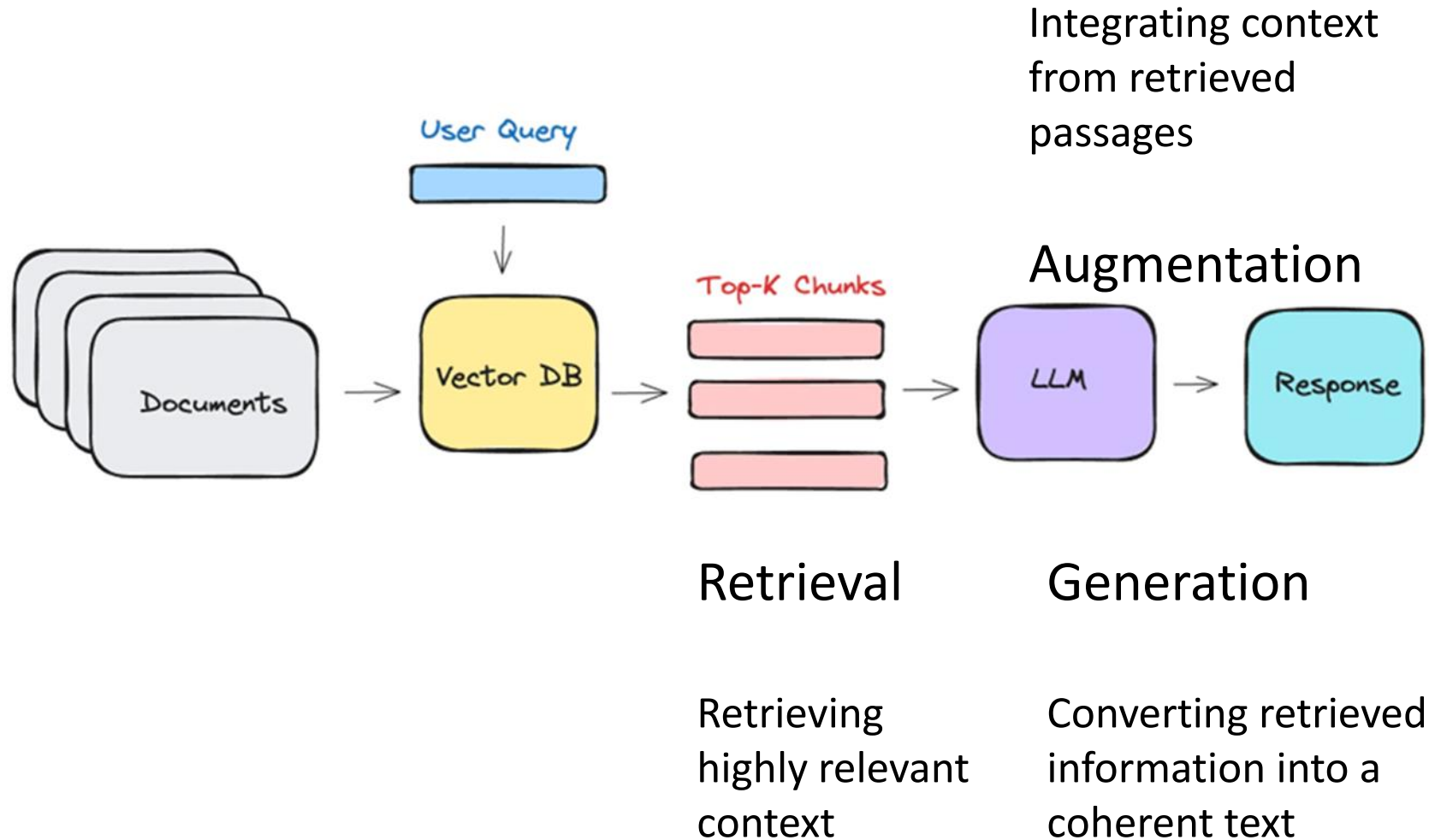


# Multi-Modal RAG

---

# Standard RAG

# What is Retrieval Augmented Generation (RAG)?



## Reduce Hallucinations

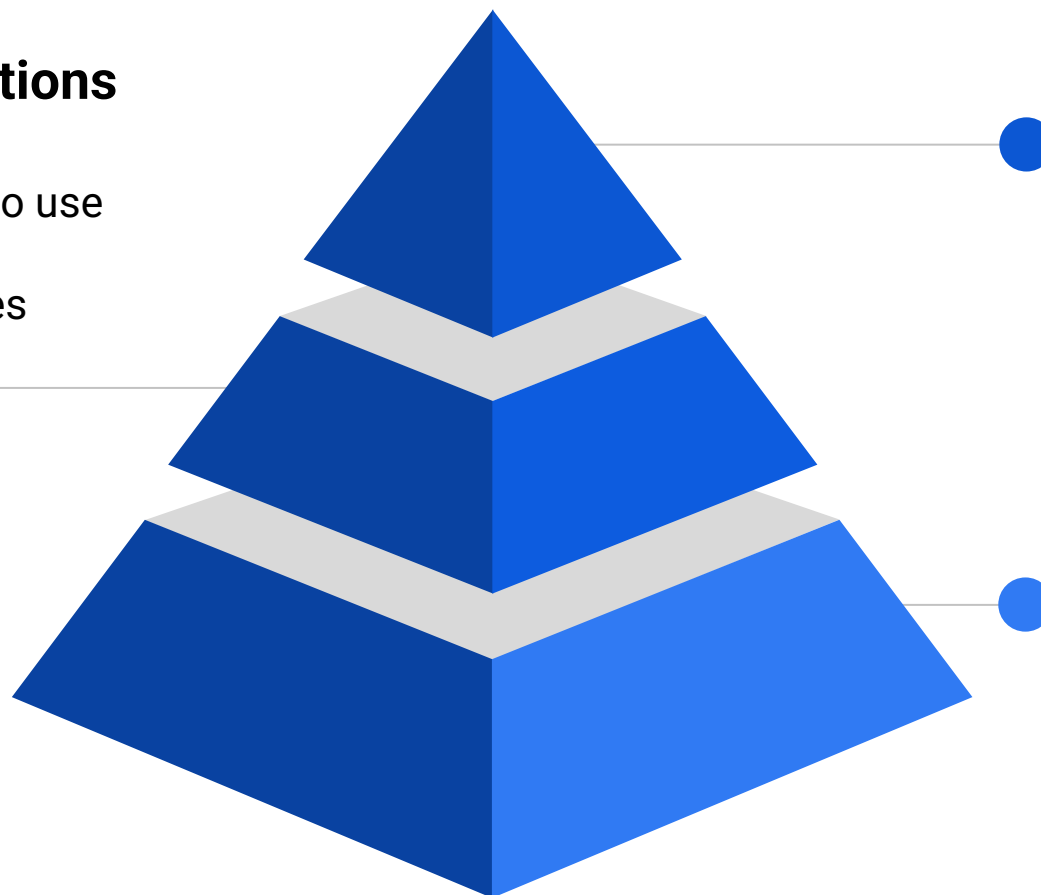
Model is conditioned to use provided context  
It can even cite sources

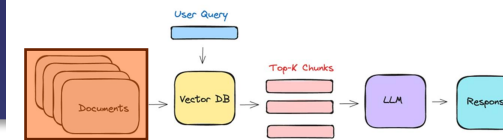
## Overcome Domain Knowledge Gaps

Include information the model was not trained on

## Improve Performance

Overcoming factuality issues by providing more faithful answers





## Document

26

Swiss National Sound Archives

### Collection

The National Sound Archives acquired 1,075 new audio carriers in 2024 (2023: 1,020) and added 7,043 new bibliographic records to the catalogue (2023: 5,467). The acquisition of the holding *Schweizerisches Cabaret- Chanson- Mimen- und Mundartrock-Archiv*, compiled by the patron of the arts and politician **Hansueli von Allmen** from Thun since 1972, which documents the history of Swiss cabaret in the four national languages merits special mention. A further noteworthy acquisition was the collection of Bernese cellist **Rudolf von Tobel**.



**Hansueli von Allmen**  
Materials from the archive

### Outreach

The Swiss National Sound Archives loaned out its works through a range of collaborations during the reporting year. For example, works went out to Sils Museum for the *Sun leivra* exhibition and to La Rada exhibition centre in Locarno for the *Sound echoes, listening spaces* exhibition.



**Rudolf von Tobel**  
Materials from the archive

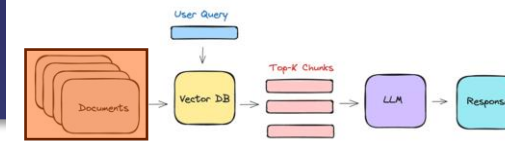


*Sound echoes, listening spaces*  
Sound artist Laurent Güdel taking recordings for the exhibition

Moreover, the National Sound Archives provided audio documents for various television and radio productions, including for the programme *Paganini* by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera, the programmes *Kulturplatz* and *Passage* by German-language broadcaster SRF Schweizer Radio und Fernseh.

## Text

Collection **The National Sound Archives acquired 1,075 new audio carriers in 2024** ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisition of the holding *Schweizerisches Cabaret- Chanson - Mimen- und Mundartrock-Archiv* , compiled by the patron of the arts and politician **Hansueli von Allmen** from Thun since 1972 , which documents the history of Swiss cabaret in the four national languages merits special mention . A further noteworthy acquisition was the collection of Bernese cellist **Rudolf von Tobel** . **Hansueli Outreach von Allmen** Materials from the The Swiss National Sound Archives archive loaned out its works through a range of collaborations during the reporting year . For example , works went out to Sils Museum for the *Sun leivra* exhibition and to La Rada exhibition centre in Locarno for the *Sound echoes* , listening spaces exhibition . *Sound echoes* , listening spaces Sound artist **Laurent Güdel** taking recordings for the exhibition Moreover , the National Sound Archives provided audio documents for various television and radio productions , including for the programme *Paganini* by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes *Kulturplatz* and *Passage* by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh



## Text

Collection The National Sound Archives acquired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisition of the holding Schweizerisches Cabaret- Chanson - Mimen- und Mundartrock-Archiv , compiled by the patron of the arts and politician Hansueli von Allmen from Thun since 1972 , which documents the history of Swiss cabaret in the four national languages merits special mention . A further noteworthy acquisition was the collection of Bernese cellist Rudolf von Tobel . Hansueli Outreach von Allmen Materials from the The Swiss National Sound Archives archive loaned out its works through a range of collaborations during the reporting year . For example , works went out to Sils Museum for the Sun leivra exhibition and to La Rada exhibition centre in Locarno for the Sound echoes , listening spaces exhibition . Sound echoes , listening spaces Sound artist Laurent Güdel taking recordings for the exhibition Moreover , the National Sound Archives provided audio documents for various television and radio productions , including for the programme Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh

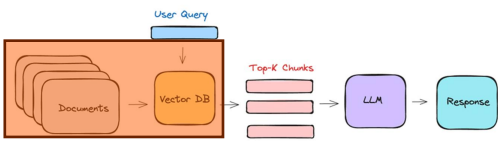
## Chunks

Collection The National Sound Archives acquired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The **acquisition of the holding**

**acquisition of the holding** Schweizerisches Cabaret- Chanson - Mimen- und Mundartrock-Archiv , compiled by the patron of the arts and politician Hansueli von Allmen from Thun since 1972 , which documents the history of Swiss cabaret in the four national languages merits special mention . A

...

Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh



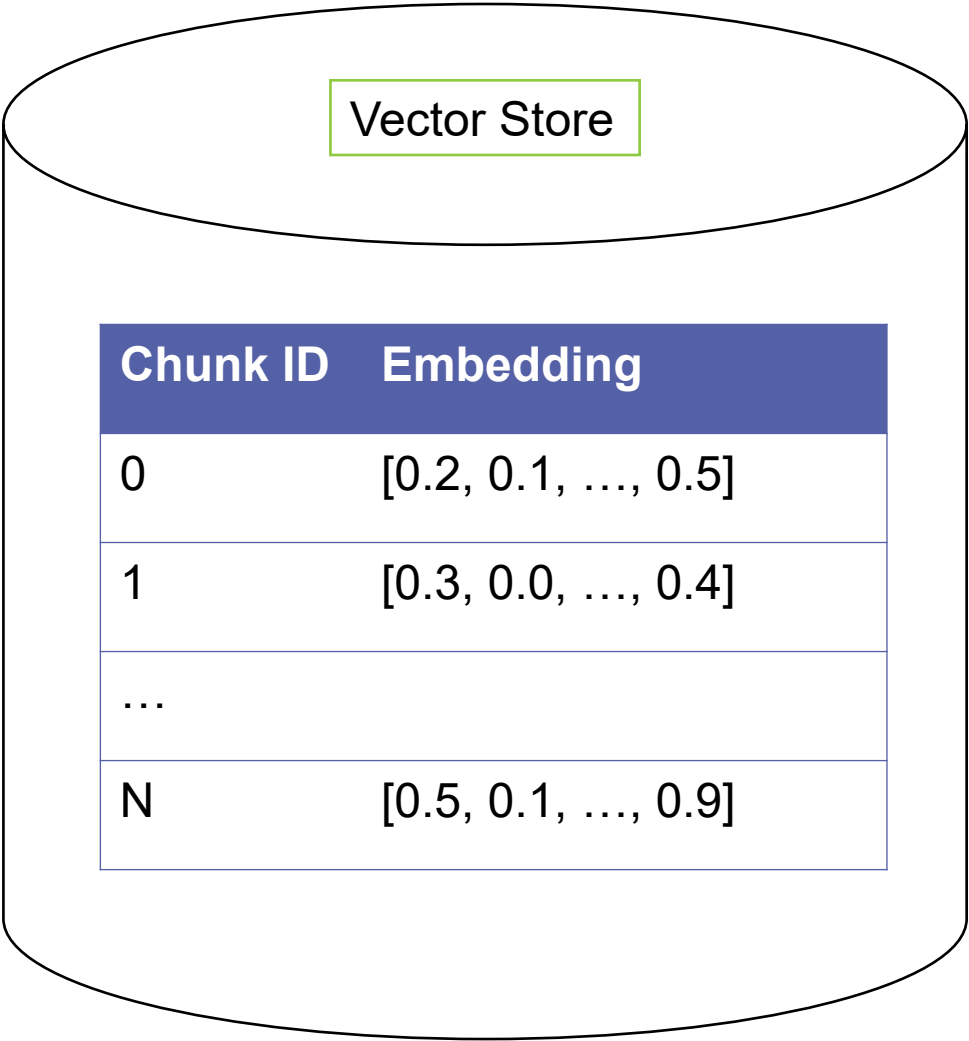
## Chunks

Collection The National Sound Archives acquired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisition of the holding

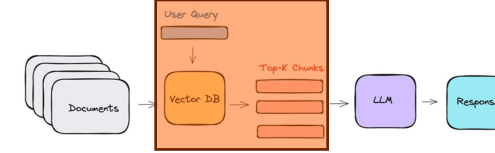
acquisition of the holding Schweizerisches Cabaret- Chanson - Mimen- und Mundartrock-Archiv , compiled by the patron of the arts and politician Hansueli von Allmen from Thun since 1972 , which documents the history of Swiss cabaret in the four national languages merits special mention . A

...

Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh







## Query

How many audio carriers have been acquired in 2024 by the Sound Archives?



## Vector Store

Chunk ID

Embedding

0 [0.2, 0.1, ..., 0.5]

1 [0.3, 0.0, ..., 0.4]

...

N [0.5, 0.1, ..., 0.9]



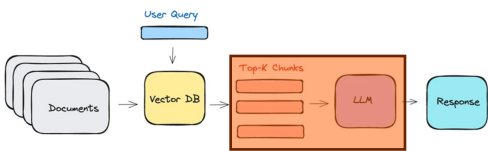
## Relevant Chunks

Collection The National Sound Archives acquired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisition of the holding

...

Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh





## Query

How many audio carriers have been acquired in 2024 by the Sound Archives?



## Relevant Chunks

Collection The National Sound Archives ac - quired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisi- tion of the holding

...

Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the pro- grammes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh



## Prompt Template

```
Answer the question solely based on
the following context:

```{context}```

Question: {question}
```

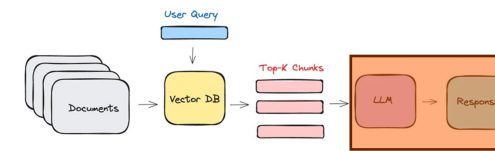


## Augmented Prompt

```
Answer the question solely based on the following
context:

...
Collection The National Sound Archives ac - quired 1,075 new
audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new
bibliographic records to the catalogue ( 2023 : 5,467 ) . The
acqui- sition of the holding
...
Paganini by Swiss Italian-language broadcaster RSI
Radiotelevisione svizzera , the pro- grammes Kulturplatz and
Passage Rudolf von Tobel by German-language broadcaster
Materials from the archive SRF Schweizer Radio und Fernseh
...

Question: How many audio carriers have been
acquired in 2024 by the Sound Archives?
```



LLM



Augmented Prompt

Answer the question solely based on the following context:

...

Collection The National Sound Archives acquired 1,075 new audio carriers in 2024 ( 2023 : 1,020 ) and added 7,043 new bibliographic records to the catalogue ( 2023 : 5,467 ) . The acquisition of the holding

...

Paganini by Swiss Italian-language broadcaster RSI Radiotelevisione svizzera , the programmes Kulturplatz and Passage Rudolf von Tobel by German-language broadcaster Materials from the archive SRF Schweizer Radio und Fernseh

...

Question: How many audio carriers have been acquired in 2024 by the Sound Archives?



How many audio carriers have been acquired in 2024 by the Sound Archives?

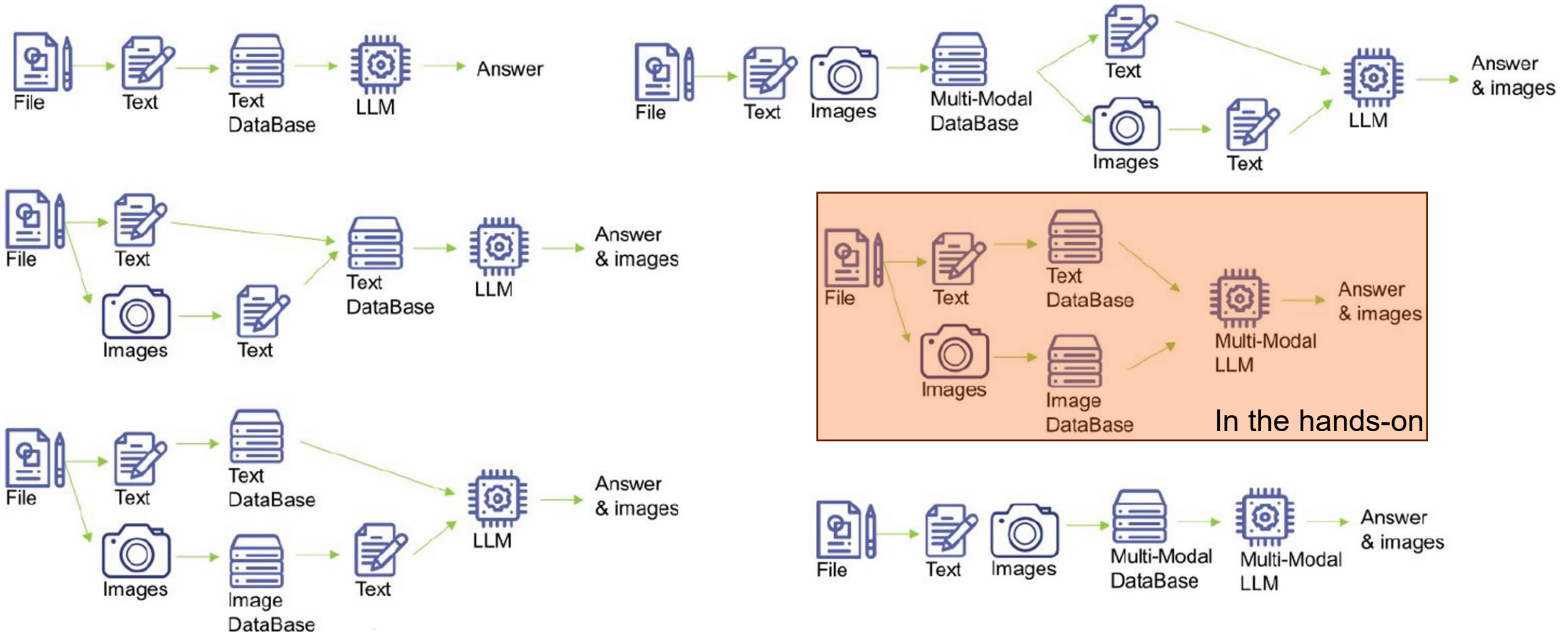
2025-06-12; 14:21

The Swiss National Sound Archives acquired 1,075 new audio carriers in 2024.

2025-06-12; 14:21

# Multi-Modal RAG

- Multi-modal RAG is a **RAG for several modalities** (text, image, sound, etc.).
- Here, we will focus on **text and images**.
- Some steps differ, but the core concept stays the same.



Another “tricky” alternative: Consider each page as an image and work with image-only RAG.

- Extraction:
  - **Images must be extracted as well.**
  - Typical challenges:
    - Images split in “layers”, cropped, etc.
    - Logos extracted on each page,
    - Captions considered as part of the image,
    - Etc.

→ This is not as easy as it looks!
  - A lot of tools can be used (MinerU, etc.).
- Chunking:
  - **Images are their own chunk.**
  - They can have their own metadata.

- Text Retrieval:

- Embed the text chunks and store them.
- Embed the query.
- Find best matches (top-k, threshold, etc.).
- Lot of different embedders (text-embedding-3-large, etc.).

- Image Retrieval:

- Embed the images and store them.
- Embed the query (! different embedder).
- Find best matches (top-k, threshold, etc.). Sparse search not possible.
- Lot of different embedders (VLM2Vec, etc.).



Relevant Chunks

Down-side of this approach, no ranking between text and images.



- Augmentation:
  - **Images are sent in the prompt as bytes.**
  - Considered as any other chunk but need to specify that the content is an image.
- Calling:
  - **The LLM can be called as usual.**
  - But it support multi-modality (*gpt-4o-mini*, etc.).

Let's see how this translates into code!

# Thank you!

[www.datascience.ch](http://www.datascience.ch)