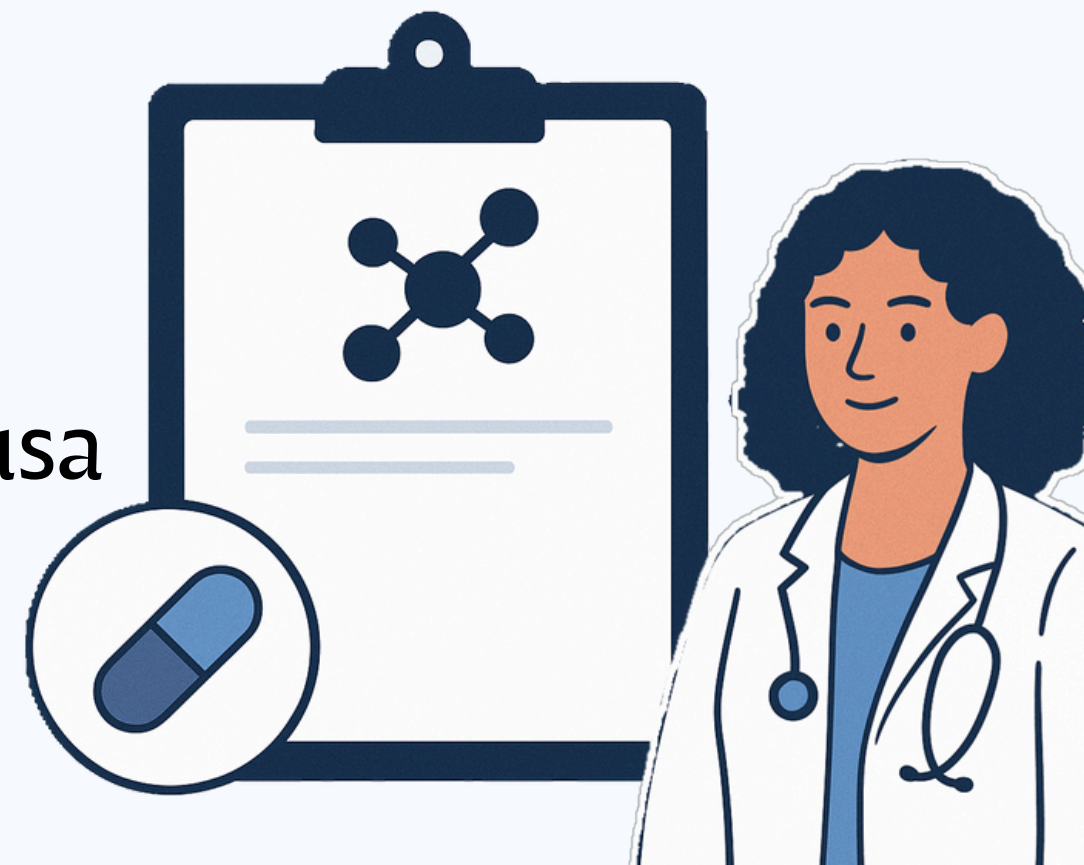

Bioinformatics Master
University of Minho, Braga
2025

Construction of a Dataset for Automated Prescription Recommendations

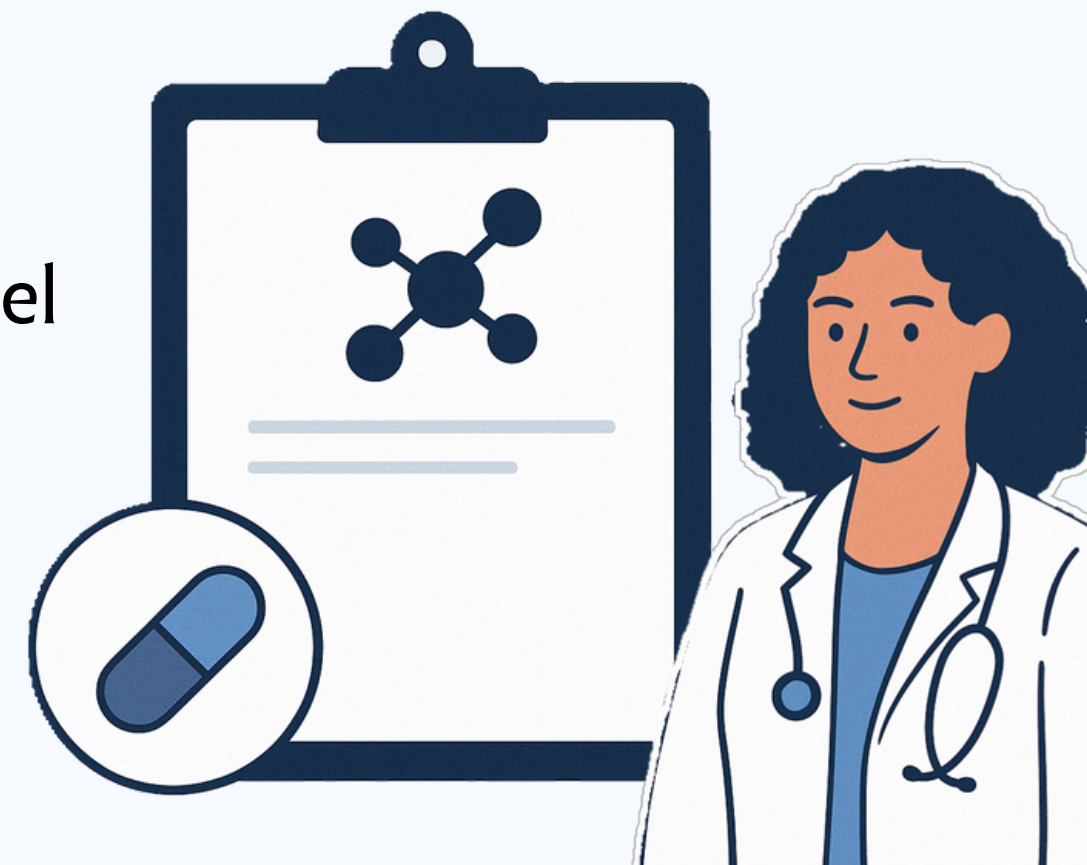
Elidiane do Rosário

Supervised by: José Machado & Regina Sousa



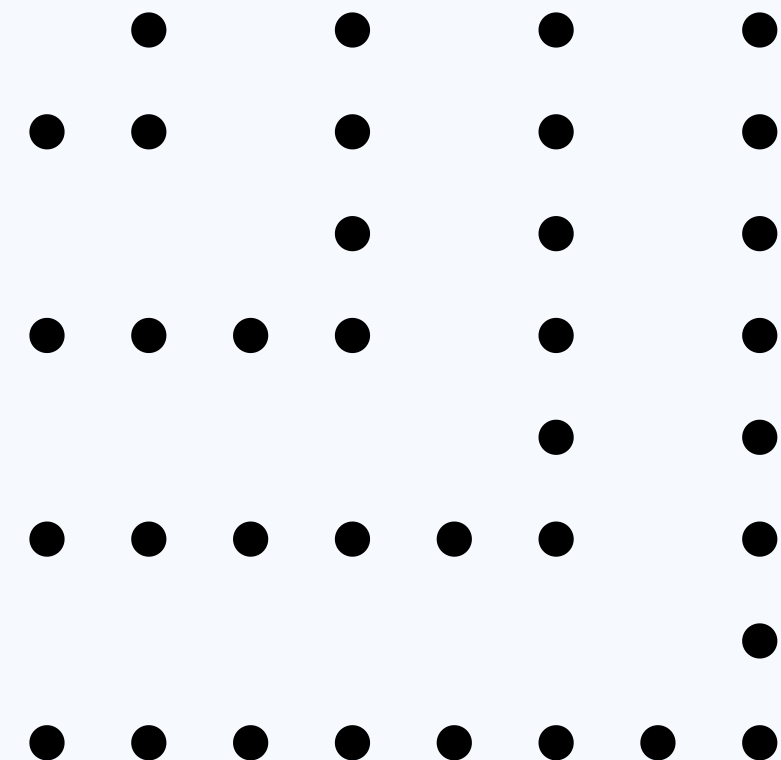
CLINICAL MOTIVATION

- Antibiotic interactions can silently compromise treatment — or even turn deadly.
- High-risk patients — like the elderly, children, and those with chronic illnesses — are the most vulnerable.
- When antibiotics fail, it's not just the infection that wins — resistance grows stronger.
- Too many drugs. Too little clarity. Polypharmacy and off-label use make safe prescribing harder than ever.



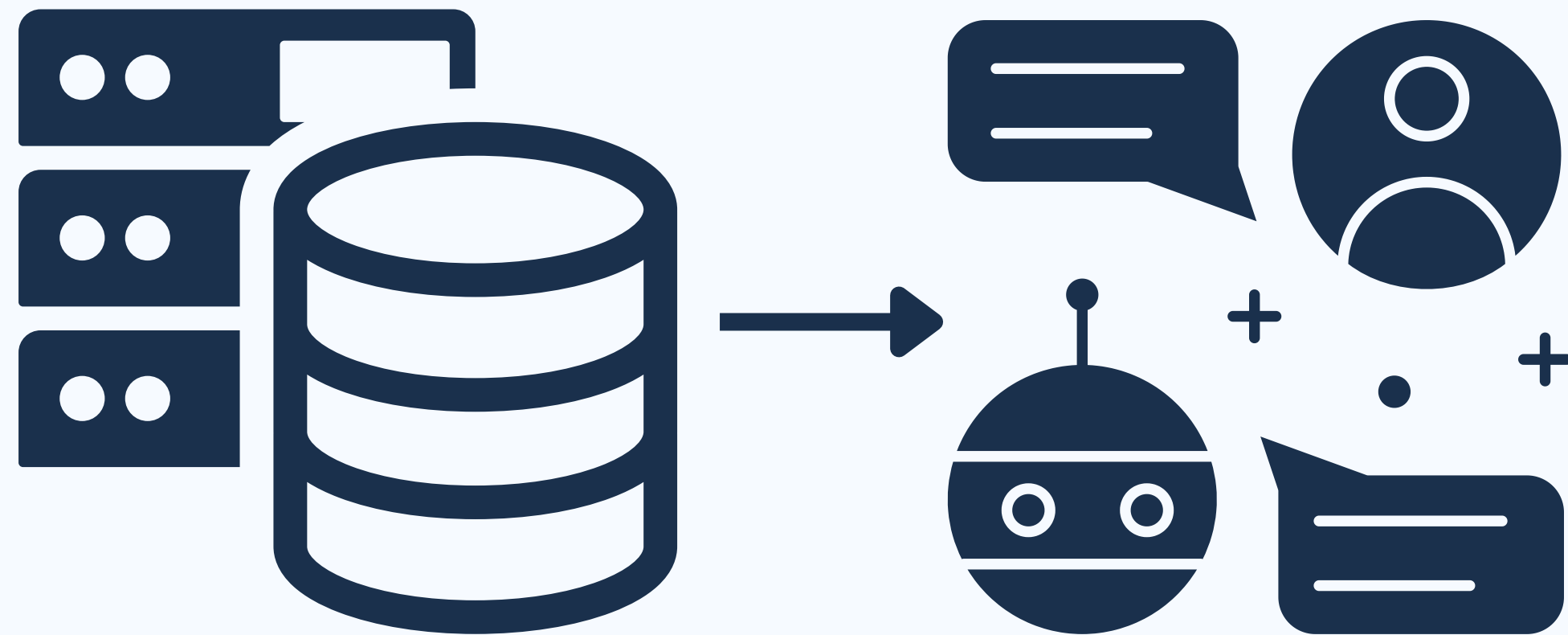
PROJECT GOALS

- Build a structured clinical dataset on antibiotics and drug interactions.
- Feed this dataset into a RAG + LLM pipeline.



PROPOSED ARCHITECTURE

Raw Data Source → ETL → Vector Store → RAG → LLM



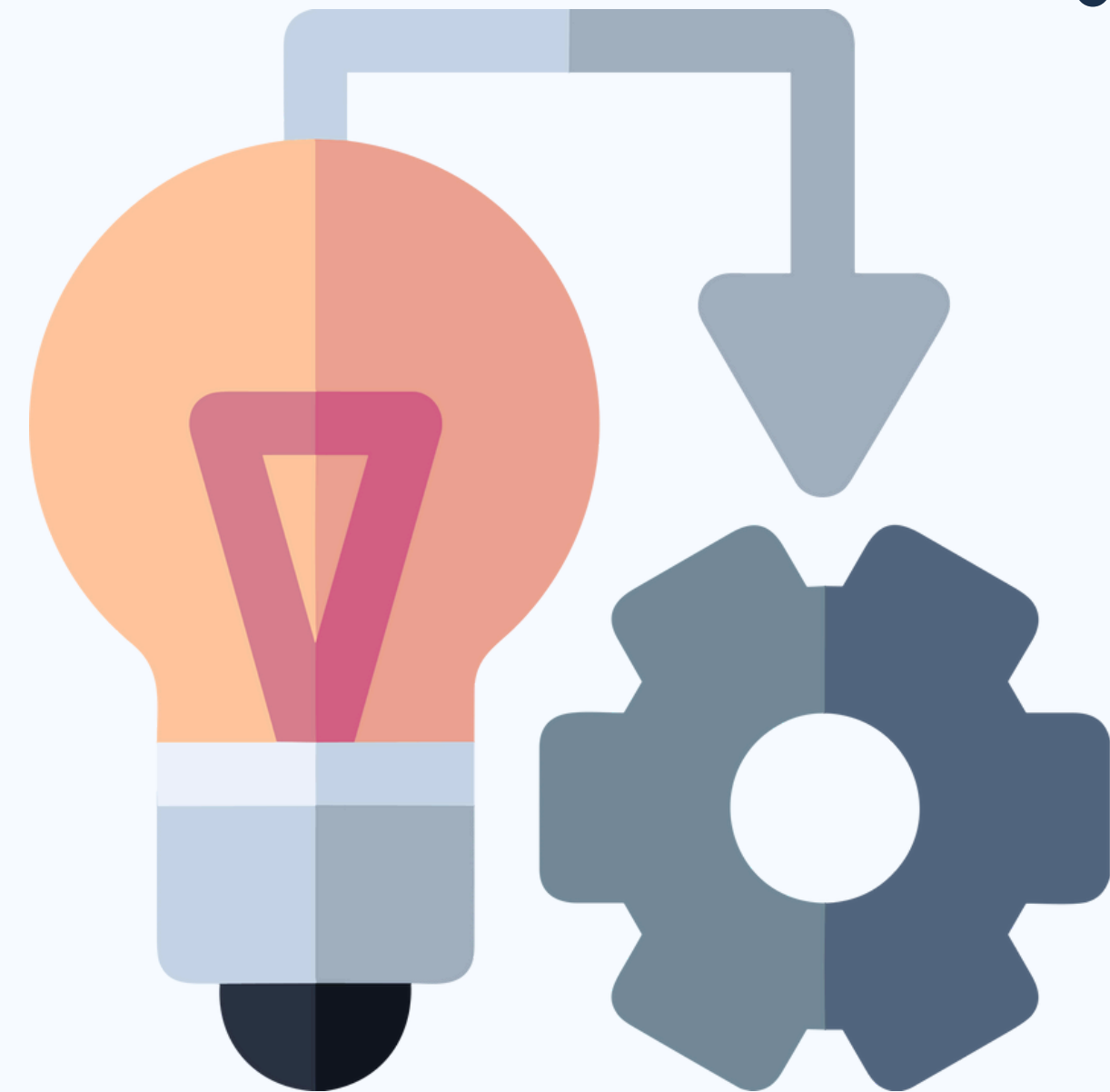
DATASET CONSTRUCTION

- Main fields:
 - Drugbank_id, name, description
 - Indications, toxicity
 - Affected organisms
 - Mechanism of action
 - Drug interactions (with description)
 - Dosages

```
"drugbank_id": "DB00080",
"name": "Daptomycin",
"description": "Daptomycin is a cyclic lipopeptide antibacterial...",
"cas_number": "103060-53-3",
...
"indication": "Daptomycin is indicated for the treatment...",
"toxicity": "Toxicity information regarding daptomycin i...",
"groups": [
    "approved",
    "investigational"
...
"categories": [
    "Agents Causing Muscle Toxicity",
    "Amino Acids, Peptides, and Proteins",
    "Anti-Bacterial Agents",
    "Anti-Infective Agents",
    "Antibacterials for Systemic Use",
...
"drug_interactions": [
    {
        "drugbank_id": "DB00243",
        "name": "Ranolazine",
        "description": "The serum concentration of Daptomycin...",
    },
    {
        "drugbank_id": "DB00175",
        "name": "Pravastatin",
        "description": "The risk or severity of myopathy can...",
    },
...
]
```

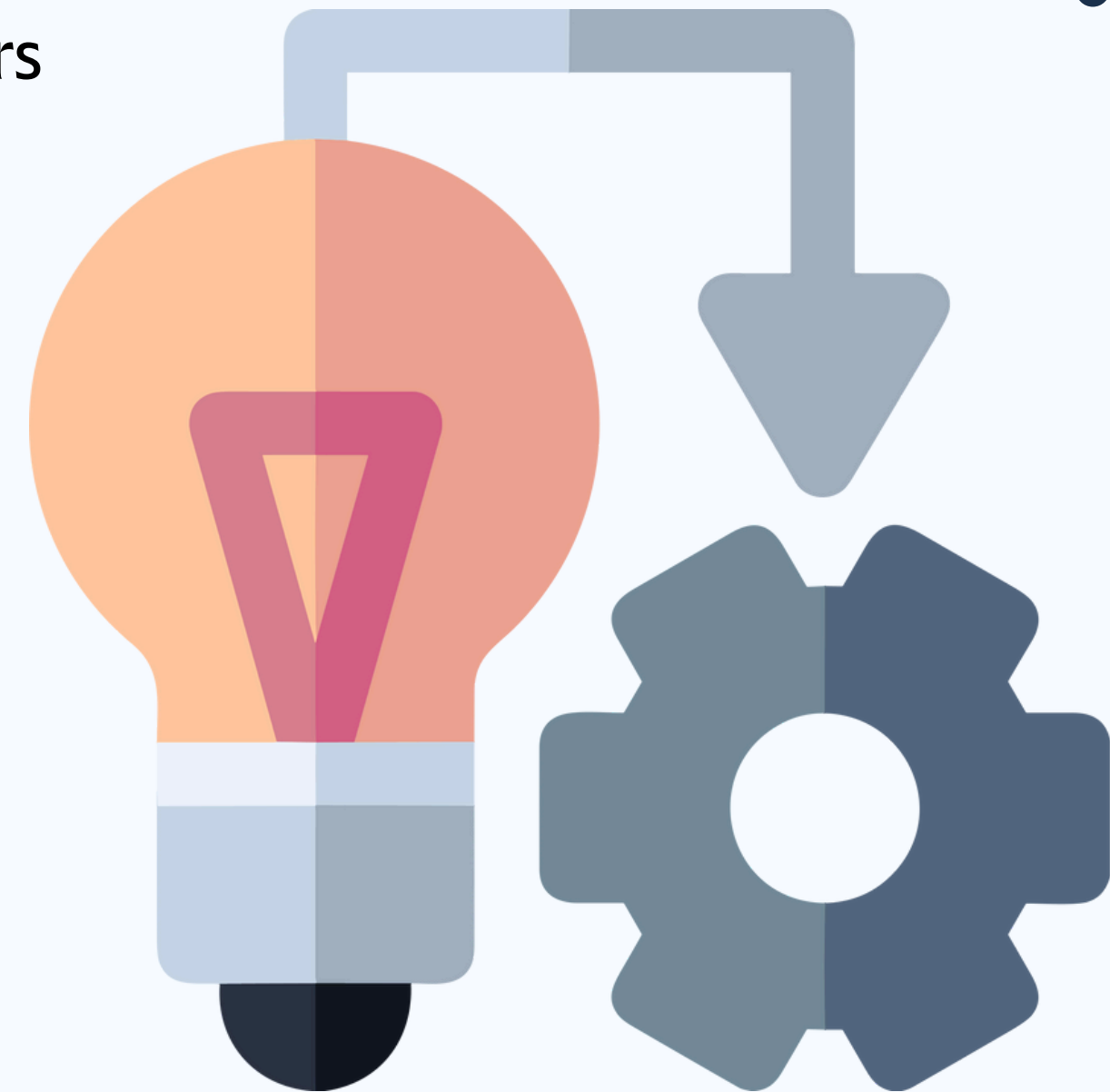
TECHNICAL IMPLEMENTATION

- 1 Parsed DrugBank XML using lxml.etree
- 2 Filtering and Extracting Relevant Data
- 3 Structuring Extracted Data as JSON
- 4 Data Chunking for RAG



TECHNICAL IMPLEMENTATION

- 5 Generating Embeddings using sentence-transformers
- 6 Indexing Embeddings in Chromadb
- 7 Querying the Vector Database (Retrieval)
- 8 Preparing Context for LLM – Mistral via Ollama –
(Generation Preparation)



PROMPT FORMATTING

```
system_prompt_rag = (  
    "Você é um assistente de informação sobre antibióticos, projetado para auxiliar médicos com base nos dados do dataset."  
    "O contexto fornecido pode estar em inglês, mas deve responder em português, traduzindo e sintetizando as informações de forma clara e precisa. "  
    "As suas respostas devem ser precisas, concisas e estritamente derivadas do 'Contexto' fornecido."  
    "É CRÍTICO que não adicione informações que não estejam explicitamente presentes no contexto, para evitar alucinações. "  
    "NUNCA forneça aconselhamento médico direto, faça diagnósticos ou prescreva tratamentos. A sua função é fornecer informações descritivas sobre os antibióticos. "  
    "Comece a sua resposta afirmando claramente que a informação é baseada nos dados do dataset e que não substitui o julgamento clínico do médico."  
    "Se o contexto fornecido não contiver a informação necessária para responder à pergunta, diga 'Não tenho informações suficientes nos dados do dataset fornecidos!'"  
    "para responder a esta pergunta."  
    " Mantenha um tom profissional e objetivo."  
)
```


LLM without RAG

Ex 1:

```
--- Consulta ao LLM PURO (Sem RAG) para: 'Um paciente usa Atorvastatina (uma estatina) diariamente e precisa de Azitromicina. Há risco de interação entre esses dois medicamentos e o que devo monitorar?' ---
```

```
--- Resposta do LLM PURO ---
```

```
Sim, existem algumas interações possíveis entre a Atorvastatina (uma estatina) e a Azitromicina. Uma delas é a potencial aumento do nível de azitromicina no sangue (farmacodinâmica), que pode levar a aumento dos efeitos colaterais, como dor de garganta ou náuseas. Outra interação é a potencial diminuição da absorção intestinal da Atorvastatina (farmacocinética), que pode levar à redução da eficácia da mesma.
```

```
Para minimizar essas interações, o médico pode ajustar as doses de cada medicamento ou recomendar um intervalo adequado entre suas administrações. É importante sempre seguir os comandos do seu médico e informá-lo se estiver tomando outros medicamentos ao mesmo tempo, pois essas interações podem ser potencialmente perigosas. Não é recomendado self-medicar sem consultar um profissional de saúde qualificado.
```

LLM with RAG

Ex 1:

```
--- Consulta RAG (Com Contexto do Dataset) para: 'Um paciente usa Atorvastatina (uma estatina) diariamente e precisa de Azitromicina. Há risco de interação entre esses dois medicamentos e o que devo monitorar?' ---  
Gerando novo embedding para a query...  
Buscando os 15 chunks mais relevantes no ChromaDB...  
Chunks recuperados (top 15):  
  - Chunk 1 (Tipo: drug_interaction, ID DrugBank: DB00207, Distância: 0.9839): Interação Medicamentosa de Azithromycin (ID DrugBank: DB00207) com Atorvastatin (ID DrugBank: DB0107...  
  - Chunk 2 (Tipo: drug_interaction, ID DrugBank: DB00355, Distância: 1.0138): Interação Medicamentosa de Aztreonam (ID DrugBank: DB00355) com Tiocloamarol (ID DrugBank: DB13451).  
  ...  
--- Resposta do LLM RAG ---  
A atenção médica deve monitorar o risco ou a gravidade de miopatia, rhabdomyolysis (quebra dos músculos) e mioglobulinúria (sangue de cor vermelha nos rins) ao combinar Azitromicina com Atorvastatina. Essa interação medicamentosa pode aumentar o risco ou a gravidade desses efeitos adversos.
```



LLM without RAG

Ex 2:

```
--- Consulta ao LLM PURO (Sem RAG) para: 'Para um paciente diabético em tratamento com Metformina e Glibenclamida, que desenvolveu uma infecção do trato urinário, posso prescrever Sulfametoxazol-Trimetoprima?' ---
```

```
--- Resposta do LLM PURO ---
```

```
Sim, o Sulfametoxazol-Trimetoprima (Bactrim ou Septra) é um antibiótico comumente usado para tratamento de infecções urinárias causadas por bactérias sensíveis. Este fármaco também pode ser prescrito ao paciente diabético que está em tratamento com Metformina e Glibenclamida, já que não interage significativamente com esses medicamentos. Porém, sempre consulte seu médico antes de começar qualquer nova terapêutica ou mudanças na dosagem de outros medicamentos, pois o tratamento ideal depende de vários fatores individuais do paciente, como a natureza da infecção e sua tolerância a determinados medicamentos.
```

LLM with RAG

Ex 2:

```
--- Consulta RAG (Com Contexto do Dataset) para: 'Para um paciente diabético em tratamento com Metformina e Glibenclamida, que desenvolveu uma infecção do trato urinário, posso prescrever Sulfametoxazol-Trimetoprima?' ---
```

```
Gerando novo embedding para a query...
```

```
Buscando os 15 chunks mais relevantes no ChromaDB...
```

```
Chunks recuperados (top 15):
```

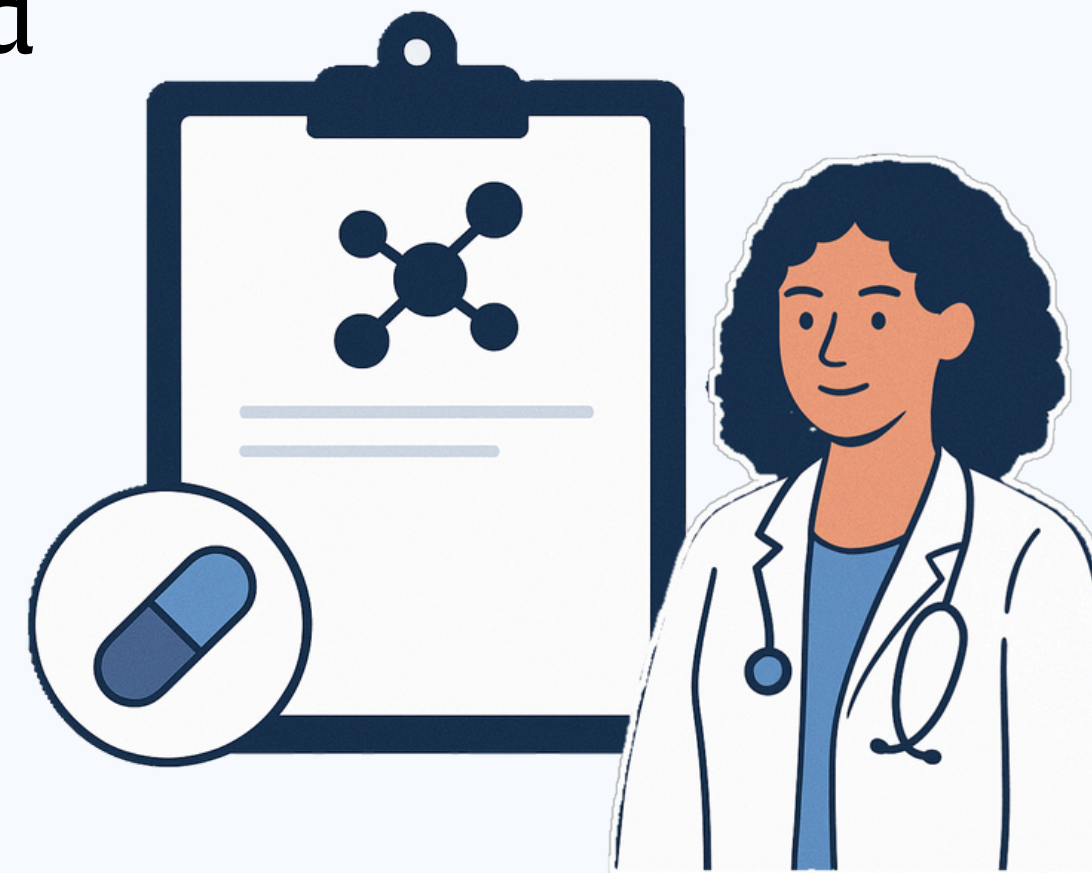
- Chunk 1 (Tipo: drug_interaction, ID DrugBank: DB01015, Distância: 0.8252): Interação Medicamentosa de Sulfamethoxazole (ID DrugBank: DB01015) com Prednisolone hemisuccinate (I...
- Chunk 2 (Tipo: drug_interaction, ID DrugBank: DB13214, Distância: 0.8337): Interação Medicamentosa de Sulfadiazine (ID DrugBank: DB13214) com Insulin tregopil (ID DrugBank: ...
- Chunk 3 (Tipo: drug_interaction, ID DrugBank: DB15975, Distância: 0.8341): Interação Medicamentosa de Sulfamethoxazole (ID DrugBank: DB15975) com Insulin tregopil (ID DrugBank: DB...

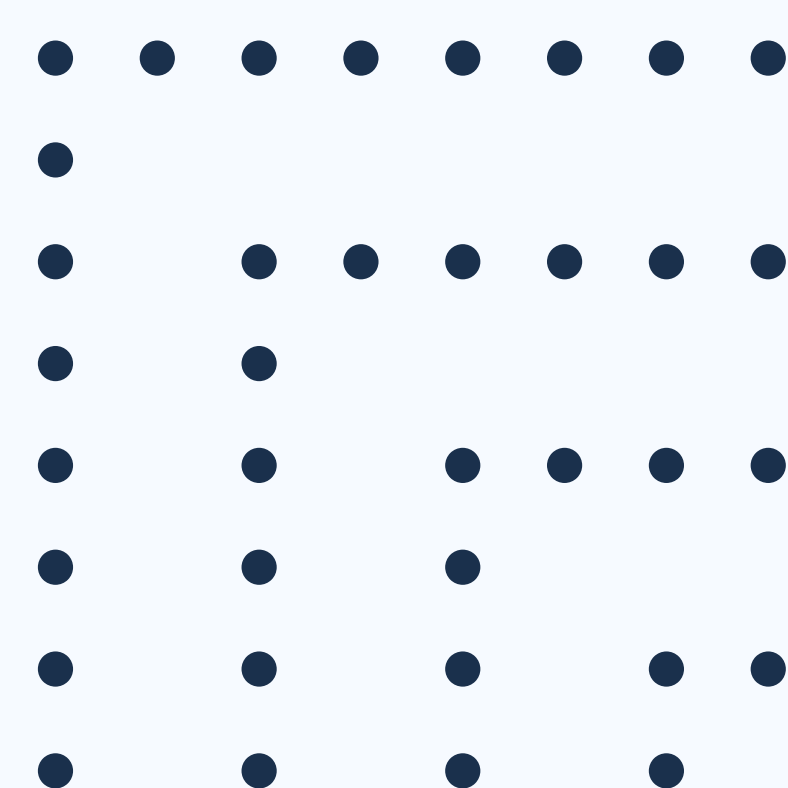
```
--- Resposta do LLM RAG ---
```

Não é possível fornecer um conselho clínico direto, mas tenho informações suficientes nos dados do dataset que podem ser úteis para a escolha de antibióticos. O Sulfametoxazol-Trimetoprima apresenta interações medicamentosas com Insulin tregopil e Insulin human, aumentando o risco ou a gravidade da hipoglicemia. Portanto, é importante que médicos tomem cuidado ao prescrever este antibiótico para pacientes diabéticos em tratamento com insulina ou outros agonistas do receptor de insulina, especialmente no caso de uso concomitante com Metformina e Glibenclamida. É importante que o médico sempre tome em consideração a natureza clínica individual do paciente para decidir sobre a escolha adequada do antibiótico e manter uma vigilância estreita da glicemia durante o tratamento.

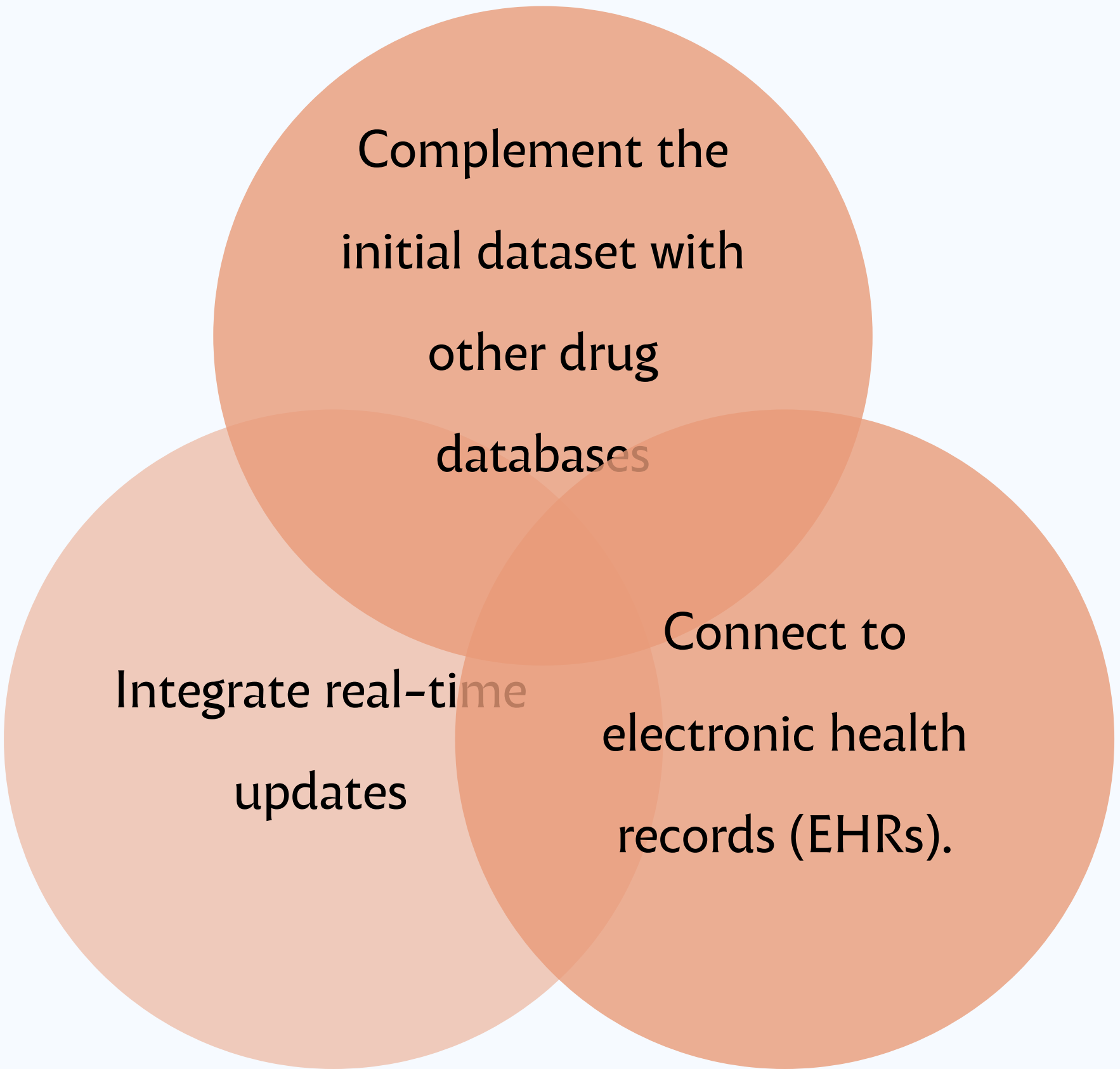
CONCLUSIONS

- A structured clinical dataset was built from DrugBank XML, processed into JSON.
- The proposed architecture was effectively implemented.
- The RAG-enabled LLM provided more precise and context-aware information on drug interactions compared to the pure LLM.





FUTURE WORK





Questions?

Elidiane do Rosário
pg55128@uminho.pt

