

Construction of a Dataset for Automated Prescription Recommendations: A Focus on Antibiotic Drug Interactions using Retrieval-Augmented Generation

Elidiane do Rosário¹, José Machado¹, and Regina Sousa¹

Universidade do Minho, Braga - Portugal

Abstract. The complexity of polypharmacy and the risk of adverse drug interactions present major challenges in clinical practice. This study proposes an AI-based system for automated prescription recommendations, focusing on antibiotic drug interactions. A structured dataset of 546 antibiotics was extracted from the DrugBank XML database, transformed into JSON, and segmented into nearly 300,000 semantically coherent text chunks. Sentence-transformers generated embeddings, stored in ChromaDB for efficient semantic retrieval. A Retrieval-Augmented Generation (RAG) framework was implemented, integrating this context into responses from the Mistral 7B large language model deployed locally via Ollama. Compared to a standalone LLM, the RAG-enhanced system showed superior accuracy and clinical relevance. It correctly flagged high-risk combinations (e.g., TMP-SMX with sulfonyleureas; azithromycin with statins) and significantly reduced hallucinations. This work demonstrates a scalable, privacy-preserving approach for AI-powered clinical decision support, improving medication safety and informed prescribing.

Keywords: Dataset Construction, Automated Prescription Recommendation, Retrieval-Augmented Generation, Large Language Models, Antibiotic Drug Interactions, AI in Medicine

1 Aims

This study aims to create a comprehensive and structured clinical dataset focused on antibiotics, including their pharmacological properties and interactions with other medications—particularly those involving other antibiotics and commonly prescribed medications. Building on this foundation, the project will implement a Retrieval-Augmented Generation (RAG) framework integrated with an LLM to improve the accuracy and safety of automated prescription recommendations. The study will also evaluate the effectiveness of LLMs, with and without RAG, in producing context-aware responses regarding drug interactions. Ultimately, the goal is to contribute to clinical decision-making by proposing a scalable, data-driven approach that can be applied to real-time healthcare applications.

2 State of the Art

2.1 Antibiotics-Drug Interactions

A drug–drug interaction (DDI) refers to a phenomenon in which the presence of one medication alters the effects of another, potentially reducing its efficacy, increasing toxicity, or, in severe instances, leading to fatal outcomes [54].

Antibiotic-drug interactions (ADIs) specifically involve these alterations when one of the drugs in question is classified as an antibiotic [7]. ADIs are a critical area of research due to the widespread use of antibiotics and the pharmacokinetic and pharmacodynamic mechanisms that make them significant contributors to Adverse Drug Reactions (ADRs) [28]. An antibiotic failure is a serious concern, as interactions not only affect the effectiveness of antibiotics but also contribute to antibiotic resistance [8].

These issues occur more frequently in vulnerable populations such as children, the elderly, and individuals with chronic diseases [33][13][48]. This heightened risk is attributed to polypharmacy, multiple comorbidities, regular off-label medication usage, and specific dosage regimens [6].

Given the growing multidrug resistance demonstrated by bacteria to nearly all classes of antibiotics, it is extremely beneficial to create a comprehensive database that thoroughly documents all potential ADIs [27].

Furthermore, advances in computational methods, especially artificial intelligence (AI), provide significant potential to improve the detection and management of ADIs in clinical practice [30].

2.2 AI in Medical Prescription Recommendation Systems

The creation of medical prescription recommendation systems designed to forecast and avert ADIs is a vital research focus, particularly as antibiotic resistance and adverse drug events become increasingly concerning [19]. These systems aim to enhance antibiotic usage while reducing harmful interactions [27].

Numerous studies have investigated the application of machine learning for predicting antibiotic combinations and their interactions. An antibiotic combination recommendation model (ACRM) was developed using a machine learning algorithm, demonstrating acceptable clinical relevance and accuracy in predictions, which suggests its potential to mitigate drug interactions [27].

Another study employed a label propagation framework to predict DDIs by incorporating clinical side effects and chemical structures, demonstrating the efficacy of computational models in forecasting harmful effects DDIs [50].

The PARS system combines semantic technologies with Multiple Criteria Decision Aiding (MCDA) to support antibiotic prescriptions[36]. Its ability to dynamically update a patient’s profile and integrate new information makes it more adaptable and reliable compared to static guidelines [36]. Similarly, AntibioHelp® functions as a clinical decision support tool that assists general practitioners in interpreting guidelines for patients without clear recommendations, ultimately boosting prescription confidence [41].

Frameworks that utilize molecular structure descriptors have been developed to effectively predict antibiotic interactions. These models can accurately anticipate interactions based on chemical structures, which helps in identifying drug pairs that work synergistically. [22]

Additionally, semi-mechanistic pharmacokinetic-pharmacodynamic (PKPD) models have been used to suggest optimal dosing regimens for antibiotic combinations, effectively addressing variability and uncertainty in parameters.[9]

2.3 Role of Large Language Models in Prescription Recommendations

Large Language Models (LLMs) exhibit considerable potential for enhancing Medical Prescription Recommendation Systems, as indicated by Sridharan (2024). This advancement has the capacity to fundamentally alter the techniques employed in prescribing and managing medications.

LLMs are advanced artificial intelligence models trained on extensive text data, allowing them to understand and generate human-like language [16]. They possess strong logic and reasoning skills and general world knowledge. Their capabilities extend to various Natural Language Processing (NLP) tasks, including understanding, summarization, question answering, creative writing, and even code generation [29].

In healthcare, LLMs present promising applications such as medication review and reconciliation where LLMs can assist in identifying errors in dosing regimens and drug-drug interactions, and suggesting dosage adjustments based on therapeutic drug monitoring and genomics [37]. Studies have assessed the abilities of models like ChatGPT, Claude-Instant, and Gemini in these processes, noting their potential to revolutionize medication management[37].

Clinical decision-making represents an additional domain in which LLMs can assist with the synthesis and analysis of patient data, enhance interaction between patients and medical professionals, and support the medical decision-making process [46].

In pharmacology and drug discovery, LLMs can transform knowledge query methods, enabling multi-round consultations on pharmacological questions [51]. They are capable of answering questions about drug properties, refining lead compound designs, and synthesizing research trends along with their limitations [51].

Personalized prescriptions using open-source LLMs, especially when combined with RAG, are being evaluated for their ability to generate tailored, patient-specific medication prescriptions [3].

However,LLMs face several limitations, including hallucinations that generate incorrect or misleading information [38], particularly in specific domains without real-time data. Outdated knowledge from old training datasets can also hinder accuracy [44]. To address these challenges, RAG has been introduced as a way to improve the medical question-answering abilities of LLMs by integrating external knowledge sources [49]. However, it may still encounter difficulties in more intricate situations that demand multiple rounds of information retrieval [45].

Furthermore, interacting with LLM can be computationally costly, which complicates the accuracy of complex tasks [53]. Ethical concerns arise from potential misuse, lack of transparency, and issues regarding patient privacy and informed consent [29].

2.4 Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) represents a methodological advancement aimed at augmenting the functionality of LLMs through the integration of pertinent information derived from an external database during the text generation process [24].

This approach alleviates hallucinations by grounding responses in factual sources [17]. In addition, RAG expands the model’s knowledge base by integrating current information, such as new research and clinical guidelines, with frameworks like Almanac [47] and MedRAG [45] to ensure more informed and contextually relevant output.

Studies show that RAG-enhanced LLMs consistently outperform standard models in terms of accuracy and evaluation metrics, such as iDISK2.0 [15], in the context of dietary supplements.

Furthermore, it improves transparency by facilitating source attribution, which allows users to verify the origins of retrieved information- an essential feature in fields such as medicine and scientific research, where credibility is crucial [52].

For elderly patients, it helps with medication management by cross-referencing curated datasets like DrugBank [40].

Additionally, systems like Chat2Data [53] utilize RAG to augment domain knowledge, enabling more efficient and low-code data analysis.

3 Problem Analysis

One of the biggest challenges in healthcare is predicting and detecting potential interactions between antibiotics and other medications, especially in patients with multiple health conditions who are taking complex medication regimens [2]. These patients are at a higher risk of experiencing harmful interactions, which can lead to negative side effects, treatment failures, and antibiotic resistance [13]. Although general-purpose LLMs can identify some drug interactions, they often miss specific interactions or provide incorrect information, posing risks in clinical settings [35]. This highlights the need for specialized models designed to detect ADIs and provide accurate, trustworthy guidance.

4 Methodology

This project focuses on designing, implementing, and evaluating an advanced system for automated prescription recommendations, utilizing a structured clinical dataset integrated with a Retrieval-Augmented Generation (RAG) [18] framework and Large Language Models (LLMs) [16]. The entire pipeline was developed using Python [26], ensuring modularity and reproducibility across experiments.

4.1 Data Acquisition and Preparation

The dataset for this study was constructed from *DrugBank* [43], a comprehensive biomedical resource that offers extensive data on drugs, including their chemical, pharmacological, and pharmaceutical properties. Its detailed coverage of drug-drug interactions made it particularly valuable for this research.

Data was initially retrieved in XML format from DrugBank, which adheres to an XML Schema Definition (XSD) for standardized representation. Due to the large file size, efficient parsing and memory management presented significant challenges. The `lxml` library [5] was utilized to facilitate efficient XML handling, specifically leveraging the `iterparse()` function for incremental parsing. This approach allowed for processing over 73,000 `<drug>` elements without overwhelming memory, with elements cleared after processing to conserve resources.

Antibiotics were identified by searching the `<categories>` element of each `<drug>` entry, using a pre-defined list of antibiotic-related keywords (e.g., “Antibiotic”, “Macrolides”, “Penicillins”). Entries matching these terms were flagged for extraction to ensure relevance.

For each antibiotic, key information was gathered, including unique identifiers, clinical properties, and interaction-related fields. Additionally, pharmacokinetic and pharmacodynamic details were collected to enhance the dataset. The resulting data were structured into Python dictionaries and serialized into a JSON file.

Text chunking was performed to prepare the data for use in a RAG system. Each antibiotic entry was divided into semantically coherent chunks, which were stored with relevant metadata. The final output was a corpus of text chunks, optimized for semantic embedding and vector database storage, providing a foundation for integrating DrugBank knowledge into an intelligent, RAG-enabled LLM pipeline.

4.2 Embedding Generation and Vector Database Establishment

To enable semantic retrieval within the RAG framework, the text data was converted into dense vector embeddings that represent the semantic meaning of each chunk in a high-dimensional space. This transformation allows for the comparison of textual similarity, facilitating more accurate information retrieval.

Embedding generation was performed using the `sentence-transformers` library [32], a widely used tool for producing high-quality sentence-level embeddings. The pre-trained model `all-MiniLM-L6-v2` [31] was selected for its computational efficiency and effectiveness in biomedical and technical contexts. This model offers a balance of speed and semantic accuracy, capturing complex relationships in biomedical language, such as drug-drug interactions and pharmacological mechanisms.

The embeddings were stored in `ChromaDB` [12], a lightweight vector database optimized for fast similarity searches. Each embedding was indexed alongside metadata, including the original `drugbank_id`, the type of text chunk (e.g., pharmacokinetics, interactions), and additional identifiers to ensure traceability and efficient retrieval.

This embedding and indexing pipeline forms the backbone of the RAG system, enabling efficient semantic retrieval and the seamless integration of domain-specific knowledge with LLMs. By combining this retrieval process with controlled natural language generation, the system delivers scalable, context-aware responses to support automated prescription recommendations in clinical and research settings.

4.3 Retrieval-Augmented Generation (RAG) System Development

The RAG pipeline was developed to enhance LLM responses by integrating context-specific data from the dataset. User queries, such as those related to drug interactions or treatment compatibility, were first converted into embeddings using the same `sentence-transformers` model employed during data preprocessing. These query embeddings were then utilized to perform semantic similarity searches within the `ChromaDB` collection, retrieving the top- k most relevant chunks (e.g., the top 15).

The retrieved text chunks were concatenated and included in a prompt alongside the original query, providing the LLM with contextual information to generate more relevant and grounded responses. A structured system prompt was carefully designed to instruct the LLM to act as an “antibiotic information assistant”, relying exclusively on the retrieved context to generate its responses. The system prompt also included safeguards to minimize hallucinations and discourage the generation of medical advice, ensuring the output maintained factual integrity.

4.4 LLM Integration and Evaluation

The final phase focused on integrating an LLM to generate responses and evaluate the effectiveness of the RAG system in a clinical context. The Mistral 7B [1] model was selected for its optimal balance of computational performance, resource efficiency, and suitability for local deployment. This choice was critical for ensuring reproducibility and controlled experimentation while maintaining complete data privacy and eliminating the need for external API dependencies.

The deployment began with the installation of the Ollama [25] platform, a lightweight environment designed for executing large language models locally. The Mistral 7B model was initialized using the command `ollama run mistral`, which downloads and activates the model for inference.

Interaction with the model was managed exclusively through Ollama’s locally hosted API, ensuring that all operations remained within the local environment. This setup allowed the Python application to submit queries and retrieve responses without external data transmission, ensuring experimental isolation and consistent access throughout the evaluation phase.

The effectiveness of the RAG framework was assessed by comparing the performance of the LLM with and without the RAG integration.

5 Results and Discussion

5.1 Dataset Construction and Characteristics

The resulting dataset, derived from DrugBank, comprises structured and semantically rich information on antibiotic medications and their interactions. The extraction process yielded over 73,000 <drug> entries, from which antibiotics were filtered using domain-specific keyword heuristics. This dataset, processed into a JSON format, contains detailed information for **546 identified antibiotic drugs**, each annotated with a wide range of pharmacological and clinical characteristics, including drug interactions, pharmacodynamics, pharmacokinetics, toxicity, dosage information, associated commercial products, and more.

A comprehensive description of the data fields included in each entry can be found in **Table 1 – Drug Information Fields**, which summarizes the main attributes extracted for each antibiotic in the constructed corpus.

Field	Description
drugbank_id	Primary DrugBank ID.
name	Name of the drug.
description	Description of the drug.
cas_number	Chemical Abstracts Service (CAS) registry number.
unii	Unique Ingredient Identifier.
indication	Medical indications for the drug’s use.
toxicity	Information regarding the drug’s toxicity.
groups	A list of drug groups the drug belongs to.
categories	A list of categories the drug falls under.
affected_organisms	A list of organisms affected by the drug.
pharmacodynamics	Pharmacodynamic properties.
mechanism_of_action	Description of how the drug produces its effect.
metabolism	Information on the drug’s metabolism.
absorption	Details on drug absorption.
half_life	The drug’s half-life.
protein_binding	Information on protein binding.
route_of_elimination	How the drug is eliminated from the body.
volume_of_distribution	Volume of distribution.
clearance	Drug clearance.
drug_interactions	A list of dictionaries, each containing the drugbank_id, name, and description of an interacting drug.
food_interactions	A list of known food interactions.
targets	A list of dictionaries, each containing the uniprot_id and name of the drug’s polypeptide targets.
dosages	A list of dictionaries, each with form, route, and strength information for dosages.
products	A list of dictionaries, each detailing a commercial product, including name, labeller, dosage_form, strength, route, generic status, approved status, and country.
synonyms	A list of alternative names or synonyms for the drug.
classification	A dictionary that may contain subclass, alternative_parent, and substituent.
external_identifiers	A list of dictionaries, each specifying an external resource and its corresponding identifier for the drug.

Table 1. Drug Information Fields

Although the dataset provides substantial coverage of clinically relevant information, some entries contain incomplete fields or missing values. These gaps present an opportunity for future enrichment by integrating complementary biomedical databases, thereby enhancing the completeness and utility of the dataset for downstream clinical and research applications.

To facilitate the integration of the dataset into a RAG framework, each drug entry was segmented into semantically coherent textual chunks through an automated preprocessing pipeline implemented in Python. A total of 298,811 chunks were generated, each categorized by thematic content—such as *Summary*, *Drug Interactions*, and *Dosage*—and enriched with structured metadata to support efficient semantic retrieval. This step was the most time-consuming stage of the data preparation process, as it involved parsing, classifying, and organizing large volumes of heterogeneous biomedical data.

A detailed classification of the chunk types and their descriptions is presented in **Table 2 – Chunk Types**.

Chunk Type	Description
Summary	Contains the general information about the drug, including its name, ID, description, indication, groups, categories, and affected organisms.
Pharmacology	Provides information on the pharmacodynamics and mechanism of action of the drug.
Pharmacokinetics	Contains data on the drug’s metabolism, absorption, half-life, protein binding, elimination route, volume of distribution, and clearance.
Toxicity	Provides details on the toxicity or adverse effects associated with the drug.
Drug Interaction	Contains details of drug interactions, including the interacting drug and a description of the interaction.
Food Interaction	Describes any food interactions with the drug.
Target	Contains information about the molecular targets of the drug, such as target name and UniProt ID.
Dosage	Provides dosage information for the drug, including the form, route, and strength.
Products	Contains information about products that contain the drug, including product names and manufacturers.
Synonyms	Lists synonyms for the drug.
Classification	Provides classification information, including the kingdom, superclass, class, subclass, and direct parent of the drug.
External Identifiers	Contains external identifiers for the drug from various resources.

Table 2. Chunk Types

Subsequently, each text segment was transformed into a high-dimensional vector representation. The resulting embeddings were stored in a Chroma vector database (ChromaDB)[12], which enables efficient semantic similarity searches during inference. This modular architecture supports the precise retrieval of contextually relevant information and contributes to the factual grounding and clinical relevance of language model outputs.

5.2 Qualitative Comparison Between LLM and RAG+LLM

A comparative evaluation was conducted to assess the impact of RAG on the performance of the LLM in answering specific clinical questions, particularly regarding drug interactions, which is the core objective of this study.

In the first example, the query investigated the safety implications of prescribing Sulfamethoxazole-Trimethoprim (TMP-SMX) to a diabetic patient who is already being treated with Metformin and Glibenclamide (Glyburide).

User Query Example 1

Para um paciente diabético em tratamento com Metformina e Glibenclamida, que desenvolveu uma infecção do trato urinário, posso prescrever Sulfametoxazol-Trimetoprima? Qual o risco de interação?

In response to the question, the baseline LLM provided a superficial and potentially misleading answer.

LLM-Only Response

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 Glibenclamina, 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 tratamento ideal depende de vários fatores individuais do paciente, como a natureza da infecção e sua intolerância a determinados medicamentos.

It stated that TMP-SMX “does not significantly interact” with these antidiabetic agents, without acknowledging the well-documented risks of hypoglycemia and altered drug clearance. The response failed to mention patient-specific risk factors—such as renal function—or offer any pharmacological rationale. It also gave general medical advice, contrary to safe clinical communication practices, despite not being based on precise sources.

LLM-RAG Response

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.

On the other hand, the LLM + RAG model delivered a much more context-aware and medically cautious output. The response clearly stated that it was based on information from the dataset and avoided making direct clinical recommendations, in strict accordance with the system prompt designed to limit hallucinations and prohibit unsolicited advice. Importantly, it accurately noted that TMP-SMX presents interaction risks with insulin and insulin-receptor agonists, increasing the chance of hypoglycemia—an observation that aligns with the pharmacological action of sulfamethoxazole, which can mimic sulfonylureas by stimulating insulin secretion.

Furthermore, the RAG-enhanced response acknowledged that, while it could not provide direct clinical guidance, it possessed sufficient dataset information to raise critical safety considerations, such as the need for glycemic monitoring and the importance of physician judgment when prescribing antibiotics in polypharmacy contexts. This response demonstrated a high degree of compliance with the structured system prompt, including accurate translation from English context into clear, professional Portuguese, avoiding speculative content, and adhering strictly to the dataset as a source of truth.

When comparing these results with the scientific literature, the RAG model showed a strong alignment. Peer-reviewed case studies and pharmacological reviews confirm that TMP-SMX can elevate the risk of hypoglycemia when used alongside sulfonylureas such as glibenclamide, particularly in patients with renal impairment [21][42]. Similarly, trimethoprim is known to inhibit metformin renal excretion by blocking OCT2, which can increase systemic levels of the drug and increase the risk of lactic acidosis [14]. These interactions, though relatively rare, are clinically relevant and require a patient-specific risk assessment, especially in elderly or really compromised individuals.

The LLM + RAG model succeeded not only in identifying these mechanistic risks but also in framing them within a cautious, evidence-based, and non-prescriptive communication style that mirrored best practices in pharmacovigilance. In contrast, the baseline LLM offered an oversimplified and potentially unsafe answer that contradicted established evidence. This result highlights the value of RAG architectures in biomedical settings - not just for factual completeness, but also for safety, alignment with system constraints, and clinical contextualization.

When questioned about the safety of combining azithromycin with atorvastatin, the LLM-only response provided an oversimplified and partially inaccurate summary. It mentioned potential increases in azithromycin blood levels and decreases in atorvastatin absorption—neither of which are supported by pharmacokinetic evidence. It failed to highlight the primary clinical concern: statin-induced muscle toxicity, including the possibility of rhabdomyolysis and myoglobinuria. Moreover, the baseline model once again ventured into speculative medical advice, suggesting dose adjustments and generalized warnings without appropriate contextual grounding in the mechanisms or severity of the interaction.

User Query Example 2

Um paciente usa Atorvastatina (uma estatina) diariamente e precisa de Azitromicina. Há risco de interação entre esses dois medicamentos e o que devo monitorar?

LLM-Only Response

Sim, existem algumas interações possíveis entre a Atorvastatina (uma estatina) e a Azitromicina. Uma delas é o 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-medicação sem consultar um profissional de saúde qualificado.

LLM-RAG Response

A atenção médica deve monitorar o risco ou a gravidade de miopatia, rhabdomyolysis (quebra dos músculos) e mioglobinúria (sangue de cor vermelha nos rins) ao combinar Azitromicina com Atorvastatina. Essa interação medicamentosa pode aumentar o risco ou a gravidade desses efeitos adversos.

By contrast, the LLM + RAG output was significantly more accurate, clinically focused, and aligned with established pharmacovigilance data. It identified the core safety issue—the increased risk of myopathy, rhabdomyolysis, and myoglobinuria—without venturing into unsupported claims. The response effectively avoided prescriptive advice and adhered to the guidelines set by the structured system prompt, which instructs the model to extract only from the available dataset content and provide factual, non-interventional insights.

Scientific evidence corroborates the concerns raised by the RAG-enhanced model. Although azithromycin is a relatively weak inhibitor of cytochrome P450 3A4 (CYP3A4) - and does not significantly alter plasma levels of atorvastatin in controlled trials - it may still increase the risk of muscle injury through other mechanisms [4]. These include inhibition of hepatic organic anion transporting polypeptides (OATP1B1), which facilitate statin uptake and clearance [34]. Furthermore, genetic polymorphisms such as variants SLCO1B1 can compound this risk, leading to increased systemic statin exposure and an increased probability of rhabdomyolysis and myoglobinuria, particularly in susceptible individuals [34].

Evidence from pharmacovigilance databases, including Vigibase, reports at least 53 cases of rhabdomyolysis associated with azithromycin-statin combinations, with 23% of cases occurring within ten days of initiating azithromycin [39]. While some controlled clinical studies report no significant interaction between azithromycin and atorvastatin in healthy volunteers [23], real-world cases indicate that individual risk factors such as age, renal function, polypharmacy, and genetic predisposition are critical modifiers of interaction severity [39].

In light of this evidence, current clinical guidelines advise caution when prescribing azithromycin to patients receiving statins, particularly those at increased risk for muscle toxicity [39]. Strategies such as temporarily discontinuing statins during short-term azithromycin therapy, or substituting with statins

less dependent on CYP3A4/OATP1B1 metabolism (e.g., fluvastatin or pravastatin), are commonly recommended in high-risk scenarios [10].

In summation, the RAG-augmented model provided a more accurate and clinically useful response compared to the standalone LLM. It emphasized the correct class of adverse events, avoided introducing false or misleading pharmacokinetic claims, and complied with rigorous content constraints. This reinforces the effectiveness of RAG systems in medical contexts—particularly when safety-sensitive interactions require grounded, context-specific reasoning.

6 Conclusion

This project successfully constructed a comprehensive dataset and developed an advanced AI-assisted system for prescription support, with a particular focus on antibiotic drug interactions. The overarching goal was to enhance the precision, reliability, and contextual understanding of automated decision-making tools in clinical pharmacology.

At the core of this work lies a custom-built dataset derived from DrugBank [43], a well-established biomedical resource. The dataset was curated to focus exclusively on antibiotics and their pharmacological properties, ultimately detailing 546 antibiotic drugs. The data extraction and cleaning process yielded over 73,000 `<drug>` entries, which were transformed into a structured JSON format. To prepare the dataset for semantic search and inference, these entries were segmented into nearly 300,000 semantically coherent text chunks. Chunks were metadata-enriched and embedded using the `all-MiniLM-L6-v2` model [31]. These embeddings were stored in ChromaDB[12], forming the foundation for rapid, context-aware retrieval.

Building on this infrastructure, a RAG[18] framework was implemented using the Mistral 7B language model[1]. The RAG architecture overcame key limitations of standalone LLMs—particularly hallucinations and reliance on outdated information—by grounding responses in factual data retrieved from the structured dataset. The assistant was governed by a strict system prompt, instructing it to avoid direct medical advice and produce responses that are accurate, traceable, and suitable for clinical support. This grounding also supports source attribution, a critical feature for transparency and trust in healthcare applications.

To evaluate the practical performance of the system, a qualitative comparison was conducted between the RAG-enhanced model and a baseline standalone LLM. The differences were significant. The standalone model often returned superficial, incomplete, or occasionally incorrect answers—sometimes downplaying well-documented clinical risks.

In contrast, the RAG-augmented system consistently generated medically cautious, contextually accurate, and literature-aligned outputs. Importantly, the model adhered strictly to the constraints imposed by its system prompt: it avoided unsupported claims, clearly flagged instances of insufficient information, and consistently refrained from offering direct clinical advice. These results underscore the critical value of combining semantic retrieval with generative reasoning in biomedical applications.

Despite these promising outcomes, several limitations must be acknowledged. The dataset, while robust, is derived solely from DrugBank[43] and therefore inherits its structural gaps and content limitations. Some entries may lack comprehensive data on pharmacokinetics, rare interactions, or contraindications, potentially affecting the model’s completeness. Furthermore, the system was evaluated on a limited set of manually selected queries; broader and more rigorous assessments—such as expert clinical reviews or real-world simulations—are needed for comprehensive validation.

Looking ahead, future work will focus on expanding both the dataset and the deployment capabilities of the system. A key area of development involves integrating real-time updates from biomedical literature and pharmacovigilance systems to ensure that the assistant reflects the most current evidence and safety data [11]. Another important direction is to connect the system to electronic health records (EHRs), enabling personalized and context-sensitive retrieval based on individual patient profiles—while maintaining strict adherence to privacy and compliance standards [20].

At the dataset level, further enrichment is planned through the integration of complementary biomedical resources. These additions aim to fill existing gaps, particularly regarding off-label uses, rare drug interactions, and underrepresented pharmacokinetic data.

Finally, deploying in controlled clinical settings under expert oversight will be essential for validating the system’s real-world utility, safety, and interpretability. The long-term vision is to establish a scalable, transparent, and trustworthy AI infrastructure for clinical decision support—one that aligns with best practices in biomedical knowledge representation and promotes patient safety.

References

1. Mistral AI. Mistral 7b. <https://mistral.ai/news/announcing-mistral-7b/>, 2023. Accessed: 2025-06-15.
2. Labib AL-Musawe, Carla Torre, Jose Pedro Guerreiro, Antonio Teixeira Rodrigues, Joao Filipe Raposo, Helder Mota-Filipe, and Ana Paula Martins. Polypharmacy, potentially serious clinically relevant drug-drug interactions, and inappropriate medicines in elderly people with type 2 diabetes and their impact on quality of life. *Pharmacology Research and Perspectives*, 8, 8 2020.
3. Mahdi Alkaeed, Sofiat Abioye, Adnan Qayyum, Yosra Magdi Mekki, Ilhem Berrou, Mohamad Abdallah, Ala Al-Fuqaha, Muhammad Bilal, and Junaid Qadir. Open foundation models in healthcare: Challenges, paradoxes, and opportunities with genai driven personalized prescription. 2 2025.
4. Guy W. Amsden, Olatunde Kuye, and Greg C.G. Wei. A study of the interaction potential of azithromycin and clarithromycin with atorvastatin in healthy volunteers. *Journal of Clinical Pharmacology*, 42:444–449, 2002.
5. Stefan Behnel, Martijn Faassen, and Ian Bicking. lxml: Xml and html with python, 2005.
6. Sara Bettonte, Mattia Berton, and Catia Marzolini. Magnitude of drug-drug interactions in special populations. *Pharmaceutics*, 14, 4 2022.
7. Mathieu S. Bolhuis, Prashant N. Panday, Arianna D. Pranger, Jos G.W. Kosterink, and Jan Willem C. Alffenaar. Pharmacokinetic drug interactions of antimicrobial drugs: A systematic review on oxazolidinones, rifamycines, macrolides, fluoroquinolones, and beta-lactams, 12 2011.
8. Bojana Božić-Cvijan, Miljana Labović, Marija Kukurić, and Milica Bajčetić. Drug-drug interactions of the reserve antibiotics: A narrative review. *Medicinska istraživanja*, 57:149–161, 2024.
9. M. J.E. Brill, A. N. Kristoffersson, C. Zhao, E. I. Nielsen, and L. E. Friberg. Semi-mechanistic pharmacokinetic–pharmacodynamic modelling of antibiotic drug combinations, 7 2018.
10. Hailey Burns, Landon Russell, and Zachary L. Cox. Statin-induced rhabdomyolysis from azithromycin interaction in a patient with heterozygous slco1b1 polymorphism. *Journal of Clinical Pharmacy and Therapeutics*, 46:853–855, 6 2021.
11. Jihye Choi, Nils Palumbo, Prasad Chalasani, Matthew M. Engelhard, Somesh Jha, Anivarya Kumar, and David Page. Malade: Orchestration of llm-powered agents with retrieval augmented generation for pharmacovigilance, 2024.
12. Chroma. Chroma: The ai-native open-source embedding database. <https://www.trychroma.com/> or <https://github.com/chroma-core/chroma>, 2024. Accessed: 2025-06-15.
13. Ann S. Doherty, Fiona Boland, Frank Moriarty, Tom Fahey, and Emma Wallace. Adverse drug reactions and associated patient characteristics in older community-dwelling adults: a 6-year prospective cohort study. *British Journal of General Practice*, 73:E211–E219, 3 2023.
14. Barbara Grün, Michael K. Kiessling, Jürgen Burhenne, Klaus Dieter Riedel, Johanna Weiss, Geraldine Rauch, Walter E. Haefeli, and David Czock. Trimethoprim-metformin interaction and its genetic modulation by oct2 and mate1 transporters. *British Journal of Clinical Pharmacology*, 76:787–796, 11 2013.
15. Yu Hou, Jeffrey R Bishop, Hongfang Liu, and Rui Zhang. Improving dietary supplement information retrieval: Development of a retrieval-augmented generation system with large language models. *Journal of medical Internet research*, 27:e67677, 3 2025.
16. Kyu Hwan Jung. Large language models in medicine: Clinical applications, technical challenges, and ethical considerations. *Healthcare Informatics Research*, 31:114 – 124, 2025.
17. Seonho Kim and Juntae Yoon. Vaiv bio-discovery service using transformer model and retrieval augmented generation. *BMC Bioinformatics*, 25, 12 2024.
18. Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.
19. Tai Han Lin, Hsing Yi Chung, Ming Jr Jian, Chih Kai Chang, Hung Hsin Lin, Ching Mei Yu, Ching Lih Perng, Feng Yee Chang, Chien Wen Chen, Chun Hsiang Chiu, and Hung Sheng Shang. Artificial intelligence-clinical decision support system for enhanced infectious disease management: Accelerating ceftazidime-avibactam resistance detection in klebsiella pneumoniae. *Journal of Infection and Public Health*, 17, 10 2024.
20. Siru Liu, Allison B. McCoy, and Adam Wright. Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines. *Journal of the American Medical Informatics Association*, 32:605–615, 4 2025.
21. Jordan Kit Mah, Daniel Negreanu, Suhaib Radi, and Stavroula Christopoulos. Trimethoprim-sulfamethoxazole-induced refractory hypoglycaemia successfully treated with octreotide. *BMJ Case Reports*, 14, 5 2021.
22. Daniel J. Mason, Ian Stott, Stephanie Ashenden, Zohar B. Weinstein, Idil Karakoc, Selin Meral, Nurdan Kuru, Andreas Bender, and Murat Cokol. Prediction of antibiotic interactions using descriptors derived from molecular structure. *Journal of Medicinal Chemistry*, 60:3902–3912, 5 2017.
23. Abdallah Abu Mellal, Nadia Hussain, and Amira S.A. Said. The clinical significance of statins-macrolides interaction: Comprehensive review of in vivo studies, case reports, and population studies, 2019.

24. Jing Miao, Charat Thongprayoon, Supawadee Suppadungsuk, Oscar A. Garcia Valencia, and Wisit Cheungpasitporn. Integrating retrieval-augmented generation with large language models in nephrology: Advancing practical applications, 3 2024.
25. Ollama. Run large language models locally. <https://ollama.com/>, 2024. Accessed: 2025-06-15.
26. Python Software Foundation. Python language reference, version 3.11. <https://www.python.org/>, 2023. Accessed: 2025-06-15.
27. Jia'an Qin, Yuhe Yang, Chao Ai, Zhaoshuai Ji, Wei Chen, Yingchang Song, Jiayu Zeng, Meili Duan, Wenjie Qi, Shutian Zhang, Zhuoling An, Yang Lin, Sha Xu, Kejun Deng, Hao Lin, and Dan Yan. Antibiotic combinations prediction based on machine learning to multicentre clinical data and drug interaction correlation. *International Journal of Antimicrobial Agents*, 63, 5 2024.
28. Andrei Flavius Radu, Simona Gabriela Bungau, Raluca Anca Corb Aron, Alexandra Georgiana Tarce, Ruxandra Bodog, Teodora Maria Bodog, and Ada Radu. Deciphering the intricate interplay in the framework of antibiotic-drug interactions: A narrative review, 10 2024.
29. Mohaimenul Azam Khan Raiaan, Md Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Saddam Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access*, 12:26839–26874, 2024.
30. Timothy M. Rawson, Nina Zhu, Ronald Galiwango, Derek Cocker, Mohammad Shahidul Islam, Ashleigh Myall, Vasin Vasikasin, Richard Wilson, Nusrat Shafiq, Shampa Das, and Alison H. Holmes. Using digital health technologies to optimise antimicrobial use globally, 12 2024.
31. Nils Reimers. all-minilm-l6-v2. <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>, 2021. Accessed: 2024-06-15.
32. Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 3982–3992. Association for Computational Linguistics, 2019.
33. Michael Rieder. Adverse drug reactions in children: Pediatric pharmacy and drug safety. *The journal of pediatric pharmacology and therapeutics : JPPT : the official journal of PPAG*, 24:4–9, 2019.
34. Annick Seithel, Sonja Eberl, Katrin Singer, Daniel Auge, Georg Heinkele, Nadine B. Wolf, Frank Dörje, Martin F. Fromm, and Jörg König. The influence of macrolide antibiotics on the uptake of organic anions and drugs mediated by oatp1b1 and oatp1b3. *Drug Metabolism and Disposition*, 35:779–786, 5 2007.
35. Justine Sicard, François Montastruc, Coline Achalme, Annie Pierre Jonville-Bera, Paul Songue, Marina Babin, Thomas Soeiro, Pauline Schiro, Claire de Canecaude, and Romain Barus. Can large language models detect drug–drug interactions leading to adverse drug reactions? *Therapeutic Advances in Drug Safety*, 16, 1 2025.
36. Souhir Ben Souissi, Mourad Abed, Lahcen El Hiki, Philippe Fortemps, and Marc Pirlot. Pars, a system combining semantic technologies with multiple criteria decision aiding for supporting antibiotic prescriptions. *Journal of Biomedical Informatics*, 99, 11 2019.
37. Kannan Sridharan and Gowri Sivaramakrishnan. Unlocking the potential of advanced large language models in medication review and reconciliation: A proof-of-concept investigation. *Exploratory Research in Clinical and Social Pharmacy*, 15, 9 2024.
38. Eric Steimetz, Jeremy Minkowitz, Elmer C. Gabutan, Joan Ngichabe, Hagar Attia, Mordechai Hershkop, Fatih Ozay, Matthew G. Hanna, and Raavi Gupta. Use of artificial intelligence chatbots in interpretation of pathology reports. *JAMA Network Open*, 7:E2412767, 5 2024.
39. Johanna Strandell, Andrew Bate, Staffan Hägg, and I. Ralph Edwards. Rhabdomyolysis a result of azithromycin and statins: An unrecognized interaction. *British Journal of Clinical Pharmacology*, 68:427–434, 9 2009.
40. Parinya Thetbanthad, Benjaporn Sathanarugsawait, and Prasong Praneetpolgrang. Application of generative artificial intelligence models for accurate prescription label identification and information retrieval for the elderly in northern east of thailand. *Journal of imaging*, 11, 1 2025.
41. Rosy Tsopra, Karima Sedki, Mélanie Courtine, Hector Falcoff, Antoine De Beco, Ronni Madar, Frédéric Mechaï, and Jean Baptiste Lamy. Helping gps to extrapolate guideline recommendations to patients for whom there are no explicit recommendations, through the visualization of drug properties. the example of antibiohelp® in bacterial diseases. *Journal of the American Medical Informatics Association*, 26:1010–1019, 6 2019.
42. Chunjiang Wang, Weijin Fang, Zuojun Li, and Linli Sun. Clinical features, risk factors, diagnosis, and treatment of trimethoprim-sulfamethoxazole-induced hypoglycemia. *Frontiers in Endocrinology*, 14, 2 2023.
43. David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, Jason R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam MacLejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, and Michael Wilson. Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Research*, 46:D1074–D1082, 1 2018.
44. Tongtong Wu, Linhao Luo, Yuan-Fang Li, Shirui Pan, Thuy-Trang Vu, and Gholamreza Haffari. Continual learning for large language models: A survey. 2 2024.
45. Guangzhi Xiong, Qiao Jin, Xiao Wang, Minjia Zhang, Zhiyong Lu, and Aidong Zhang. Improving retrieval-augmented generation in medicine with iterative follow-up questions. 8 2024.

46. Xiongwen Yang, Yi Xiao, Di Liu, Yun Zhang, Huiyin Deng, Jian Huang, Huiyou Shi, Dan Liu, Maoli Liang, Xing Jin, Yongpan Sun, Jing Yao, Xiao Jiang Zhou, Wankai Guo, Yang He, Wei Juan Tang, and Chuan Xu. Enhancing doctor-patient communication using large language models for pathology report interpretation. *BMC medical informatics and decision making*, 25:36, 12 2025.
47. Cyril Zakka, Rohan Shad, Akash Chaurasia, Alex R Dalal, Jennifer L Kim, Michael Moor, Robyn Fong, Curran Phillips, Kevin Alexander, Euan Ashley, Jack Boyd, Kathleen Boyd, Karen Hirsch, Curt Langlotz, Rita Lee, Joanna Melia, Joanna Nelson, Karim Sallam, Stacey Tullis, Melissa Ann Vogelsong, John Patrick Cunningham, and William Hiesinger. Almanac - retrieval-augmented language models for clinical medicine. *NEJM AI*, 1, 2 2024.
48. Maria Beatrice Zazzara, Katie Palmer, Davide Liborio Vetrano, Angelo Carfi, and Onder Graziano. Adverse drug reactions in older adults: a narrative review of the literature, 6 2021.
49. Fanlong Zeng, Wensheng Gan, Yongheng Wang, Ning Liu, and Philip S. Yu. Large language models for robotics: A survey. 11 2023.
50. Ping Zhang, Fei Wang, Jianying Hu, and Robert Sorrentino. Label propagation prediction of drug-drug interactions based on clinical side effects. *Scientific Reports*, 5, 7 2015.
51. Yingbo Zhang, Shumin Ren, Jiao Wang, Junyu Lu, Cong Wu, Mengqiao He, Xingyun Liu, Rongrong Wu, Jing Zhao, Chaoying Zhan, Dan Du, Zhajun Zhan, Rajeev K. Singla, and Bairong Shen. Aligning large language models with humans: A comprehensive survey of chatgpt’s aptitude in pharmacology. *Drugs*, 2 2024.
52. Yingbo Zhang, Shumin Ren, Jiao Wang, Junyu Lu, Cong Wu, Mengqiao He, Xingyun Liu, Rongrong Wu, Jing Zhao, Chaoying Zhan, Dan Du, Zhajun Zhan, Rajeev K Singla, and Bairong Shen. Aligning large language models with humans: A comprehensive survey of chatgpt’s aptitude in pharmacology. *Drugs*, 85:231–254, 2 2025.
53. Xinyang Zhao, Xuanhe Zhou, and Guoliang Li. Chat2data: An interactive data analysis system with rag, vector databases and llms. In *Proceedings of the VLDB Endowment*, volume 17, pages 4481–4484. VLDB Endowment, 2024.
54. Zhehuan Zhao, Zhihao Yang, Ling Luo, Hongfei Lin, and Jian Wang. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network. *Bioinformatics*, 32:3444–3453, 11 2016.