Experiences Developing an AI Chatbot in the Pharmaceutical Industry

Samuel Abedu
samuel.abedu@mail.concordia.ca
Concordia University
Montreal, QC, Canada

Saviour Owolabi
saviour.owolabi@ucalgary.ca
University of Calgary
Calgary, AB, Canada

Mayra Sofia Ruiz Rodriguez
mayrasofia.ruizrodriguez@mail.concordia.ca
Concordia University
Montreal, QC, Canada

Adam Yuen
adam.yuen@ucalgary.ca
University of Calgary
Calgary, AB, Canada

Cedric Lim Ah Tock
cedric.limahtock@mail.concordia.ca
Concordia University
Montreal, QC, Canada

Ali Zaraket
ali.zaraket@sandoz.com
Sandoz
Montreal, QC, Canada

Ahmad Abdellatif
ahmad.abdellatif@ucalgary.ca
University of Calgary
Calgary, AB, Canada

Emad Shihab
emad.shihab@concordia.ca
Concordia University
Montreal, QC, Canada

Nahal Nasseri
nahal.nasseri@sandoz.com
Sandoz
Montreal, QC, Canada

Abstract

Chatbots have become prevalent in industries such as health and pharmaceutical. They make access to streamlined information easier and faster. With recent advances in large language models (LLMs), chatbots powered by LLMs offer new opportunities to improve access to information, including information about drugs. However, developing chatbots in pharmaceutical practice remains challenging due to safety and regulatory compliance requirements and the unique nature of the data in this domain. In this paper, we share our experience deploying an LLM-based chatbot to answer drug-related questions. We highlight the challenges we encountered developing the chatbot for a global pharmaceutical company. Among these challenges are ensuring that our chatbot retrieves reliable, up-to-date information from trusted sources and that its responses are trustworthy. We also share the strategies we adopt to overcome these challenges and the lessons we learn from deploying the chatbot. We believe these insights can guide the BoatSE and the broader software engineering community when deploying chatbots for highly regulated domains like pharmaceuticals.

CCS Concepts

• **Applied computing → Computers in other domains**; • **Software and its engineering →** *Software design engineering*.

Keywords

Chatbot, Artificial Intelligence, Industry, Pharmaceutical, LLM

BoatSE '26, Rio De Janeiro, Brazil

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Samuel Abedu, Saviour Owolabi, Mayra Sofia Ruiz Rodriguez, Adam Yuen, Cedric Lim Ah Tock, Ali Zaraket, Ahmad Abdellatif, Emad Shihab, and Nahal Nasseri. 2025. Experiences Developing an AI Chatbot in the Pharmaceutical Industry. In *Proceedings of the 7th International Workshop on Bots and Agents in Software Engineering (BoatSE '26).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Chatbots are increasingly becoming important tools for organizations, offering conversational interfaces that automate information retrieval and other routine interactions [1]. The recent advances in LLMs have equipped chatbots with cutting-edge capabilities, which have contributed to the rise in adoption of chatbots in various domains such as customer support [21], financial assistants [7, 23], programming assistants [22], and in e-commerce [17]. The pharmaceutical and medical domains have also seen growth in the interest of chatbots. For instance, Google's Med-PaLM 2 was trained to answer medical questions, and it demonstrated superior performance over other models on healthcare QA benchmarks [24]. Steybe et al. [25] also introduced GuideGPT, a context aware chatbot for answering clinical questions on osteonecrosis medications.

These capabilities prove attractive to a global pharmaceutical company that receives a high volume of inquiries about dosage, contraindications, or storage of their products from health professionals (e.g., doctors and pharmacists), patients and the general public. Responding to such inquiries traditionally demands significant human effort, as relevant information is often dispersed across multiple sources. In this context, a chatbot can help automate responses, improving efficiency, reducing response times, and enhancing customer satisfaction.

However, developing a chatbot in the pharmaceutical domain presents unique challenges. The pharmaceutical domain is highly regulated, and information disseminated by pharmaceutical companies (and, by extension, their chatbots) must comply with stringent regulatory standards. These regulations exist for good reason, as inaccurate information about drugs can have serious consequences for patient health and safety.

As part of our ongoing collaboration with the global pharmaceutical company to transform drug-related inquiries with artificial intelligence (AI), we developed a drug information chatbot. The chatbot combines data from a pharmaceutical company's internal databases and reputable external sources, utilizing AI to analyze and synthesize this information to deliver accurate, relevant, and compliant responses in a reduced turnaround time. Our chatbot answers questions related to drug information, such as *"Can I take drug A while taking drug B?"* or *"How should I store drug X?"*. To ensure the integrity of the information provided by our chatbot, we implement several safeguards, including strict external source selection and a confidence scoring mechanism that provides transparency to users.

In this paper, we share our experience developing an LLM-based chatbot for a global pharmaceutical company. We believe this will provide valuable insights to both researchers and practitioners in the SE community when developing chatbots for highly regulated domains.

Paper Organization. The remainder of this paper is organized as follows. Section 2 presents the background, and Section 3 reviews related work. Section 4 describes the high-level architecture of our chatbot, while Section 5 discusses the technical and domain-specific challenges encountered during development and the strategies adopted to address them. In section 6, we share the lessons learned, and Section 7 concludes the paper.

2 Background

The pharmaceutical industry is highly regulated across the world. [12]. In the Canadian context, the regulatory authority, Health Canada (hereafter referred to as the regulator), maintains strict rules governing the communication of information about drug products. These requirements extend, by implication, to any chatbot endorsed by a pharmaceutical company, as it serves as a channel of communication with users. [4].

One key requirement is that communication about drug products be consistent with information in the product monograph [4]. The product monograph is an authoritative, publicly available document that provides comprehensive and factual information about a specific drug product. It is a mandatory component of market authorization and follows a standardized structure consisting of three parts:

(1) *Health Professional Information*, covering indications, contraindications, warnings and precautions, adverse reactions, drug interactions, dosage and administration instructions, pharmacological properties, and other clinical and safety data;
(2) *Scientific Information*, detailing clinical pharmacology, toxicology studies, product composition, and stability; and
(3) *Patient Medication Information*, providing information on the product's uses, correct administration, potential side effects, and when to seek medical attention.

The product monographs are lengthy documents with the *Health and Professional Information* and *Scientific Information* sections written in technical language intended for healthcare professionals and may include tables, figures, and specialized terminology to summarize clinical and pharmacological data. By contrast, the *Patient Medication Information* section is written in plain language, targeting a Grade 6–8 reading level to ensure accessibility for patients and caregivers.

In addition to these product monographs, the pharmaceutical company maintains non-public reference materials, such as Frequently Asked Questions (FAQs) and marketing documents, which contain supplementary information used by correspondents to provide accurate and consistent responses to inquiries from healthcare professionals or patients. Together, these materials form the knowledge base for a drug product, and all communication regarding that product (whether by humans or chatbots) must remain consistent with the information they contain.

LLMs are the state of the art in conversational AI, including chatbot applications. However, despite their remarkable general knowledge and strong performance on benchmarks such as MMLU and MedQA [19, 24], they cannot independently satsisfy these requirements. While they demonstrate a broad understanding of generic molecules and popular product offerings, they lack reliable awareness of specific product-level details, such as brand names, packaging forms, or dosage forms [14].

A separate, but equally important requirement of the regulator is that serious adverse events associated with the use of a drug are reported within a defined time frame, as part of its pharmacovigilance framework[13]. While not required, the regulator also encourages the reporting of all adverse events associated with the use of a drug product. Among other required data points, such reports must include the drug product suspected and the adverse reaction described. These requirements mean that beyond question-answering the chatbot must incorporate a system for detecting and taking appropriate action in cases where a user expresses an adverse event in the course of their interaction with the chatbot. Given the potentially large volume of user interactions, manually reviewing conversations for adverse event detection would be impractical, thus motivating the development of automated systems for detecting adverse events and drug names.

3 Related Works

Within the pharmaceutical domain, AI systems have long been explored for medication management, dosage optimization, and adverse event detection [5]. More recent studies have evaluated generative models specifically: Al-Dujaili et al. [2] assessed ChatGPT's accuracy in pharmacotherapy decision-making, finding moderate reliability across repeated sessions. Beavers et al. [3] compared chatbot responses to those from clinical pharmacists, concluding that while LLMs can produce clinically acceptable information, they fall short in completeness and safety. Han [11] and Li et al. [16] further identified risks of misinformation in specialized contexts such as prescription review. More positively, de Jesus et al. [8] demonstrated that retrieval-augmented generation (RAG) using official patient information leaflets can improve factual correctness and clarity in medication instructions.

Outside the pharmaceutical context, several case studies have reported successful deployment of retrieval-augmented chatbots for specialized industrial domains, including software engineering

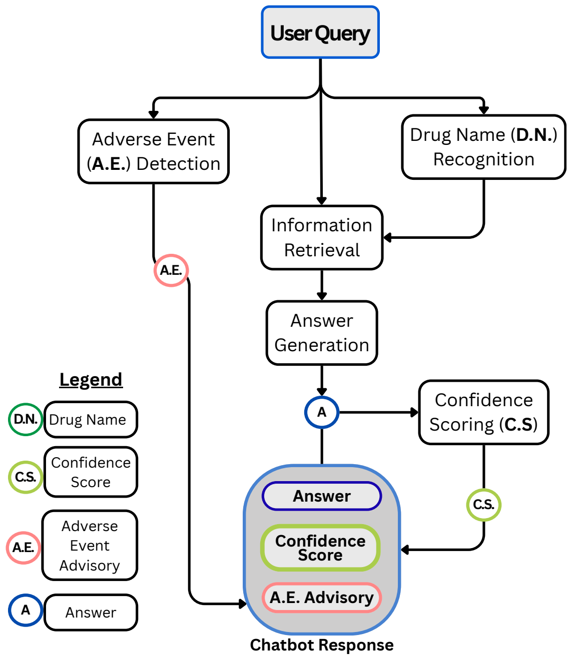

Figure 1: Flow of the Chatbot Interaction from User Question to Response Generation.

at Ericsson [6], aerospace [26], and tourism [15]. These efforts show how domain-grounded retrieval can mitigate hallucinations and improve contextual relevance, an insight increasingly applied to regulated settings such as pharmacovigilance, where Painter et al. [20] explored LLMs for drug-safety document retrieval.

These studies demonstrate the utility of LLMs for pharmaceutical-oriented tasks and highlight the promise of chatbots built around them in industrial contexts. In this paper we share our experience developing an LLM-based chatbot to address user inquiries about the drug products offered by a global pharmaceutical company.

4 Chatbot Design

We design the pharmaceutical chatbot to reduce the turnaround time for pharmaceutical questions, thereby improving the efficiency of health professionals by providing trusted, timely responses to their questions. Figure 1 shows the flow of the chatbot. The chatbot is designed using a retrieval-augmented generation (RAG) pipeline, augmented by specialized components such as custom entity recognizers trained on domain-specific data, a hybrid retrieval component that retrieves information from internal databases and the web, and a confidence scoring component to measure the trustworthiness of the response.

4.1 Corpus Creation for Internal Documents

Data Extraction. We ingest two data sources, i.e., the product monographs and the FAQ documents, as presented in section 2. When extracting information from the product monograph, we first extract all the top-level sections. If a section contains subsections, we also extract the subsections and create a reference for each subsection to their top-level section. Then we extract the tables as separate entities and link them to their captions as metadata. The FAQ questions are also extracted and mapped to their corresponding answers.

Data Preprocessing and Indexing. When preprocessing the monographs, the extracted sections, subsections, and tables are processed as individual data chunks. We summarize each data chunk and use the summary for embedding creation. The embeddings of the summaries are mapped to the original chunks and are indexed in a vector store for embedding-based retrieval.

For data from the FAQ documents, each question forms a unique chunk, and a direct reference is created to its corresponding answer. These question–answer pairs are indexed in a vector store, allowing the chatbot to perform embedding-based retrieval.

4.2 Chatbot Walkthrough

The chatbot accepts the user's question in natural language and maintains conversational state across multiple turns. Each request is processed by handling the session state of the conversation. When a user submits a question, the chatbot attaches a unique session identifier and retrieves any existing conversation context. This allows the chatbot to recall previous questions and responses, maintaining continuity across turns.

The chatbot applies custom-trained Named Entity Recognition (NER) models to extract drug names, variants, and adverse events from the user's question. If multiple formulations or dosage forms exist for the identified drug, the chatbot prompts the user to select the correct variant in its response.

Based on the recognized drug, the chatbot retrieves relevant information from multiple sources. First, it checks the FAQ documents of the identified drug, and uses a semantic similarity search to find a question among the FAQs that matches the user's question. If a match to a question in the FAQ is found (i.e. semantic similarity higher than the specified threshold), the corresponding answer forms the basis of the chatbot's response to the user's question.

If the user's question does not match a question in the FAQ document (i.e. semantic similarity lower than the specified threshold), the retrieval component expands the search to the product monograph of the drug and a web search. The retrieval component uses similarity search to find the most relevant paragraphs in the monograph to answer the question. When conducting the web search, the retrieval component restricts the search to trusted domains (e.g., official regulatory agencies).

The retrieved paragraphs and the web results, along with the user's question and session context, are passed to the LLM to generate a response. We construct a prompt that instructs the LLM to prioritize information from the monograph, to avoid speculation, and to produce a concise, well-structured answer. After the LLM generates an answer, we compute a confidence score based on factors such as the model's internal certainty (log probabilities), the similarity score between the answer and retrieved context, and the sources of the information retrieved from the web. The final response, with the references and the confidence score, is returned to the user as shown in Figure 2. The adverse events, if detected are recorded separately for patients safety reports.

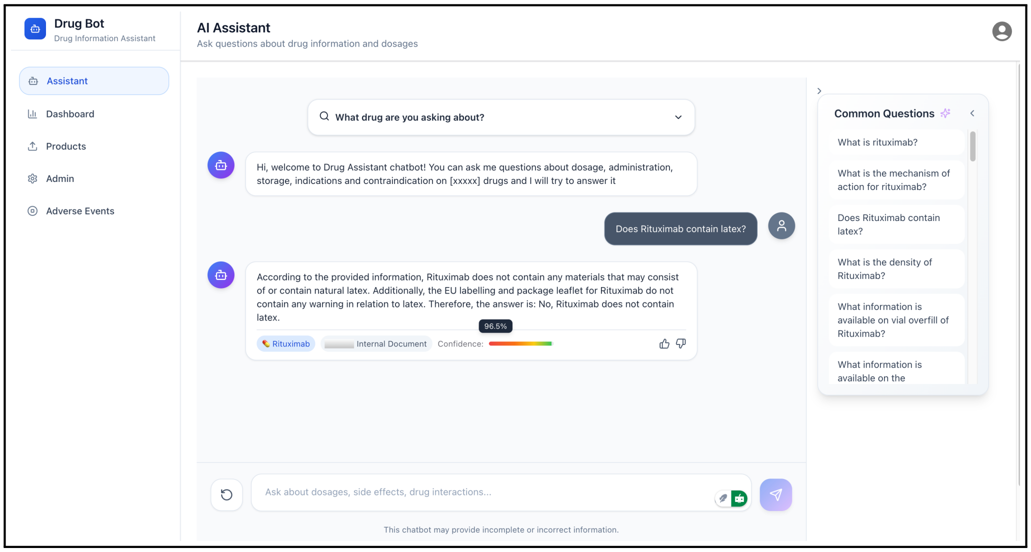

Figure 2: User Interface of the Pharmaceutical Chatbot. The UI shows the questions asked by the user, the response by the chatbot, the identified drug in the question, the source of the information used for the answer and the confidence score for the response.

5 Challenges and Mitigations

During the development of the chatbot for the pharmaceutical company, we had to address both technical and domain-specific challenges to meet regulatory and functional requirements. In this section, we discuss the challenges encountered and the strategies adopted to mitigate them.

5.1 Data Retrieval Challenges

Challenge I. Retrieving information accurately from product monographs.

Our experience in the early stages of developing the chatbot revealed that retrieving accurate information from the knowledge base for a drug posed a significant challenge. This challenge was particularly evident with the product monographs, which contain both unstructured text (often in long continuous paragraphs) and structured data in the form of tables. We found that the naive approach of using conventional dense retrieval with vector embeddings was inadequate, as simply partitioned chunks could omit important contextual information, reducing the effectiveness of the retrieval step. Moreover, direct embeddings of tables fail to capture their underlying structure and meaning, which limits their semantic representation. This limitation in the retrieval step affected the accuracy of the chatbot's responses.

Mitigation: To address this challenge and improve the performance of the retrieval step (and the subsequently generated answer), we implemented a summarization strategy that uses an LLM to summarize lengthy sections of the product monograph and generate descriptions of tabular data, prior to embedding. The resulting summaries and descriptions improve the effectiveness of the retrieval process, as the embeddings created from them capture more context. Importantly, while we use embeddings of the summaries for the retrieval step, we maintain a map of each summary to its original text, which is used in the generation step (rather than the summary itself). This approach is similar to the strategy decribed by Liu [18] and Eibich et al. [9].

Challenge II. Retrieving reliable and current web information from reputable web source.

After implementing and integrating the web search component into the chatbot, we observed that the search process frequently returned information from unverified sources such as blogs. Given that the pharmaceutical domain is highly regulated, and responses from the chatbot must be accurate, trustworthy, and meet regulatory standards, we cannot not rely on information from such sources. Doing so risks the chatbot producing answers based on outdated, speculative, or non-factual information, which can also violate regulatory requirements.

Mitigation: To address this, we implemented a **domain whitelisting and ranking system**. We consulted with our domain experts to make a list of reputable web domains from which we want the chatbot to retrieve information. We whitelisted these websites so that our web search component only retrieves and uses information from these pre-approved, reputable sources, such as Health Canada and official company publications. Also, in consultation with the domain experts, we implemented a ranking system to rank the pre-approved domains by reliability, ensuring that regulatory data took precedence over secondary literature or public repositories. For instance, information from Health Canada (which maintains a repository of verified information about drug products in Canada) is given higher priority and credence than information from PubMed, despite the latter's strong reputation. This filtering and weighting mechanism reduced noise and improved the factual integrity of generated responses.

5.2 Response Generation Challenges

Challenge III. Ensuring responses from the chatbot are trustworthy.

In designing our chatbot, an important consideration is that users must find the answers from our chatbot useful [10, 27]. For a critical domain like pharmaceuticals, the answers from our chatbot must be factual and trustworthy for users to find them useful. In this context, the utilization of LLMs presents a challenge as they are prone to *hallucination*, potentially generating uncertain and factually incorrect responses. In the pharmaceutical domain, such hallucinations can have adverse consequences for users and erode trust in the chatbot's reliability. As such, answers generated by the chatbot must be verifiable by users to build and maintain trust in the chatbot's responses.

Mitigation: To address this, we introduced a **confidence scoring strategy** that quantifies the trustworthiness of the chatbot's response. The chatbot computes a confidence score based on the source of information used to answer the question. For instance, if the question is answered using information in the FAQ document, the confidence score is calculated from the combination of the retrieval similarity scores between the user's question and the FAQ entries with token-level log probabilities from the LLM. If the question is answered using information from monographs and web content, it combines the question complexity, the semantic similarity score of the retrieved documents, source reliability from web retrievals, and the token-level log probabilities from the LLM to compute the confidence score. The confidence score is returned with each response, allowing users to interpret the chatbot's certainty in the response. The user might consider a response with a 95% confidence score as trustworthy, while a response with a 50% confidence score would be considered less trustworthy and therefore the user will not rely on it for decision-making. By combining multiple parameters to quantify the confidence of a response, the confidence score ensures transparency and prevents our chatbot from overconfidently providing low-certainty responses. In addition to confidence scores, the chatbot returns a list of consulted documents, including hyperlinks to them where available, for each response. This enables users to verify the chatbot's answer against the underlying source materials.

5.3 Challenges Related to User Interaction

Challenge IV. Robust recognition of drug names in questions.

Users may make mistakes when typing drug names, such as misspellings or incomplete names, which can hinder accurate retrieval. This presents a challenge for our chatbot, as inaccurate or incomplete drug names can hamper the information retrieval process. At the same time, requiring users to input drug names precisely imposes a practical burden that may hinder the user experience. In addition to the potential for mistyped drug names, some drug products have multiple names: a **generic name**, shared by all products containing the same active molecule, and a **brand name** that uniquely identifies the specific offering of the pharmaceutical company. A user may use either of these names in a question, necessitating a mechanism that consistently maps all known names to the same underlying product and knowledge base.

Mitigation: To address this challenge, we implemented a **type-ahead recommendation** feature in the chatbot interface that suggests drug names as the user types. This feature reduces the cognitive and typing burden on users, particularly for complex drug names, and helps minimize input errors that could otherwise hinder accurate retrieval. In addition, we trained our NER model for drug names to tolerate minor misspellings and integrated a spell correction mechanism that works at inference time to map misspelled drug names to their correct forms. Finally, to address cases where a drug has multiple names, we implemented a **normalization pipeline** that links brand and generic names of the company's products. During data ingestion and processing the user's question, all drug mentions are normalized to a single canonical form, ensuring that they resolve to the same underlying data collection and responses remain consistent regardless of which variant of the drug name appears in a question.

Challenge V. Handling multiple dosage forms and concentration levels of the same drug.

Some drugs in the database of the pharmaceutical company have multiple dosage forms and/or concentration levels. The different dosage forms sometimes have different concentration levels. For example, drug A has both tablets (with a concentration level of 50 mcg) and injections (concentration level of 100 mg/mL) as dosage forms. In some cases, users ask about the drug without specifying the variant. For example, *"How should I store drug A?"*. This situation leaves the chatbot uncertain about which variant to reference, increasing the risk of mixing up information in its responses.

Mitigation: To handle this ambiguity, we designed the chatbot with an **interactive clarification mechanism**. When a query about a product with multiple variants is presented, the chatbot responds with a clarification prompt that lists the available dosage forms or concentration levels. Once the user selects the relevant form, that choice is stored in the session context and persists until the user switches to another drug or formulation. This prevents misinterpretation and ensures that accurate information is provided to the user about the variant of interest.

5.4 Challenges Related to Compliance

Challenge VI. Identifying adverse events in user questions.

As part of regulatory compliance for monitoring and patient safety, our pharmaceutical chatbot must detect when users describe possible adverse drug reactions, as such cases require escalation or proper guidance. In line with pharmacovigilance responsibilities—and recognizing that regulators require prompt reporting of *serious* and *serious and unexpected* adverse reactions—the chatbot must maintain the capability to identify potential adverse events in user interactions. Each drug has its adverse event catalogue in the product monograph; however, relying on the LLM alone to detect adverse events in users' questions is not always accurate, especially because adverse event descriptions may be implicit or ambiguous and some reactions are uncommon and specific to a given product. For example, for a drug administered as a patch, patients might experience adverse reactions if the patch falls off frequently, and such incidents have to be reported.

Mitigation: To address this challenge, we curated an **adverse event corpus** derived from product monographs and reported adverse-event databases, and trained a specialized NER model to recognize adverse-event mentions in questions. When such events are detected, the chatbot invokes a safety workflow that advises the user on appropriate reporting procedures (through a formal channel) and prevents the generation of potentially unsafe recommendations. In addition, a record of the adverse event, including the drug discussed in the conversation, is securely stored and shared with the appropriate stakeholders for review and action.

6 Lessons Learned

Our experience building the chatbot for information inquiries in the pharmaceutical domain yielded valuable lessons. In this section, we share these lessons, as we believe they offer valuable insights to the BoatSE and broader software engineering communities.

Designing such chatbots requires an interdisciplinary collaboration. It is important to have knowledge and inputs from domain experts when building domain-specific chatbots. During the development, our domain experts observed that our chatbot missed details like the brand names of drugs and adverse events associated with some drugs. With guidance from our domain experts, we curated domain-specific training data, built custom entity recognizers, and ran continuous reviews with the domain experts. This ensured that the chatbot could accurately answer questions about specific drug brand names and accurately identify adverse events, which is required for regulatory compliance.

Also, within our trusted, whitelisted domains, our domain experts found that information retrieved when using a drug's generic name can include details from brands other than our partner pharmaceutical company. The experts explained the issue with this is the same drug from a different manufacturers can have different inactive ingredients and concentration levels. Thus, using this information to answer questions can lead to inconsistent and incorrect responses. As a lesson, we always ensure that priority is given to the most trusted web domains. In our chatbot, results, with information from our partner pharmaceutical company ranked highest, followed by regulatory agencies and then secondary literature.

Prompting in Regulated Domains Requires Explicit Guardrails. Define the role and scope of the LLM when developing LLM-based chatbots. When developing chatbots for specific domains, it is essential to guard the chatbot from responding to questions not related to the domain to prevent abuse of the chatbot. For instance, in our chatbot, we have a default response when users ask questions that are irrelevant, like *"what is the recipe for apple pie"*. Also, we explicitly define additional guardrails in our prompt to safeguard the chatbot to ensure the responses are safe. For example, we instruct the model to use information from the approved sources only, giving priority to those from highly rated pages and not rely on its internal knowledge, which could be at risk of being outdated or incorrect.

7 Conclusion

In this paper, we share our experience developing and deploying a retrieval-augmented (RAG) LLM-based chatbot for pharmaceutical question answering. Our chatbot integrates domain-specific entity recognition to identify names of drugs in users' questions, embedding-based retrieval to obtain information from internal documents like monographs and FAQs for answer generation, performs web searches on a curated whitelist of domains, and uses a confidence-scoring framework to enhance the trustworthiness of the chatbot's responses. We also implement a patient safety feature that detects and reports any adverse events in the user's question to align with regulatory compliance.

In the paper, we highlight some of the challenges we encountered while deploying the chatbot and share the strategies we adopted to mitigate these challenges. We highlight that domain expert guidance is key when building safety-critical systems, that summary-based retrieval can improve performance, and that ensuring information for answering users' questions is sourced from reputable domains is essential. Our mitigation strategies and the lessons we share can serve as a reference and guidance for software engineers building chatbots in highly regulated domains and for the BoatSE community.

References

[1] Eleni Adamopoulou and Lefteris Moussiades. 2020. Chatbots: History, Technology, and Applications. *Machine Learning with Applications* 2 (Dec. 2020), 100006. doi:10.1016/j.mlwa.2020.100006

[2] Zahraa Al-Dujaili, Sarah Omari, Jey Pillai, and Achraf Al Faraj. 2023. Assessing the Accuracy and Consistency of ChatGPT in Clinical Pharmacy Management: A Preliminary Analysis with Clinical Pharmacy Experts Worldwide. *Research in Social and Administrative Pharmacy* 19, 12 (Dec. 2023), 1590–1594. doi:10.1016/j.sapharm.2023.08.012

[3] Jennifer Beavers, Ryan F. Schell, Halden VanCleave, Ryan C. Dillon, Austin Simmons, Huiding Chen, Qingxia Chen, Shilo Anders, Matthew B. Weinger, and Scott D. Nelson. 2023. Evaluation of Inpatient Medication Guidance from an Artificial Intelligence Chatbot. *American journal of health-system pharmacy: AJHP: official journal of the American Society of Health-System Pharmacists* 80, 24 (Dec. 2023), 1822–1829. doi:10.1093/ajhp/zxad193

[4] Health Canada. 2016. Guidance Document: Product Monograph. https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/product-monograph/guidance-document-product-monograph.html.

[5] Sri Harsha Chalasani, Jehath Syed, Madhan Ramesh, Vikram Patil, and T.M. Pramod Kumar. 2023. Artificial Intelligence in the Field of Pharmacy Practice: A Literature Review. *Exploratory Research in Clinical and Social Pharmacy* 12 (Oct. 2023), 100346. doi:10.1016/j.rcsop.2023.100346

[6] Daksh Chaudhary, Sri Lakshmi Vadlamani, Dimple Thomas, Shiva Nejati, and Mehrdad Sabetzadeh. 2024. Developing a Llama-Based Chatbot for CI/CD Question Answering: A Case Study at Ericsson. In *2024 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE Computer Society, 707–718. doi:10.1109/ICSME58944.2024.00075

[7] Zhisheng Chen. 2025. Revolutionizing Finance with Conversational AI: A Focus on ChatGPT Implementation and Challenges. *Humanities and Social Sciences Communications* 12, 1 (March 2025), 388. doi:10.1057/s41599-025-04725-y

[8] Davi Dos Reis de Jesus, Antônio Pereira de Souza Júnior, Elisa Tuler de Albergaria, Adriana Silvina Pagano, Isaias Jose Ramos de Oliveira, Cristiane Dos Santos Dias, Eura Martins Lage, Flavia Ribeiro de Oliveira, Juliana Almeida Oliveira, Igor de Carvalho Gomes, Leonardo Chaves Dutra da Rocha, and Zilma Silveira Nogueira Reis. 2025. Enhanced LLM-supported Instructions for Medication Use through Retrieval-Augmented Generation. *Computers in Biology and Medicine* 198, Pt A (Oct. 2025), 111135. doi:10.1016/j.compbiomed.2025.111135

[9] Matouš Eibich, Shivay Nagpal, and Alexander Fred-Ojala. 2024. ARAGOG: Advanced RAG Output Grading. doi:10.48550/arXiv.2404.01037 arXiv:2404.01037

[cs].

[10] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience* 5, 1 (April 2020), 3. doi:10.1007/s41233-020-00033-2

[11] Jung Yeol Han. 2025. Usefulness and Limitations of Chat GPT in Getting Information on Teratogenic Drugs Exposed in Pregnancy. *Obstetrics & Gynecology Science* 68, 1 (Jan. 2025), 1–8. doi:10.5468/ogs.24231

[12] Shweta Handoo, Vandana Arora, Deepak Khera, Prafulla Kumar Nandi, and Susanta Kumar Sahu. 2012. A Comprehensive Study on Regulatory Requirements for Development and Filing of Generic Drugs Globally. *International Journal of Pharmaceutical Investigation* 2, 3 (2012), 99–105. doi:10.4103/2230-973X.104392

[13] Health Canada. 2018. Reporting Adverse Reactions to Marketed Health Products: Guidance Document for Industry. https://www.canada.ca/en/health-canada/services/drugs-health-products/public-involvement-consultations/medeffect-canada/consultation-reporting-adverse-reactions-marketed-health-products-guidance-document-industry/document.html#s1 Guidance document, Government of Canada.

[14] Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large Language Models Struggle to Learn Long-Tail Knowledge. In *Proceedings of the 40th International Conference on Machine Learning (ICML'23, Vol. 202)*. JMLR.org, Honolulu, Hawaii, USA, 15696–15707.

[15] Pattama Krataithong, Watchira Buranasing, Marut Buranarach, Theerawat Wutthitasarn, Pattaraporn Meeklai, and Patipat Tumsangthong. 2025. Tourism Chatbot Framework: Enhancing Visitor Experience Through GraphRAG and AI Chatbot. In *2025 IEEE International Conference on Cybernetics and Innovations (ICCI)*. 1–6. doi:10.1109/ICCI64209.2025.10987390

[16] Lulu Li, Pengqiang Du, Xiaojing Huang, Hongwei Zhao, Ming Ni, Meng Yan, and Aifeng Wang. 2025. Comparative Analysis of Generative Artificial Intelligence Systems in Solving Clinical Pharmacy Problems: Mixed Methods Study. *JMIR medical informatics* 13 (July 2025), e76128. doi:10.2196/76128

[17] Xiangci Li, Zhiyu Chen, Jason Ingyu Choi, Nikhita Vedula, Besnik Fetahu, Oleg Rokhlenko, and Shervin Malmasi. 2025. Wizard of Shopping: Target-Oriented E-commerce Dialogue Generation with Decision Tree Branching. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 13095–13120. doi:10.18653/v1/2025.acl-long.641

[18] Jerry Liu. 2023. A New Document Summary Index for LLM-powered QA Systems. https://www.llamaindex.ai/blog/a-new-document-summary-index-for-llm-powered-qa-systems-9a32ece2f9ec

[19] Harsha Nori, Nicholas King, Scott Mayer McKinney, Dean Carignan, and Eric Horvitz. 2023. Capabilities of GPT-4 on Medical Challenge Problems. arXiv:2303.13375 [cs] doi:10.48550/arXiv.2303.13375

[20] Jeffery L. Painter, Olivia Mahaux, Marco Vanini, Vijay Kara, Christie Roshan, Marcin Karwowski, Venkateswara Rao Chalamalasetti, and Andrew Bate. 2023. Enhancing Drug Safety Documentation Search Capabilities with Large Language Models: A User-Centric Approach. In *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*. 49–56. doi:10.1109/CSCI62032.2023.00015

[21] Philipp Reinhard, Mahei Manhai Li, Christoph Peters, and J. M. Leimeister. 2024. Generative AI in Customer Support Services: A Framework for Augmenting the Routines of Frontline Service Employees. social science research network:4862940 doi:10.2139/ssrn.4862940

[22] Steven I. Ross, Fernando Martinez, Stephanie Houde, Michael Muller, and Justin D. Weisz. 2023. The Programmer's Assistant: Conversational Interaction with a Large Language Model for Software Development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 491–514. doi:10.1145/3581641.3584037

[23] Kuldeep Singh, Simerjot Kaur, and Charese Smiley. 2024. FinQAPT: Empowering Financial Decisions with End-to-End LLM-driven Question Answering Pipeline. In *Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF '24)*. Association for Computing Machinery, New York, NY, USA, 266–273. doi:10.1145/3677052.3698682

[24] Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R. Pfohl, Heather Cole-Lewis, Darlene Neal, Qazi Mamunur Rashid, Mike Schaekermann, Amy Wang, Dev Dash, Jonathan H. Chen, Nigam H. Shah, Sami Lachgar, Philip Andrew Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera y Arcas, Nenad Tomašev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale R. Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. 2025. Toward Expert-Level Medical Question Answering with Large Language Models. *Nature Medicine* 31, 3 (March 2025), 943–950. doi:10.1038/s41591-024-03423-7

[25] David Steybe, Philipp Poxleitner, Suad Aljohani, Bente Brokstad Herlofson, Ourania Nicolatou-Galitis, Vinod Patel, Stefano Fedele, Tae-Geon Kwon, Vittorio Fusco, Sarina E. C. Pichardo, Katharina Theresa Obermeier, Sven Otto, Alexander Rau, and Maximilian Frederik Russe. 2025. Evaluation of a context-aware chatbot using retrieval-augmented generation for answering clinical questions on medication-related osteonecrosis of the jaw. *Journal of Cranio-Maxillofacial Surgery* 53, 4 (April 2025), 355–360. doi:10.1016/j.jcms.2024.12.009

[26] Surendra Yadav. 2024. AeroQuery RAG and LLM for Aerospace Query in Designs, Development, Standards, Certifications. In *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. 1–6. doi:10.1109/CONECCT62155.2024.10677028

[27] Jennifer Zamora. 2017. I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations. In *Proceedings of the 5th International Conference on Human Agent Interaction (HAI '17)*. Association for Computing Machinery, New York, NY, USA, 253–260. doi:10.1145/3125739.3125766