Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240639

Developing a Generative AI-Powered Chatbot for Analyzing MAUDE Database

Yue YU[a], Yuheng SHI[a], Yubo FENG[b] and Yang GONG[a,1]
[a]*University of Texas Health Science Center at Houston, Houston, Texas, USA*
[b]*Vanderbilt University, Nashville, Tennessee, USA*
ORCiD ID: Yang Gong https://orcid.org/0000-0002-0864-8368

Abstract. This paper presents a chatbot that simplifies accessing and understanding the open-access records of adverse events related to medical devices in the MAUDE database. The chatbot is powered by generative AI technology, enabling count and search queries. The chatbot uses the openFDA API and GPT-4 model to interpret users' natural language queries, generate appropriate API calls, and summarize adverse event reports. The chatbot also provides a downloadable link to the original reports. The model's performance in generating accurate API calls was assessed and improved by training it with few-shot examples of query-URL pairs. Additionally, the quality of content-based summaries was evaluated by human expert ratings. This initiative is a significant step towards making patient safety data accessible, replicable, and easily manageable by a broader range of researchers.

Keywords. Patient safety, incident reports, data extraction, information retrieval

1. Introduction

There is a growing interest in analyzing adverse events related to medical devices using open-access data from the Manufacturer and User Facility Device Experience (MAUDE) database [1]. The number of MAUDE-based studies has tripled from 26 in 2020 to 78 in 2023, with a total of 413 reports since 2000 [2]. Users can access the reports through a web-based search interface by applying filters such as product problem, event type, and brand name and specifying a time window. However, selecting a proper filter up to a 10-year timespan and limiting the search to 500 reports per query poses challenges [2].

OpenFDA is a platform that provides the public with access to FDA-regulated product data [3], aiming to promote transparency and enable research, analysis, and innovation for public health and safety. OpenFDA offers an application programming interface (API) that developers can use to retrieve MAUDE reports, but it may be challenging for non-technical lay users. Only two of the 413 MAUDE-based studies have utilized the OpenFDA API and mostly used the basic or advanced web interface, lacking replicability and sustainability for data sharing [2].

Early studies applying hybrid models to understand the MAUDE dataset exhibit good performance but require intense labor in developing customized training sets [4, 5]. ChatGPT, a pre-trained generative AI model, is transforming healthcare data management and communication [6]. This project aims to utilize ChatGPT's language

[1] Corresponding Author: Yang Gong, MD, PhD, FIAHSI. 7000 Fannin St, Ste 600, Houston, Texas, USA 77030; E-mail: Yang.Gong@uth.tmc.edu.

comprehension and generative capabilities to translate natural language queries into API calls and summarize retrieved reports. The chatbot is proposed as an easy-to-use entry point for users to access the MAUDE database via the openFDA API, enabling users to define and refine their inquiries and enhance the database's accessibility and openness.

2. Methods

To create a generative AI-powered chatbot for the MAUDE database, a structured approach was taken to understand natural language queries and provide users with content-based concise summaries of adverse event reports. The methodology consisted of three primary steps: *transforming text to URL, fetching data, and summarizing MAUDE reports* (**Figure 2**). We adopted LangChain, a Python package for building large language model applications, to generate prompt templates, perform similarity searches, and send API calls. Additionally, Chainlit, an open-source package, is used to build customized chatbot interfaces. The temperature parameter of GPT-4 is set to 0.0 to minimize variation between summaries and ensure reproducibility.

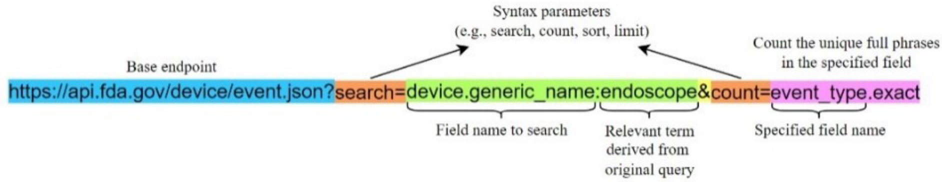

Figure 1. An illustration of a query URL.

Transforming text to URL: Query URLs with parameters such as search, count, sort, and limit are generated to match the components of the user query using GPT-4 (**Figure 1**). Dynamic few-shot examples and a search field schema are used to enhance the accuracy of the generated URLs, which are reported to be effective in Text-to-SQL studies [7]. A dataset of 80 question-URL pairs covering the most popular MAUDE domains [8] was created to implement dynamic few-shot examples and stored in a vector database. During the query phase, three examples are dynamically selected from the dataset based on the semantic matching between the user's question and the examples' questions [9]. The search field schema, downloaded from the openFDA website, imposes a constraint on the available search fields [10] for the model to choose.

Fetching data: an API call based on the URL is then sent to the openFDA to fetch reports. All fetched data are exhibited in JSON formats with structured and easy-to-process data fields. The retrieved reports are then filtered by data fields “date_received”, “patient_problems”, “device_problems”, and descriptive texts, offering details about the adverse event and the manufacturer's response. Given a response error such as an invalid URL, the error message is added to the prompt query during the transforming step to reflect on the causes and refine the URL generation with at most three iterations.

Summarizing MAUDE reports: To generate a concise summary of the retrieved data based on the user's query, we created a template for describing the specific details of the retrieved report, including event date, device, patient's problem, device's problem, a brief description of the cause of the event, and so on. A word count limit of 150 has been set to ensure brevity. A template and prompt are shown in **Figure 2**.

To evaluate the model's performance in generating URLs, a separate test set of 20 question-URL pairs was thoughtfully created to represent a diverse range of queries [11].

We compared using this test set with and without the few-shot examples. Two evaluators rated the model's performance in term of quality of generated summaries of the reports based on criteria of readability, correctness, coherence, and usefulness using a 5-point Likert scale, ranging from “Strongly disagree” as (1) to “Strongly agree” as (5) [12].

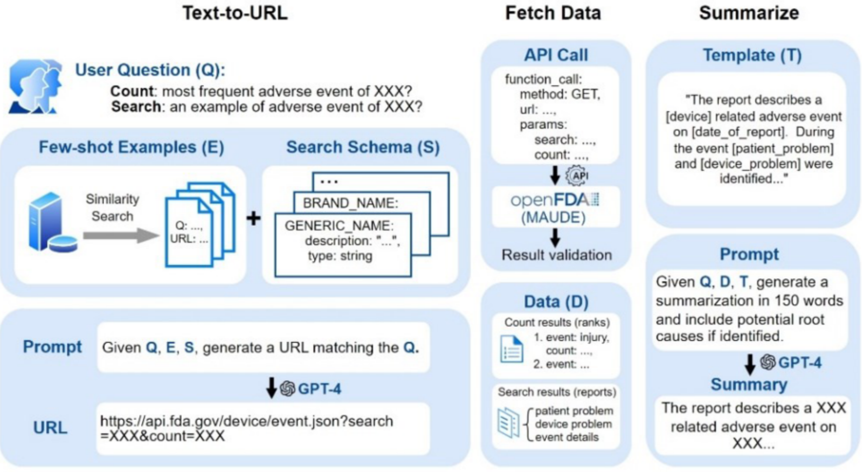

Figure 2. A workflow of the generative AI-powered chatbot.

3. Results

The chatbot connected by the openFDA API can execute two types of queries: count and search queries, as shown in **Figure 3**. Count queries provide information on the frequency of adverse events associated with a medical device, while search queries offer specific examples. Both queries enable lay users to access and download relevant information easily.

Our initial evaluation showed that the model's URL generation performance increased by 35% when dynamic few-shot examples were added to the prompt. The quality of GPT-4 generated summaries is deemed easy to follow and understand (readability, Mean [M]=5.00, Standard Deviation [SD]=0.00), using appropriate terminologies and concepts with occasional misinterpretation of the relations (correctness, M=4.52, SD=0.57). The majority of key information was successfully captured from the original report (coherence, M=4.35, SD=1.32), offering helpful information that matched the intent of the query well (usefulness, M=4.48, SD=1.23).

4. Discussion

This paper introduces our effort to utilize an innovative chatbot designed to enable lay users to search the MAUDE database. This tool simplifies search strategy development and identifies frequent events or specific reports. It even provides a URL for users to access and save the retrieved reports. The chatbot is in the prototype phase and subject to evaluation and iterative redesign through practical use.

Although the model accurately generated most URLs for test questions, complex search terms and missing generic names posed challenges. Further research is necessary to improve the few-shot dataset and expand its coverage or create a reference table connecting different MAUDE topics with their predefined search criteria. This could be useful in developing a hybrid approach that combines rule-based and generative methods to generate more accurate URLs. Allowing users to customize search criteria by offering them suggested search field options represents another promising avenue for improving the system's functionality and usability.

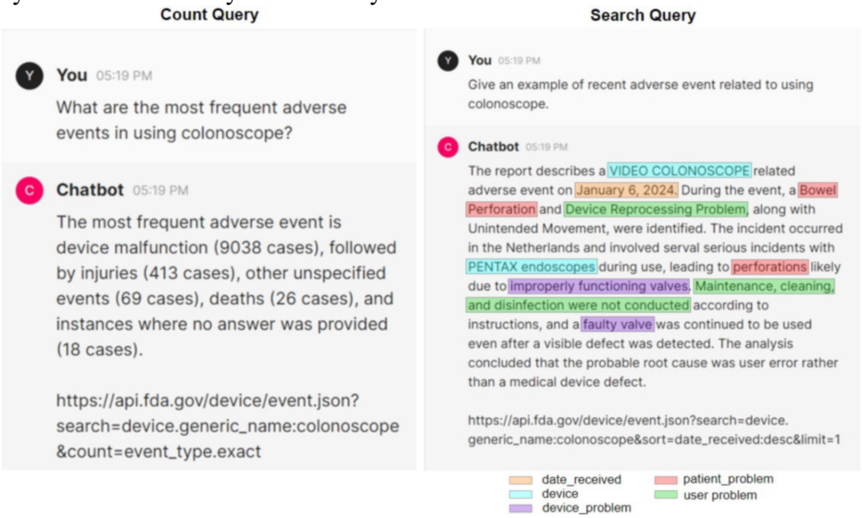

Figure 3. Illustrations of a count query and a search query. Two types of query responses are described. A count query shows the most frequent adverse events. A search query highlights structured fields and free-text fields, and provides a summary of a specific report. A URL linking to the original data or report is included.

During the data fetching step, it is essential to carefully examine the response that has been fetched to identify any errors. To ensure a successful search, it is crucial to strike a balance between freedom and restriction. When an error message indicates a faulty URL, the system should refine the URL iteratively until the desired results are retrieved. However, it is important to be cautious of being overly restrictive with search criteria, such as narrowing the search date range or omitting potential generic names, as this can lead to instances where no reports can be identified. A promising direction for improvement is establishing a workflow with two specialized agents: one dedicated to generating URLs and another to critically evaluating responses and offering iterative improvements of the URLs or providing explanations and suggestions to the user, ultimately enhancing URL accuracy and user understanding.

Although a summary template delineates critical data fields for generating a content-based report summary, the summary quality is contingent upon multiple factors, including the quality of the original report and its representativeness of the event. An iterative evaluation and user feedback mechanism is expected to keep users and domain experts in the loop, identifying the shortcomings and assessing the quality of the summaries. In the long run, this mechanism would facilitate the creation of a repository of "gold standard" or best cases, enabling a retrieve-reuse-revise-retain (4R) cycle driven by case-based reasoning, promoting learning and quality improvement.

Advanced functions beyond the basic count and search queries are currently under development. It is important to prioritize finding solutions to technical barriers and challenges that lay users face when using the search interface. Medical informaticians are encouraged to make publicly available databases more accessible to lay users due to the growing prevalence of MAUDE-based studies. However, we should not overlook the fact that underreporting and low quality of raw reporting continues to be significant obstacles in open-access datasets of patient safety events.

5. Conclusions

A chatbot powered by generative AI that leverages the openFDA API simplifies the process of accessing and understanding adverse event reports in the MAUDE database. This initiative represents our effort to enhance the credibility and trustworthiness of MAUDE-based studies in terms of data extraction, information retrieval, and knowledge acquisition from lessons learned, promote open-access data exchange, and advance replicability and reproducibility, thus improving scientific rigor in data-driven patient safety research.

Acknowledgement

This project is supported by UTHealth Houston Innovation for Cancer Prevention Research Training Program Postdoctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP210042)

References

[1] Wang E, Kang H, Gong Y. Generating a Health Information Technology Event Database from FDA MAUDE Reports. Stud Health Technol Inform. 2019;264:883-7.
[2] Li X, Feng Y, Gong Y, Chen Y. Assessing the Reproducibility of Research Based on the Food and Drug Administration Manufacturer and User Facility Device Experience Data. J Patient Saf. 2024.
[3] Kass-Hout TA, Xu Z, Mohebbi M, Nelsen H, Baker A, Levine J, et al. OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data. J Am Med Inform Assoc. 2016;23(3):596-600.
[4] Kang H, Gong Y. Creating a database for health IT events via a hybrid deep learning model. J Biomed Inform. 2020;110:103556.
[5] Kang H, Gong Y, editors. Creating a Database for Health IT Event Reports by Using a Hybrid Deep Learning Model. AMIA; 2019.
[6] Yu P, Xu H, Hu X, Deng C. Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. Healthcare (Basel). 2023;11(20).
[7] Tarbell R, Choo KR, Dietrich G, Rios A. Towards Understanding the Generalization of Medical Text-to-SQL Models and Datasets. AMIA Annu Symp Proc. 2023;2023:669-78.
[8] Ramai D, Deliwala SS, Mozell D, Facciorusso A, Chandan S, Persaud A, et al. Adverse Events of Endoscopic Clip Placement: A MAUDE Database Analysis. J Clin Gastroenterol. 2024;58(1):76-9.
[9] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.
[10] Fda.gov. Searchable Fields 2014 [Available from: https://open.fda.gov/apis/device/event/searchable-fields/.
[11] Yuan C, Ryan PB, Ta C, Guo Y, Li Z, Hardin J, et al. Criteria2Query: a natural language interface to clinical databases for cohort definition. J Am Med Inform Assoc. 2019;26(4):294-305.
[12] Park J, Fang Y, Ta C, Zhang G, Idnay B, Chen F, et al. Criteria2query 3.0: Leveraging Generative Large Language Models for Clinical Trial Eligibility Query Generation. Available at SSRN 4637800. 2024.