Innovation in Applied Nursing Informatics
G. Strudwick et al. (Eds.)

doi:10.3233/SHTI240155

Enhancing MAUDE Database Utility by GPT-4 and Cause-Effect Visualization

Yue YU[a], Yuheng SHI[a], Yubo FENG[b], and Yang GONG[a,1]
[a] *University of Texas Health Science Center at Houston*
[b] *Vanderbilt University*

Abstract. The MAUDE database is a valuable public resource for understanding malfunctions and adverse events related to medical devices and health IT. However, its extensive data and complex structure pose challenges. To overcome this, we have developed an automated analytical pipeline using GPT-4, a cutting-edge large language model. This pipeline is intended to efficiently extract, categorize, and visualize safety events with minimal human annotation. In our analysis of 4,459 colonoscopy reports from MAUDE (2011-2021), the events were categorized into operational, human factor, and device-related. Ishikawa diagrams visualized a subset stored in a vector database for easy retrieval and comparison through a similarity search. This innovative approach streamlines access to vital safety insights, reducing the workload on human annotators, and holds promise to enhance the utility of the MAUDE database.

Keywords. Patient safety events, GPT4, Ishikawa diagram, colonoscopy, MAUDE

1. Introduction

The Manufacturer and User Facility Device Experience (MAUDE) database, maintained by the FDA, catalogs safety events related to medical devices and Health IT [1], supporting informed decision-making for healthcare professionals and the public. However, it faces challenges due to its large number of reports, unique data structure and existing search interface. With millions of entries, the task of efficiently extracting and categorizing data is complex [2]. The database currently includes approximately 1,000 product problem categories and over 3,000 product categories, complicating the retrieval of relevant information [3]. Leveraging the utility of the MAUDE database is crucial for understanding Health IT safety, which would enable timely, replicable, and sharable event extraction and categorization [2]. To further enhance usability, organizing safety events into use cases based on users' selection would help visualize and refine search strategies, aggregate similar events in a holistic view. These features are expected to help users find event information quickly and easily and offer a structured overview of events that match their interests.

Ensuring patient safety is of utmost importance during gastrointestinal endoscopy procedures, as it guarantees a high-quality and cost-efficient examination [4]. Thus,

[1] Corresponding Author: Yang Gong, MD, PhD, FIAHSI. 7000 Fannin St, Ste600, Houston, Texas, 77030 USA. Yang.Gong@uth.tmc.edu

colonoscopy events in MAUDE were chosen to demonstrate our insights into the data analysis process and benefits brought up by the proposed user-friendly features.

The use of Large Language Models (LLMs) known as GPT-4 has increased in popularity in the field of natural language processing (NLP) due to their ability to understand context and perform various tasks [4]. We propose to use the text analysis and categorization capabilities of GPT-4 to extract event information from the reports and assign them to predefined categories to reduce human workload [5]. We also use the Ishikawa diagram [6], a classic root-cause representation, to automatically visualize the categorized events and contributing factors into multi-faceted diagrams. Toward a user-friendly interface for the MAUDE database, we propose to design a data analytical pipeline that includes data extraction, categorization, and visualization. Finally, we use a vector database to store the categorized events, enabling a similarity search feature for advanced queries and ensuring the sustainability of the proposed analytical pipeline.

2. Methods

We employed a three-stage review process of the MAUDE database [5] to systematically identify generic, brand, and manufacturing names, apply stratified sampling and analytical merging procedures, which consolidated 112,000 original generic categories into 4,471 refined categories. Further, 4,459 reports under the colonoscopy category (2011-2021) were identified for cluster sampling and specific category refinement.

We utilized GPT-4, an advanced language model by OpenAI, to automate the analysis of the colonoscopy reports. This process involved customizing a prompt defining the tasks, feeding prepared reports into GPT-4 API [7], extracting and categorizing the events into predefined categories and sub-categories. We restricted the output of model as JSON, detailing each event with its assigned category, sub-categories, and associated case numbers of the original reports. We chose JSON for storing the events and their categorizations due to its flexibility and hierarchical data structure, which mirrors the multi-level cause-effect relationship of the events. The JSON data is then saved in a relational database for future retrieval. The JSON data can be directly retrieved and transformed into the Ishikawa diagram for visualization, with each 'bone' representing a pre-defined event category or sub-category identified by GPT-4. The entire visualization process was fully automated using JavaScript for enhancing interactivity and user engagement. This visualization approach is intended to present a clear and structured understanding of the issues in colonoscopy procedures, highlighting potential areas for improvement in a visually intuitive and analytically rich format.

The events, along with their categories, were further transformed into a high-dimensional vector using OpenAI embedding models and stored in a specialized vector database for similarity searches [8]. This vector database enables users to quickly identify and compare similar event instances across different reports, thus facilitating a deeper understanding of recurring issues and patterns in colonoscopy procedures.

3. Results

The cluster sampling on the colonoscopy reports created 112 reports, resulting into three main categories and nine sub-categories based on human factors (reprocessing, handling, and training), device-related issues (manufacturing, maintenance, and failure), and operational period (pre-op, during-op, and post-op). The three distinct categories serve as the foundation for data extraction, categorization, and visualization, as shown in Figure 1 workflow.

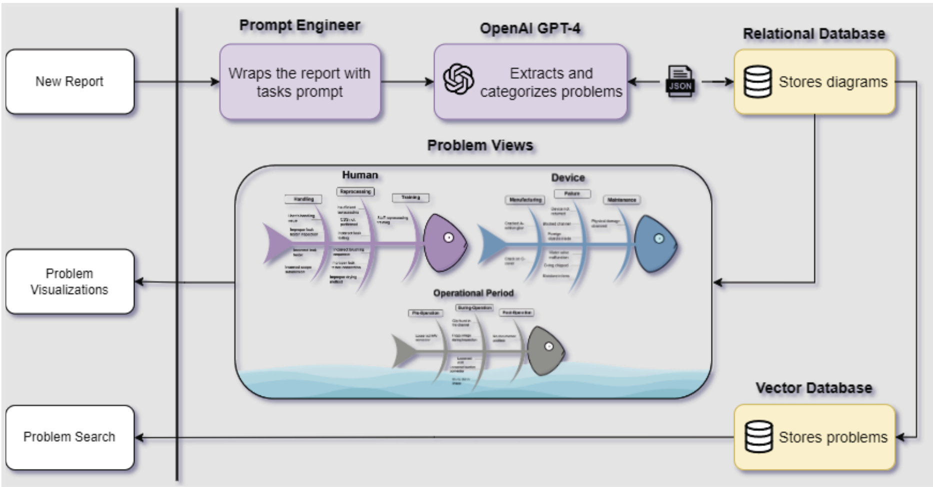

Figure 1. Workflow of analysis pipeline of the MAUDE reports related to colonoscopy events.

Every report that enters the analytical pipeline undergoes processing by GPT-4 modules (indicated by purple blocks). Subsequently, the reports are updated in the databases (indicated by yellow blocks), which store the data permanently for easy visualization (using Ishikawa diagrams) and searching (using the search platform). The modular design of the pipeline enables straightforward customization and upgrading of each step in the analytical procedures. The GPT-4 modules allow testing and defining of the prompts that dictate the analytical tasks that the model performs on the reports. The output of the model, in JSON format, can be further modified and stored in the relational and vector databases for straightforward retrieval.

Diagram data stored in the relational database can be programmatically retrieved and transformed into structured views represented by the Ishikawa diagrams. Notably, some sub-categories have more types of problems than others, such as the “Failure” category under the “Device” view and the “Reprocessing” category under the “Human” view.

Supported by the vector database, the user interface allows free text query (Figure 2). Users can directly search for their queries in the search bar, which will then respond with related issues. Users can select from the listed issue categories and click the case number to reveal the original case report in MAUDE. The free text queries can be stored and accumulated for designing an advanced user interface with more prompts and interaction to help users define and refine their search strategy.

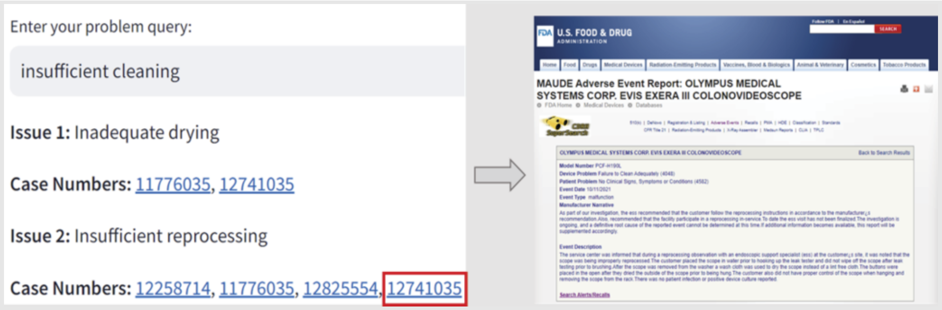

Figure 2. **A** search interface designed for accessing safety events stored in the vector database.

4. Discussion

This study demonstrates a data analytical pipeline enabled by integrating GPT-4 and Ishikawa diagrams to enhance the utility of the MAUDE database. This novel pipeline transforms raw reports into a user-friendly and interactive format, offering intuitive insights into multi-faceted adverse events. Compared to traditional NLP models [1,2], which require extensive labeled data for training and significant maintenance efforts, GPT-4 offers out-of-the-box functionality for the extraction and categorization tasks. Nonetheless, it remains essential to keep the balance between GPT-4's automated analysis and the vital input of human expert in defining initial event categories and validating model outputs, and ensure relevance and accuracy in the specific context, such as colonoscopy events.

This study represents the first time that Ishikawa diagrams have been integrated into an LLM-powered analytical pipeline, despite their traditional use in root-cause analysis visualization. Three 'fishbone' diagrams, which are derived from the categorization structure defined manually, present distinct yet complementary perspectives of the events in a structured format. This helps users to perceive a holistic view of the safety events.

Figure 2 displays the search interface, which aims to present GPT-4 results and assist in improving search strategies for relevant events. This new interface is an addition to the current MAUDE search interfaces that offer basic and advanced search capabilities. The proposed interface shows potential in providing easy access to valuable event reports and making them more user-friendly. Over time, the compiled reports could reveal patterns in specific categories, drawing attention to issues that may contribute to adverse events. The pipeline is designed to allow the continuous integration of new reports into the MAUDE database, enabling the extraction and classification of emerging events, thus keeping the visual diagrams and search interface up to date.

Our demonstration focused solely on colonoscopy cases. However, the analytical pipeline we designed can be used for other event categories identified by experts, indicating its potential usefulness in addressing various safety issues within the MAUDE database. It is important to note that investigating patient safety events should never overlook the issue of underreporting, which can affect both the quantity and quality of reports.

Future developments aim to routinely update event categories, utilizing GPT model analysis and human review, fostering a human-in-the-loop automated pipeline. This adaptability can be further enhanced by insights from user search logs within the interface, informing potential categorization refinements.

5. Conclusions

The study demonstrates our effort to improve the utility of the MAUDE reports by developing an automated data analytics pipeline that combines GPT-4 and Ishikawa diagrams. The process of collecting event data in MAUDE automatically is expected to reduce the need for manual labor in the traditional NLP process, improving its efficiency in tracking new events.

With the help of an intuitive search interface, the pipeline not only extracts and categorizes event data but also provides an interactive platform for developing and refining event search strategies and structured visualization in Ishikawa diagrams. The pipeline can be adapted to other event categories and can offer event updates, identify trends, patterns, or recurrence of events per similarity, making it a valuable tool for a wide range of applications in the MAUDE database.

References

[1] E. Wang, H. Kang, and Y. Gong, Generating a health information technology event database from FDA maude reports, in: Stud Health Technol Inform, IOS Press, 2019: pp. 883–887. doi:10.3233/SHTI190350.

[2] H. Kang, and Y. Gong, Creating a database for health IT events via a hybrid deep learning model, *J Biomed Inform*. **110** (2020). doi:10.1016/j.jbi.2020.103556.

[3] U.S. Food and Drug Administration, MAUDE - Manufacturer and User Facility Device Experience, (2023). https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm.

[4] H. Ji, S. Wang, and Y. Gong, A Descriptive Analysis of Capsule Endoscopy Events in the FDA Manufacturer and User Facility Device Experience (MAUDE) Database, *Journal of Digestive Endoscopy*. **12** (2021) 071–077. doi:10.1055/s-0041-1731960.

[5] OpenAI, GPT-4 Technical Report, (2023). http://arxiv.org/abs/2303.08774.

[6] H. Kang, J. Wang, B. Yao, S. Zhou, and Y. Gong, Toward safer health care: A review strategy of FDA medical device adverse event database to identify and categorize health information technology related events, *JAMIA Open*. **2** (2018) 179–186. doi:10.1093/jamiaopen/ooy042.

[7] G. Ilie, and Carmen Nadia Ciocoiu, Application of Fishbone Diagram to Determine the Risk of an Event with Multiple Causes, *Management Research and Practice*. **2** (2010) 1–20.

[8] B. Greg, W. Peter, M. Mira, and OpenAI, OpenAI:OpenAIAPI, (2020). https://openai.com/blog/openai-api.

[9] Y. Han, C. Liu, and P. Wang, A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge, (2023). http://arxiv.org/abs/2310.11703.