Analyzing Temporal and Associative Relationships of Vaccine Adverse Events

Gunik Luthra, gluthra@stevens.edu **Sneha Dharne, sdharne@stevens.edu**

1 Introduction

The Vaccine Adverse Event Reporting System (VAERS), collaboratively managed by the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA), is one of the most comprehensive repositories for vaccine safety monitoring. Established in 1990, VAERS has served as a passive surveillance system, enabling the identification of potential safety signals from millions of reports on adverse events following vaccination. Its structured fields capture demographic, vaccine, and event details, while unstructured narrative fields often contain critical contextual information, including symptom descriptions and timelines.

While VAERS data is indispensable for vaccine safety research, its utility is hampered by challenges related to the scale, complexity, and heterogeneity of its content. Analyzing unstructured text fields demands advanced computational techniques, as traditional statistical methods often overlook nuanced relationships within narrative data, such as symptom progression and co-occurrence. Moreover, the variability in terminology, medical abbreviations, and reporting styles exacerbates the difficulty of extracting meaningful insights.

Recent Advances in Computational Approaches

Traditional approaches to VAERS data analysis have largely relied on statistical models, such as the Bayesian data mining technique by Evans et al. (2001), which focuses on detecting disproportionality in structured fields. However, these methods often fail to utilize the richness of unstructured text. The rise of Natural Language Processing (NLP) techniques and Large Language Models (LLMs) has transformed the landscape of biomedical text analysis. LLMs, such as BERT and GPT-3, excel in understanding semantic relationships within large-scale text corpora, making them suitable for addressing the challenges posed by VAERS data.

Notably, NLP-based frameworks are capable of performing tasks such as Named Entity Recognition (NER), temporal analysis, and association mining. Despite these advancements, the lack of integration between traditional statistical models and LLMs represents a critical research gap. This project seeks to bridge this gap by leveraging LLMs for symptom extraction and statistical models for analyzing temporal and associative relationships, providing a holistic view of vaccine safety data.

1.1 Literature Review

Analyzing vaccine adverse events has been a critical focus of medical informatics, with significant contributions spanning traditional statistical methods, natural language processing (NLP) advancements, and hybrid approaches combining the two. This section synthesizes relevant literature that informs the current research.

1.1.1 Statistical Approaches

Early methods, such as Proportional Reporting Ratios (PRRs) introduced by Evans et al. (2001), provided a statistical foundation for identifying safety signals from structured data.[3] Harpaz et al. (2012) extended these techniques by incorporating novel data-mining methodologies to improve adverse event detection.[7] However, these models fall short in utilizing unstructured text for richer insights.

1.1.2 Advancements in NLP for Biomedical Applications

Advancements in NLP have revolutionized the processing of biomedical text. Chen et al. (2019) introduced BioSentVec, enabling effective embedding representations for biomedical sentences.[1] Domain-specific pretraining techniques, as demonstrated by Gu et al. (2021), further enhanced NLP performance on medical data.[5] Qi et al. (2020) developed Stanza, a comprehensive toolkit for biomedical NLP, facilitating processing of biomedical narratives.[10] Neumann et al. (2019) and Eyre et al. (2022) introduced SciSpacy and MedSpaCy, respectively, for fast, domain-specific text processing, aiding tasks like Named Entity Recognition (NER).[9][4]

1.1.3 Applications to VAERS

VAERS data analysis has increasingly leveraged NLP. Zhao et al. (2021) used natural language techniques to assess cardiovascular risks linked to COVID-19 vaccines.[13] Guo et al. (2022) employed statistical and ontological approaches to categorize adverse events, emphasizing structured data but highlighting gaps in temporal and associative analyses.[6] Zhang et al. (2019) demonstrated the importance of text mining for symptom extraction but lacked temporal insights.[12]

1.1.4 Hybrid Approaches

Combining NLP and statistical methods bridges the gap between structured and unstructured data analysis. Harpaz et al. (2012) and "Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis" highlighted how integrating Bayesian methods with NLP can enhance signal detection.[7] Zhang et al. (2020) emphasized temporal symptom analysis, advocating for hybrid frameworks to capture progression and associations.[11]

1.1.5 Research Gaps and Opportunities

Despite these advancements, several limitations persist in the existing literature:

- Most studies do not fully utilize the unstructured narrative data in VAERS reports, missing opportunities to analyze temporal progression and co-occurrence of symptoms.
- Few studies integrate LLMs with statistical methodologies, limiting the scalability and robustness of adverse event analyses.
- Temporal analysis remains an underexplored area in vaccine safety research, with current methods failing to capture symptom progression effectively.

This study addresses these gaps by proposing a hybrid model that leverages the strengths of LLMs for unstructured text processing and statistical models for temporal and associative analysis. By integrating these approaches, the study aims to enhance the accuracy and scalability of vaccine safety monitoring.

1.2 Challenges Faced

- **Data Quality and Preprocessing:**The VAERS dataset contains a large volume of unstructured text, making it difficult to extract meaningful information. Many entries include medical jargon, abbreviations, and inconsistent terminology that complicate symptom extraction.Removing duplicate entries proved to be a labor-intensive process. Ensuring that the dataset was clean and representative required careful attention to detail.
- **Named Entity Recognition (NER) Limitations:**The initial NER models (e.g., en_ner_bc5cdr_md,medspacy NER) struggled to capture complex terms and multi-word symptoms, such as "heart attack" or "hospitalized," limiting the comprehensiveness of the extracted data.Many symptom entries contained medical abbreviations that were not universally recognized by the NER models. Establishing a comprehensive dictionary of abbreviations became essential to improve extraction accuracy
- **Integration of Diverse Methodologies:** Designing an effective comparative framework between traditional NLP methods and LLM-based approaches poses challenges in terms of defining evaluation metrics and ensuring a fair comparison.Utilizing LLMs like Ollama for symptom extraction is resource-intensive. The need for efficient computation and the

potential costs associated with processing large datasets raised concerns about feasibility, especially when scaling to more than 10,000 data points.

- **Scalability:** Processing large datasets with LLMs is resource-intensive and requires efficient computational strategies.

1.3 Contributions

This project offers several novel contributions:

- **Proposing a New Framework:** Development of a hybrid model integrating LLMs and statistical methods for temporal and associative analysis of VAERS symptoms.
- **Evaluation of Symptom Extraction Models:** Comparing NER and LLM approaches to assess accuracy, scalability, and computational efficiency.
- **Enhanced Data Processing:** Addressing challenges in handling medical abbreviations and synonyms through preprocessing techniques and customized dictionaries.
- **Comprehensive Analysis:** Bridging gaps in existing research by combining LLMs' strengths in text understanding with statistical models' robustness for association rule mining.
- **Enhanced Data Processing:** Addressing challenges in handling medical abbreviations and synonyms through pre-processing techniques and customized dictionaries.

2 Problem Formulation

The project is structured to compare Named Entity Recognition (NER)-based models combined with temporal and associative relationship analysis against Large Language Model (LLM)-based models integrated with the same analysis techniques. The project is divided into six key tasks as outlined below.

2.1 Task 1:EDA and Data Preprocessing:

The first task involves filtering COVID-19-related data from the VAERS dataset, followed by essential data wrangling and cleaning. This includes expanding abbreviations for better comprehension. The abbreviation dataset used for this task is sourced from a GitHub repository[8]. Once the dataset is preprocessed, exploratory data analysis (EDA) will be conducted to gain a better understanding of the data and generate initial insights.

2.2 Task 2: Applying NER:

This task focuses on extracting symptom-related terms from the free text fields of VAERS reports using NLP-based Named Entity Recognition (NER) methods. The goal is to identify and extract relevant symptoms mentioned in the unstructured text.

2.3 Task 3: LLM extraction:

This task involves extracting symptom-related terms using a large language model (LLM). The challenge here is to engineer the correct prompt to effectively extract symptoms from the free text data, leveraging the power of pre-trained language models like GPT or similar LLMs.

2.4 Task 4: Temporal Relationship Analysis of Symptoms

The objective of this task is to investigate the progression of symptoms over time after vaccination, examining the typical sequence of adverse symptoms.

- Manually annotate the data to establish a baseline for symptom progression. This annotated data will serve as the reference for comparing the performance of the NER-extracted and LLM-extracted symptom sequences.

- Use Kendall's Tau, Longest Common Subsequence (LCS), and Dynamic Time Warping (DTW) to assess the similarity between observed symptom progressions and model predictions.

2.5 Task 5:Associative Relationship Analysis of Symptoms

This task aims to uncover patterns of co-occurrence between adverse events using NLP-based approaches, with comparisons against an LLM model for potential improvements.

Apriori Algorithm: Apriori is a classical algorithm used to identify frequent itemsets in a dataset and derive association rules based on these itemsets. In this context, the "products" refer to symptoms, and the goal is to identify clusters of symptoms that frequently co-occur in VAERS reports.

2.6 Task 6: Evaluation

2.6.1 For Task 4(Temporal Analysis):

The model's ability to predict the temporal order of symptoms will be evaluated using the following metrics:

Kendall's Tau Rank Correlation Coefficient: Measures the correlation between two symptom sequences based on the relative ordering of symptoms. It counts the number of concordant and discordant pairs to evaluate how well the predicted order aligns with the true order.

$$\tau = \frac{C - D}{\sqrt{(C + D + T_1)(C + D + T_2)}}$$

Where:

- C: Number of concordant pairs.
- D: Number of discordant pairs.
- T_1: Number of ties in the first ranking.
- T_2: Number of ties in the second ranking.

Longest Common Subsequence (LCS): Identifies the longest subsequence of symptoms that appears in the same order in both the predicted and true sequences. The longer the LCS, the more similar the sequences are. This is particularly useful for partial matches of symptom progressions.

$$\text{LCS}(i, j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0, \\ \text{LCS}(i-1, j-1) + 1 & \text{if } x_i = y_j, \\ \max(\text{LCS}(i-1, j), \text{LCS}(i, j-1)) & \text{otherwise.} \end{cases}$$

Where:

- i: Index of the first sequence $X = (x_1, x_2, \ldots, x_m)$.
- j: Index of the second sequence $Y = (y_1, y_2, \ldots, y_n)$.

Dynamic Time Warping (DTW):A flexible distance measure used to evaluate the similarity between two sequences that may vary in time or speed. It allows stretching or compressing of sequences to find the optimal alignment, making it ideal for analyzing symptom progression even when symptoms appear at different rates in different cases.

$$\text{DTW}(X, Y) = \min \left(\sum_{i=1}^{m} \sum_{j=1}^{n} d(x_i, y_j) \right)$$

Where:

- $d(x_i, y_j)$: Distance (e.g., Euclidean distance) between the points x_i and y_j.
- $X = (x_1, x_2, \ldots, x_m)$: First sequence.
- $Y = (y_1, y_2, \ldots, y_n)$: Second sequence.

The DTW algorithm finds the optimal alignment path that minimizes the distance.

The task is to calculate these metrics with respect to ground truth for NER generated symptoms and LLM genrated symptoms and compare the results of both.

2.6.2 For Task 5 (Associative Analysis):

The performance of the Apriori Algorithm and the LLM-based pattern discovery will be evaluated using the following metrics:

Support: The probability that a symptom (or set of symptoms) appears in the dataset.

$$\text{Support}(A) = \frac{\text{Number of transactions containing } A}{\text{Total number of transactions}}$$

Where:

- A: The item or itemset.

Confidence: The conditional probability that one symptom appears given the occurrence of another.

$$\text{Confidence}(A \to B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)}$$

Where:

- $A \cup B$: The union of items A and B occurring together.
- Support(A): The support of item A.

Lift: A measure of how much more likely two symptoms are to occur together compared to their independent occurrence.

$$\text{Lift}(A \to B) = \frac{\text{Confidence}(A \to B)}{\text{Support}(B)}$$

Where:

- Confidence$(A \to B)$: The confidence of rule $A \to B$.
- Support(B): The support of item B.

- **Support:** Higher support indicates that an item or itemset frequently appears in the dataset.
- **Confidence:** Higher confidence means a stronger conditional relationship between items A and B.
- **Lift:** A lift value greater than 1 indicates a positive association, while a value less than 1 suggests a negative association.

The task is to use the symptom data from VAERS data as ground truth and compare these metrics for NER extracted symptoms and LLM extracted symptoms.

3 Methods

In this project, we aim to develop a hybrid model that combines a Large Language Model (LLM), specifically Gemini 1.5 Flash, with statistical methods for analyzing vaccine adverse events. To establish the problem, we first perform Exploratory Data Analysis (EDA), Temporal Relationshhip analysis and Associative Analysis on the data extracted using NER (NLP-based) methods. We then compare the methodologies to evaluate their effectiveness.

3.1 Symptom Extraction and Preprocessing

The process of symptom extraction and preprocessing involved several systematic steps to ensure the data was clean and ready for analysis. The following subsections detail the methodologies employed:

3.1.1 Data Acquisition and Exploration

- **Dataset:** The primary dataset used for this project is the Vaccine Adverse Event Reporting System (VAERS) dataset. Initial exploration involved assessing the dataset's structure, understanding the types of adverse event reports, and identifying the narrative text that contains symptom descriptions.
- **Vaccine Symptoms Data:** The focus was primarily on extracting symptoms related to the COVID-19 vaccine. This involved filtering the dataset for relevant vaccine entries and organizing the data for further analysis.

3.1.2 Data Preprocessing

- Filtering COVID-19-specific data from the VAERS dataset.
- **Deduplication:** Duplicate entries within the VAERS dataset were identified and removed to ensure a unique set of reports. This step was crucial for maintaining the integrity of the analysis.
- **Text Normalization:** The text data underwent normalization processes, including:
  - **Lowercasing:** All text was converted to lowercase to ensure uniformity and prevent case-sensitive mismatches.
  - **Tokenization:** The narrative text was tokenized to split it into individual words or tokens for analysis.
  - **Stopword Removal:** Common stopwords (e.g., "the," "and," "is") were removed from the dataset to focus on more informative words.

3.1.3 Handling Medical Abbreviations

- **Medical Abbreviation Dictionary:** A comprehensive dictionary of medical abbreviations was established to enhance the extraction process. This dictionary helped standardize terms and improve the recognition of symptoms during analysis.
- **Text Processing:** Special attention was given to free-text entries that contained medical abbreviations, ensuring that these terms were properly recognized and extracted.

3.2 Named Entity Recognition (NER)

- **NER Model Implementation:** Two different NER models were employed to extract symptom tokens from the narrative text:
  - **SpaCy[9][4] NER Model:**
    - **Model**: `en_ner_bc5cdr_md`[9][4] was utilized to extract symptom tokens such as "fatigue," "pain," and "fever."
    - **Limitations:** This model struggled with complex symptom phrases, such as "heart attack" or "death," indicating a need for improved extraction methods.
  - **Medical Abbreviations Enhanced NER**
    - **Model:** `en_core_sci_md` was used in conjunction with the medical abbreviations dataset. This approach aimed to capture a broader range of medical terms, including patient histories and therapies.
    - **Results:** This model successfully extracted both medical terms and other relevant information; however, it also retrieved non-symptom terms such as "day," "year," "dose," and "consumer."

  Challenges: In some cases, the output included detailed descriptions such as "brain was in a fog," "thinking was slowed," and "flu symptoms (for 24hrs)," indicating a need to refine the prompt or adjust the model settings for better specificity in symptom extraction.

3.3 LLM Extraction

- Methods
  - Using pre-trained LLMs like Gemini 1.5 flash to extract symptoms from narrative text.
  - Designing optimized prompts to accurately extract symptom-related terms.
- Reason for Selection
  - LLMs exhibit strong capabilities in understanding unstructured text and extracting semantic information.
  - Prompt engineering enables customizable and adaptable symptom extraction.
- Challenges
  - LLMs are computationally intensive and limited by token size.
  - Solution: Implemented batch processing and token optimization strategies to reduce resource usage.

3.4 Temporal Relationship Analysis of Symptoms

- Methods
  - Manually annotating data to establish a baseline for symptom progression.
  - Calculating temporal similarity metrics: Kendall's Tau, Longest Common Subsequence (LCS), and Dynamic Time Warping (DTW) for NER extracted symptoms and LLM extracted symptoms with respect to baseline.
- Reason for Selection
  - These metrics effectively capture the order, sequence similarity, and temporal alignment of symptoms.

3.5 Associative Relationship Analysis of Symptoms

- Methods
  - using the pre-exisitng VAERS symptoms data as baseline.
  - Applying the Apriori Algorithm to identify frequent symptom co-occurrences and derive association rules for both NER extracted symptoms and LLM extracted symptoms.
- Reason for Selection
  - The Apriori Algorithm is a well-established method for association rule mining and is effective in identifying clusters of co-occurring symptoms.

3.6 Evaluation

- Evaluating temporal analysis with Kendall's Tau, LCS, and DTW.
- Evaluating associative analysis with support, confidence, and lift metrics.
- Compared the NER and LLM models on temporal and associative relationship analysis using baselines.

3.7 Tools used

- **NER Models:** SciSpacy, MedSpaCy.
- **LLMs:** gemini 1.5 flash.

3.8 Innovative Approaches

- Integration of custom abbreviation dictionaries to handle domain-specific medical terminology.
- Hybrid framework combining statistical methods with LLM-based extractions to leverage the strengths of both approaches.
- Prompt engineering techniques tailored for extracting biomedical information with LLMs.

4 Dataset and Experiments

4.1 Dataset

The primary dataset for this project will be the Vaccine Adverse Event Reporting System (VAERS) dataset, co-managed by the Centers for Disease Control and Prevention (CDC) and the US Food and Drug Administration (FDA). VAERS is a national early warning system that collects reports of adverse events following vaccinations to identify potential safety issues in US-licensed vaccines. The VAERS dataset consists of three key components:

- **VAERS Data:** Contains detailed reports of adverse events, including narrative descriptions and demographics of patients.
- **VAERS Symptoms:** Lists documented symptoms associated with the reported adverse events.
- **VAERS Vaccine:** Provides information on the administered vaccines, allowing for a focused analysis on specific vaccines (e.g., COVID-19).
- **Abbreviation Data:**[8] A github repositary containing 15 csv files for abbreviations and its meaning was combined to create a abbreviation dataset to map the abbreviations to its full form in VAERS_DATA for better symptoms extraction.

This dataset was used for both the temporal and associative analysis of vaccine-related adverse events.

4.2 Symptoms Extraction From NER and EDA

- Initially, symptoms were extracted by applying Named Entity Recognition (NER) techniques(en_ner_bc5cdr_md model), which involved removing stop words and ignoring abbreviations. To gain insight into the extracted symptoms, exploratory data analysis (EDA) was performed.

Figure 1: Word Cloud

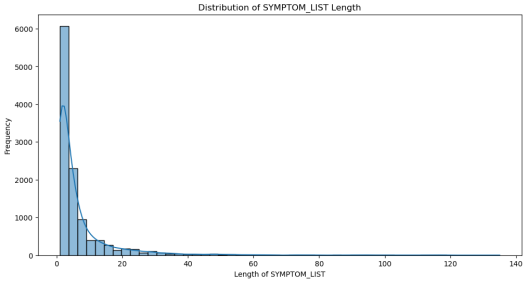

Figure 2: Number of Symptoms per data point

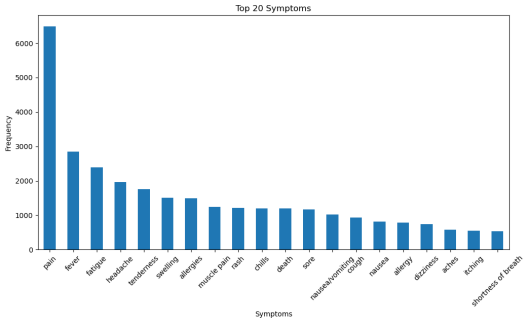

Figure 3: Distribution of Top 20 Symptoms

- Symptoms were extracted using NER techniques(en_core_sci_md) with removing STOP WORDS and replacing abbreviations(HR ,N/V, BP, etc) with their meanings. To gain insights EDA was performed.

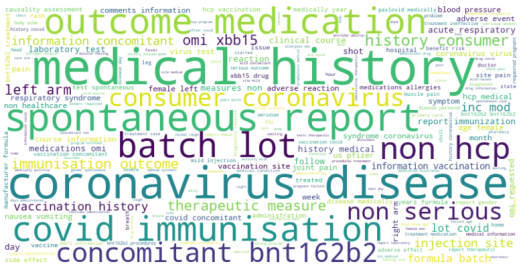

Figure 4: Word Cloud

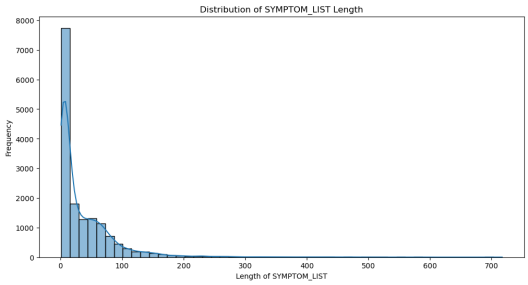

Figure 5: Number of Symptoms per data point

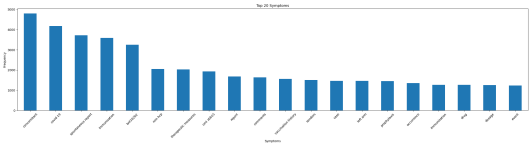

Figure 6: Distribution of Top 20 Symptoms

- Symptoms were extracted using NER Techniques(en_ner_bc5cdr_md model) without removing STOP WORDS and replacing abbreviations(HR ,N/V, BP, etc) with their meanings. To gain insights EDA was performed.

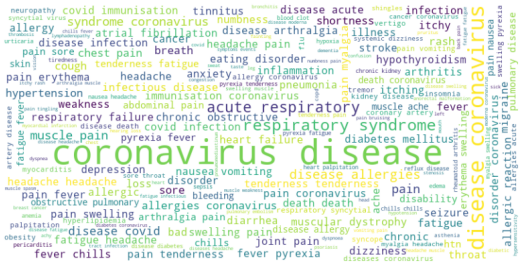

Figure 7: Word Cloud

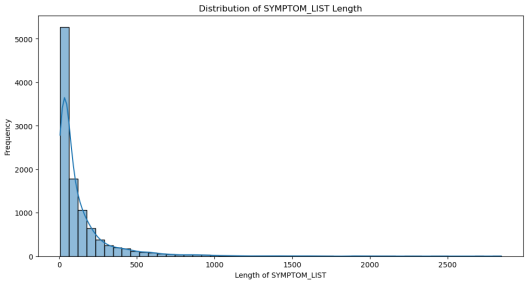

Figure 8: Number of Symptoms per data point

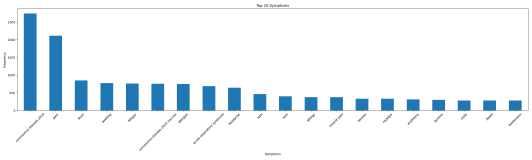

Figure 9: Distribution of Top 20 Symptoms

- As we can see, the last approach yielded the best results, producing a clearer word cloud and a more accurate distribution of symptoms. It is evident that this method focused primarily on symptom extraction, minimizing noise and leading to superior outcomes when utilizing the Apriori algorithm and its associated rules.

4.3 Temporal Relationship Analysis

For the Temporal Relationship Analysis, we selected 200 random data points from the custom covid_symptoms dataset. These data points were manually annotated to serve as the baseline for comparison. Using this baseline, we calculated various evaluation metrics to assess the performance of both the NER model and the LLM-based model, highlighting their comparative strengths and weaknesses.

4.3.1 Kendall's Tau Rank Correlation Coefficient

For Kendall's Tau which ranges from -1 to 1, we computed the aggregated values across the 200 data points as follows:

- NER Model: 0.1290
- LLM Model: 0.3393

The results from the Kendall's Tau Rank Correlation Coefficient indicate that the LLM-based model (0.3393) outperforms the NER model (0.1290) in capturing the temporal relationships within the covid_symptoms dataset. The higher Kendall's Tau value for the LLM model suggests a stronger

alignment with the manually annotated baseline, implying that the LLM model is more effective in modeling the temporal dependencies between symptoms.

4.3.2 Longest Common Subsequence (LCS)

The Longest Common Subsequence (LCS) was computed for the 200 data points to evaluate the alignment of the predicted temporal relationships with the manually annotated baseline.

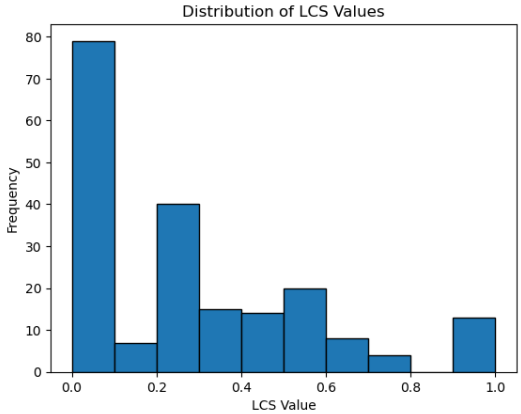

Figure 10: Longest Common Subsequence Distribution of NER model

- The LCS values have a bimodal distribution, with a large peak around 0.0 and a smaller peak around 0.8.
- This suggests the NER model performs well for some inputs, capturing a long common subsequence, but struggles with other inputs.

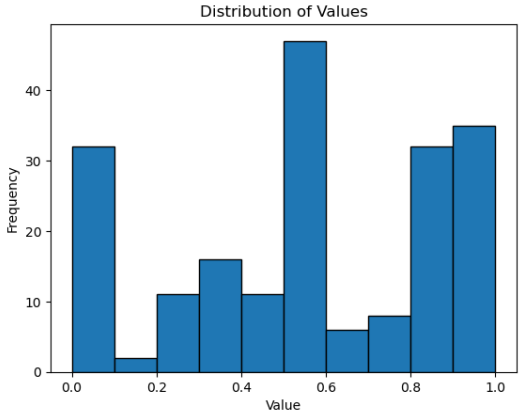

Figure 11: Longest Common Subsequence Distribution of LLM model

- The value distribution is more evenly spread out, with a large peak around 0.6 and smaller peaks at other values.
- This indicates the LLM model has a more consistent performance across the range of inputs, without as many extreme outliers as the NER model.

4.3.3 Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) was computed to measure the alignment of the predicted temporal sequences with the manually annotated baseline, considering the optimal temporal alignment.

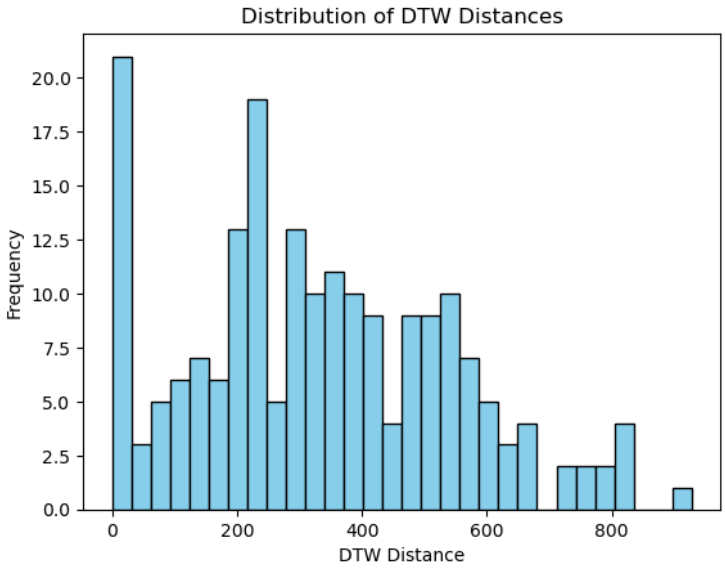

Figure 12: Dynamic Time Warping for NER Model

We realized it is distance of words so we thought rather than finding distances between number, find distances between word embeddings for better result. Used Word2Vec.

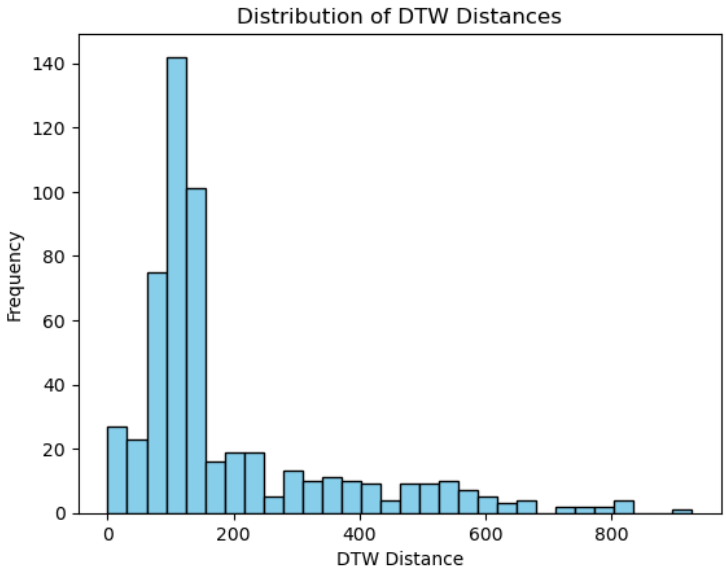

Figure 13: Dynamic Time Warping for NER Model after word2vec

- The distribution of DTW distances exhibits a clear peak around 200, indicating a significant number of samples with that DTW distance.

- The distribution also shows a wide range of DTW distances, extending from 0 to over 800, suggesting the NER model's performance varies considerably across different inputs.

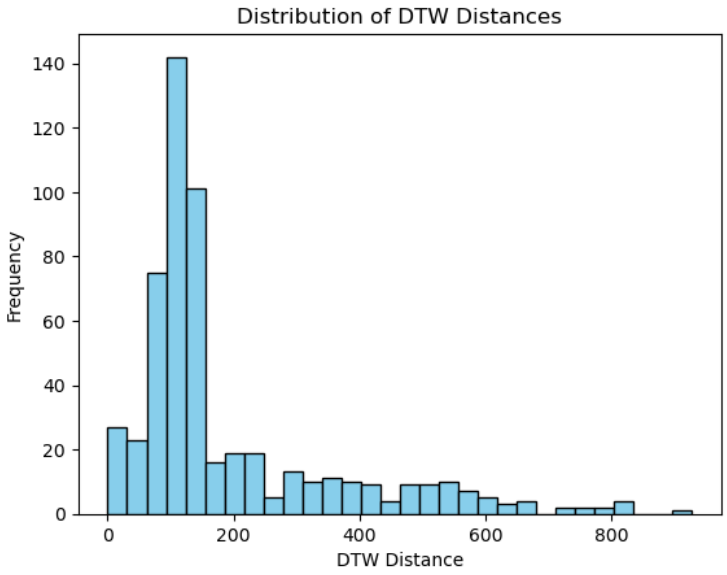

Figure 14: Dynamic Time Warping for LLM Model after word2vec

- The distribution of DTW distances for the LLM model is more evenly spread out, with a peak around 200 and smaller peaks at other distances.

- The range of DTW distances is still relatively wide, but not as extreme as the NER model, indicating more consistent performance across the inputs.

The differences in the DTW distance distributions suggest that the LLM model may have more stable and robust performance compared to the NER model when it comes to the word2vec technique. The NER model appears to struggle with certain inputs, resulting in a wider range of DTW distances, while the LLM model demonstrates a more balanced distribution, potentially indicating better overall performance.

4.4 Associative Relationship Analysis

To analyze associative relationships, we extracted the most frequent itemsets and association rules across 15,000+ data points. The analysis compared symptoms identified through Named Entity Recognition (NER) and Large Language Models (LLMs) with symptoms annotated in the VAERSSYMPTOMS dataset, which serves as a benchmark curated by medical professionals.

4.4.1 VAERSSYMPTOMS Dataset:

This dataset, annotated by medical experts with a fixed number of symptoms per data point, allowed us to establish **13 solid association rules**. These rules provided reliable insights, given the consistency and precision of the annotations.

	antecedents	consequents	support	confidence	lift
7	(Injection site erythema)	(Injection site swelling)	0.014462	0.516854	23.286246
6	(Injection site swelling)	(Injection site erythema)	0.014462	0.651558	23.286246
8	(Underdose)	(Product administered to patient of inappropri...	0.010501	0.609489	19.271002
9	(Headache, Malaise)	(Fatigue)	0.010123	0.797030	11.277545
10	(Fatigue, Malaise)	(Headache)	0.010123	0.624031	11.176339
2	(Body temperature)	(Fatigue)	0.011255	0.617241	8.733636
1	(Myalgia)	(Fatigue)	0.015279	0.532895	7.540176
0	(Malaise)	(Fatigue)	0.016222	0.522267	7.389802
4	(Vaccination failure)	(COVID-19)	0.049296	0.988651	6.040530
12	(Drug ineffective, SARS-CoV-2 test)	(COVID-19)	0.019995	0.984520	6.015293
11	(Vaccination failure, SARS-CoV-2 test)	(COVID-19)	0.019429	0.984076	6.012582
5	(Drug ineffective)	(COVID-19)	0.067027	0.939207	5.738436
3	(SARS-CoV-2 test)	(COVID-19)	0.045775	0.888889	5.430998

Figure 15: Association Rules VAERSSymptoms

4.4.2 NER-Based Annotation:

- Symptoms annotated using the NER model generated a significantly larger dictionary of terms due to the lack of nuanced semantic understanding and the inability to consolidate medical synonyms effectively.

- As a result, **647 association rules** with significant lift were established, many of which included redundant associations like "fever" and "pyrexia," despite their semantic equivalence.

- The large volume of rules, while statistically significant, reduced the practical utility of the findings, making it challenging to draw actionable insights from the symptom co-occurrence data.

	antecedents	consequents	support	confidence	lift
470	(pain, pyrexia)	(fever, erythema)	0.010312	0.689076	61.567746
473	(fever, erythema)	(pain, pyrexia)	0.010312	0.921348	61.567746
281	(pain, pyrexia)	(fever, muscle pain)	0.010123	0.676471	55.172247
283	(fever, muscle pain)	(pain, pyrexia)	0.010123	0.825641	55.172247
635	(pain, myalgia, fatigue)	(erythema, muscle pain)	0.010501	0.625468	51.809613
639	(erythema, muscle pain)	(pain, myalgia, fatigue)	0.010501	0.869792	51.809613
622	(swelling, myalgia)	(pain, muscle pain, fatigue)	0.011004	0.911458	51.770833
605	(pain, muscle pain, fatigue)	(swelling, myalgia)	0.011004	0.625000	51.770833
628	(pain, muscle pain, fatigue)	(erythema, myalgia)	0.010501	0.596429	51.552174
645	(erythema, myalgia)	(pain, muscle pain, fatigue)	0.010501	0.907609	51.552174
418	(swelling, myalgia)	(muscle pain, fatigue)	0.011129	0.921875	50.731834
423	(muscle pain, fatigue)	(swelling, myalgia)	0.011129	0.612457	50.731834
615	(muscle pain, fatigue)	(pain, swelling, myalgia)	0.011004	0.605536	50.686578
614	(pain, swelling, myalgia)	(muscle pain, fatigue)	0.011004	0.921053	50.686578
612	(pain, myalgia, fatigue)	(swelling, muscle pain)	0.011004	0.655431	50.601796
617	(swelling, muscle pain)	(pain, myalgia, fatigue)	0.011004	0.849515	50.601796
433	(muscle pain, fatigue)	(erythema, myalgia)	0.010563	0.581315	50.245825
428	(erythema, myalgia)	(muscle pain, fatigue)	0.010563	0.913043	50.245825
637	(pain, erythema, myalgia)	(muscle pain, fatigue)	0.010501	0.912568	50.219676
638	(muscle pain, fatigue)	(pain, erythema, myalgia)	0.010501	0.577855	50.219676

Figure 16: Association Rules NER Based Annotation

	antecedents	consequents	support	confidence	lift
61	(injection site swelling)	(injection site redness, injection site pain)	0.011469	0.654237	47.660195
58	(injection site redness, injection site pain)	(injection site swelling)	0.011469	0.835498	47.660195
60	(injection site redness)	(injection site pain, injection site swelling)	0.011469	0.696751	46.712845
59	(injection site pain, injection site swelling)	(injection site redness)	0.011469	0.768924	46.712845
20	(injection site redness)	(injection site swelling)	0.012836	0.779783	44.482017
21	(injection site swelling)	(injection site redness)	0.012836	0.732203	44.482017
54	(injection site pain, fatigue)	(injection site tenderness)	0.012123	0.625767	38.857583
56	(injection site tenderness)	(injection site pain, fatigue)	0.012123	0.752768	38.857583
52	(injection site redness)	(injection site pain, fatigue)	0.010459	0.635379	32.798033
51	(injection site pain, fatigue)	(injection site redness)	0.010459	0.539877	32.798033
48	(injection site swelling)	(injection site pain, fatigue)	0.011112	0.633898	32.721597
46	(injection site pain, fatigue)	(injection site swelling)	0.011112	0.573620	32.721597
31	(injection site pain, malaise)	(muscle pain)	0.010340	0.731092	29.716965
47	(injection site swelling, fatigue)	(injection site pain)	0.011112	0.963918	28.308559
50	(injection site redness, fatigue)	(injection site pain)	0.010459	0.956522	28.091357
55	(injection site tenderness, fatigue)	(injection site pain)	0.012123	0.948837	27.865676
30	(muscle pain)	(injection site pain, fatigue)	0.013073	0.531401	27.430722
28	(injection site pain, fatigue)	(muscle pain)	0.013073	0.674847	27.430722
4	(injection site tenderness)	(muscle pain)	0.010696	0.664207	26.998235
33	(malaise, muscle pain)	(injection site pain)	0.010340	0.915789	26.895123

Figure 18: Association Rules Gemini LLM Based Annotation

	antecedents	consequents	support	confidence	lift
642	(myalgia, fatigue)	(pain, erythema, muscle pain)	0.010501	0.596429	49.924211
631	(pain, erythema, muscle pain)	(myalgia, fatigue)	0.010501	0.878947	49.924211
589	(pain, myalgia, fatigue)	(muscle pain, headache)	0.011004	0.655431	49.875455
594	(muscle pain, headache)	(pain, myalgia, fatigue)	0.011004	0.837321	49.875455
544	(pain, myalgia, fatigue)	(fever, muscle pain)	0.010249	0.610487	49.790685
548	(fever, muscle pain)	(pain, myalgia, fatigue)	0.010249	0.835897	49.790685
430	(fatigue, myalgia)	(erythema, muscle pain)	0.010563	0.600000	49.700000
431	(erythema, muscle pain)	(fatigue, myalgia)	0.010563	0.875000	49.700000
608	(pain, swelling, muscle pain)	(myalgia, fatigue)	0.011004	0.870647	49.452736
620	(myalgia, fatigue)	(pain, swelling, muscle pain)	0.011004	0.625000	49.452736
547	(muscle pain, fatigue)	(pain, fever, myalgia)	0.010249	0.564014	49.286133
546	(pain, fever, myalgia)	(muscle pain, fatigue)	0.010249	0.895604	49.286133
585	(pain, muscle pain, headache)	(myalgia, fatigue)	0.011004	0.866337	49.207921
597	(myalgia, fatigue)	(pain, muscle pain, headache)	0.011004	0.625000	49.207921
540	(pain, fever, muscle pain)	(myalgia, fatigue)	0.010249	0.862434	48.986243
551	(myalgia, fatigue)	(pain, fever, muscle pain)	0.010249	0.582143	48.986243
421	(swelling, muscle pain)	(fatigue, myalgia)	0.011129	0.859223	48.803883
420	(fatigue, myalgia)	(swelling, muscle pain)	0.011129	0.632143	48.803883
553	(fever, myalgia)	(pain, muscle pain, fatigue)	0.010249	0.853403	48.473298
537	(pain, muscle pain, fatigue)	(fever, myalgia)	0.010249	0.582143	48.473298

Figure 17: Association Rules NER Based Annotation

4.4.3 Gemini Model-Based Annotation:

- The Gemini LLM overcame many limitations of the NER model by handling medical synonyms, annotating abbreviations without manual preprocessing, and maintaining a consolidated dictionary.
- Despite requiring minimal preprocessing (only table merging), the Gemini model yielded **62 association rules** with significant lift, striking a balance between semantic accuracy and analytical usability.
- These results demonstrated greater alignment with the VAERSSYMPTOMS dataset while maintaining efficiency and scalability.

The analysis highlights the limitations of rule-based approaches like NER when applied to nuanced medical datasets and emphasizes the advantages of using advanced LLMs like Gemini. While the VAERSSYMPTOMS dataset remains the gold standard for reliability, the Gemini model provides a scalable and effective alternative for annotating symptoms with minimal preprocessing and high semantic accuracy.

	antecedents	consequents	support	confidence	lift
19	(injection site tenderness)	(injection site pain)	0.014678	0.911439	26.767360
57	(injection site redness, injection site swelling)	(injection site pain)	0.011469	0.893519	26.241064
25	(malaise, fatigue)	(muscle pain)	0.010459	0.637681	25.920045
17	(injection site swelling)	(injection site pain)	0.014916	0.850847	24.987890
18	(injection site redness)	(injection site pain)	0.013727	0.833935	24.491202
29	(muscle pain, fatigue)	(injection site pain)	0.013073	0.794224	23.324954
32	(injection site pain, muscle pain)	(malaise)	0.010340	0.696000	22.875563
44	(malaise, fatigue)	(injection site pain)	0.012657	0.771739	22.664618
43	(injection site pain, fatigue)	(malaise)	0.012657	0.653374	21.474573
26	(muscle pain, fatigue)	(malaise)	0.010459	0.635379	20.883123
2	(joint pain)	(muscle pain)	0.011944	0.505025	20.527930
16	(redness)	(swelling)	0.015391	0.609412	19.168563
3	(muscle pain)	(injection site pain)	0.014856	0.603865	17.734443
15	(vomiting)	(nausea)	0.017174	0.621505	14.983800
24	(malaise, muscle pain)	(fatigue)	0.010459	0.926316	12.149682
42	(injection site pain, malaise)	(fatigue)	0.012657	0.894958	11.738389
27	(injection site pain, muscle pain)	(fatigue)	0.013073	0.880000	11.542198
40	(injection site pain, headache)	(fatigue)	0.011766	0.860870	11.291281
53	(injection site pain, injection site tenderness)	(fatigue)	0.012123	0.825911	10.832758
36	(injection site pain, fever)	(fatigue)	0.010043	0.820388	10.760324

Figure 19: Association Rules Gemini LLM Based Annotation

5 Conclusion

The Vaccine Adverse Event Reporting System (VAERS) contains valuable yet complex datasets, often requiring advanced methodologies to extract meaningful insights from unstructured text.The comprehensive analysis conducted on the VAERS dataset has provided valuable insights into the effectiveness of different techniques for extracting, modeling, and analyzing symptom-related data. The study highlights the advantages of using advanced language models, such as the Gemini LLM, over traditional rule-based approaches like Named Entity Recognition (NER). The LLM-based methods demonstrated superior performance in capturing temporal relationships, as evidenced by the higher Kendall's Tau correlation and more consistent Longest Common Subsequence distributions. Furthermore, the Gemini LLM approach generated a more semantically accurate and consolidated set of association rules, offering greater alignment with the benchmark VAERSSYMPTOMS dataset while maintaining efficiency and scalability. These findings underscore the potential for LLM-powered techniques to provide more insightful and actionable analysis in the context of vaccine-related adverse event monitoring and research, ultimately contributing to enhanced patient safety and better-informed decision-making.The code and implementation details for this analysis can be found at `https://github.com/SnehaDharne/TemporalAndAssociativeRelationships-VAERS`[2]

6 Project Management

This project will be carried out by two members:

- **Gunik Luthra:**
  - Data merging from all sources
  - Extraction of data using two NER models with abbreviation dictionary
  - EDA on the same models
  - Manual annotation for temporal analysis
  - Evaluation using Temporal Relationship Analysis
  - Documentation and citations
- **Sneha Dharne:**
  - Extraction of data using two NER models without abbreviation dictionary
  - EDA on extracted symptoms from all models
  - Ollama implementation for data extraction (not scalable or satisfactory)

- Gemini implementation for data extraction
- Evaluation using Associative Relationship Analysis
- Maintaining code on git repository

References

[1] Qingyu Chen, Yifan Peng, and Zhiyong Lu. Biosentvec: creating sentence embeddings for biomedical texts. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–5. IEEE, 2019.

[2] Sneha Dharne. Temporal and associative relationships in vaers, 2024.

[3] Stephen JW Evans, Patrick C Waller, and S Davis. Use of proportional reporting ratios (prrs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10(6):483–486, 2001.

[4] Hannah Eyre, Alec B Chapman, Kelly S Peterson, Jianlin Shi, Patrick R Alba, Makoto M Jones, Tamara L Box, Scott L DuVall, and Olga V Patterson. Launching into clinical space with medspacy: A new clinical text processing toolkit in python. In *AMIA Annual Symposium Proceedings*, volume 2021, page 438, 2022.

[5] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(1):1–23, 2021.

[6] Wenxin Guo, Jessica Deguise, Yujia Tian, Philip Chi-En Huang, Rohit Goru, Qiuyue Yang, Suyuan Peng, Luxia Zhang, Lili Zhao, Jiangan Xie, et al. Profiling covid-19 vaccine adverse events by statistical and ontological analysis of vaers case reports. *Frontiers in Pharmacology*, 13:870599, 2022.

[7] Rave Harpaz, William DuMouchel, Nigam H Shah, David Madigan, Patrick Ryan, and Carol Friedman. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology & Therapeutics*, 91(6):1010–1021, 2012.

[8] Imants M. Medical abbreviations, 2024. Accessed: 2024-11-04.

[9] Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. Scispacy: Fast and robust models for biomedical natural language processing. *arXiv preprint arXiv:1902.07669*, 2019.

[10] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*, 2020.

[11] Mingyu Zhang, Hua Chen, and Jian Xu. Temporal and associative analysis in medical datasets: Insights from adverse event reports. In *Proceedings of the IEEE International Conference on Health Informatics*, pages 211–220. IEEE, 2020.

[12] Wei Zhang, Ling Chen, and Xiao Li. Text mining for symptom identification in vaers reports. In *Proceedings of the IEEE International Conference on Biomedical and Health Informatics (BHI)*, pages 342–348. IEEE, 2019.

[13] Yiqing Zhao, Michael G Ison, and Yuan Luo. Covid vaccine and cardiovascular risks: A natural language analysis of vaccine adverse event reports. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 614–618. IEEE, 2021.