MUFASA – AI Tool to Transform Medical Information Data into Actionable Medical Affairs Intelligence

Karen Ka Yan Ng, PharmD, University of Toronto
LEO Pharma Medical Affairs and Education Residency Project

Keywords: Artificial Intelligence, Machine Learning, Medical Affairs Intelligence and Data Strategies, Semantic Analysis, Medical Information, Data Visualization

Abstract:

The concepts of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) have effectively made their way into pharmaceutical companies leading to many interesting discussions and promises for application in healthcare research and drug developments. Already, AI tools such as chatbots, virtual digital assistants, and research tools are being used by pharmaceutical companies. There is still opportunity to leverage AI, ML, and NLP technology in Medical Affairs activities for optimal sharing of medical insights.

The following Pharmacy Residency Project: **MI** Data **U**ses **F**or **A**I **S**emantic **A**nalysis (MUFASA), is a python coded tool designed with state-of-the-art Sentence Transformer, clustering, and visualization techniques in AI and Big Data Analysis. The tool has demonstrated promising ability to harness AI technology in semantic understanding of sentences to improve efficiency, identify emerging topics and provide better visualization infographics aimed to effectively transform medical information data into actionable Medical Affairs intelligence.

1. Introduction

1.1 The Role of Medical Information Department

One of the responsibilities of the medical information (MI) team is to handle medical inquiries for LEO Pharma Canada's marketed products in a timely manner. These inquiries can come from various Healthcare Professionals (HCPs), patients, and healthcare organizations. The inquiries include:

- Product information for specific population
- Research publications as well as other information requests following Other Learning Activities (OLA) unaccredited HCP facilitated programs/ Continuing Medical Education (CME) accredited programs
- Sales Rep/MSL contacts
- Pharmacovigilance (PV)/Adverse Events (AE) cases
- Product Complaints (PC)
- Patient Support Program (PSP) inquiries

The responsibilities for day-to-day case handling and case documentation uniquely situates the MI team as the only department able to directly collect unsolicited customer feedback and to keep the responses in a validated medical inquiry database.

1.2 The Pandemic has Expedited the Growing Expectations for the MI Department

Since the start of the Covid-19 pandemic, when remote engagements became the norm, pharmaceutical companies have found an urgent need to reinvent themselves for the digitalized new world. According to IQVIA's Channel Preferences Survey 2020 (Figure 1) [1], HCPs indicated their preference for less face-to-face interaction and increased preference for online resources. Most recently, Accenture as well as IQVIA surveys have found that HCPs have adapted and will continue to prefer the virtual interface even when the pandemic ends (Figure 2) [2].

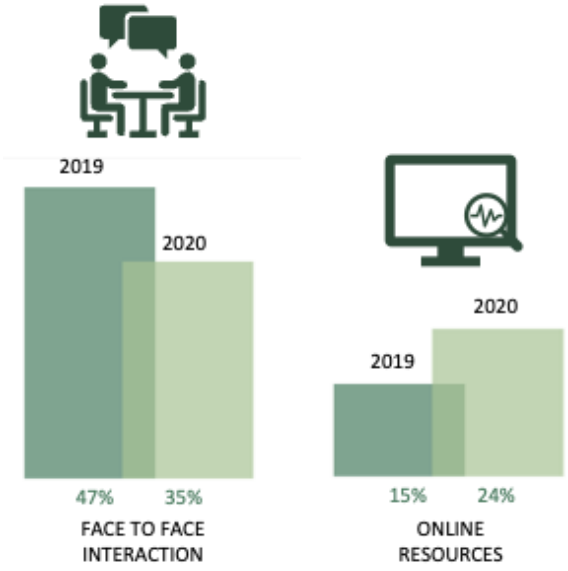

Figure 1: IQVIA's Channel Preference Survey 2020

Naturally, HCPs now have higher expectations to receive medical information with more 'on-demand' services. Faced with time pressure, HCPs lean to ask questions and share insights via their mobile devices, expect rapid response tools to provide better clinician support and patient support programs, and seek more value from meetings, regardless of which part of the organization they are dealing with.

Figure 2: Virtual Interactions are here to stay post pandemic Source: Accenture (left), IQVIA (right).

According to McKinsey's HCP survey in 2018 [3], 81 percent of physicians are dissatisfied with their interactions with biopharmaceutical companies, and over 40 percent no longer perceive a "need" for medical support from pharma. Outlined below for dermatologists, the primary dissatisfaction is a perceived lack of personalized, relevant content (28 percent) and appropriate communication channels (17 percent) (Figure 3).

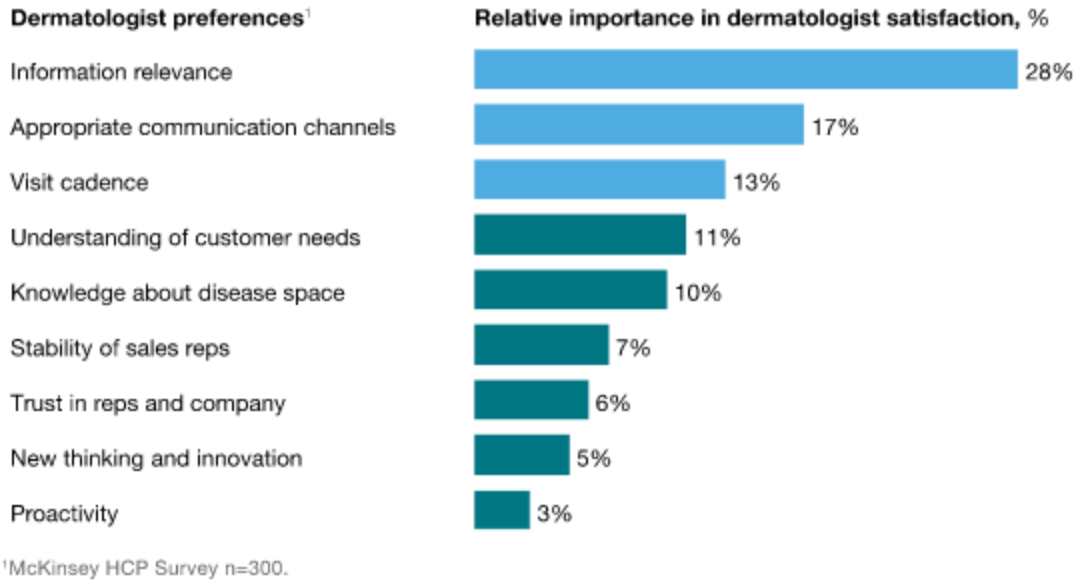

Figure 3: Personalized, relevant content and appropriate communications channels are priorities for physicians. Source: McKinsey HCP Survey

Provided these findings and consistencies, pharmaceutical companies will need to better understand what type of questions or information the HCPs require. Medical Affairs

department should therefore leverage the growing amounts of information contained in Medical Information databases to develop insights that can deliver targeted content at scale and engage meaningfully.

1.3 Invaluable Unsolicited Data Within the MI Department

Much like when a company runs an advisory board to collect insights, the MI department can capture a variety of opinions through the unsolicited inquiries received. The benefit of the data being unsolicited is that such data is free from sampling and selection bias. Previously, sending surveys for needs assessment required Medical Affairs personnel to design the questions and required physicians to fill out the survey. Physicians' responses are limited to what the company personnel thinks physicians want to learn, and what the doctor feels is most appropriate to ask rather than what they ask in practice.

Through the benefit of understanding insights from unsolicited inquiries from the MI database, field members have guided topics to ask more direct questions pertaining to topics that may not typically emerge from regular surveys. MI data will lead to better actions that addresses what we are hearing from HCPs and asking questions such as: Is there an issue with new drugs? Do we need new data? Is there a more recent study the Medical Affairs team should be aware of?"

Ultimately, the wealth of these various direct points of contact can be utilized to find areas to improve case handling efficiency for MI and MSLs, identify product improvement areas, improve customer satisfaction, and monitor Medical Affairs initiatives that can lead to better programs for HCPs and patients. All these topics will result in better strength, weaknesses, opportunities, and threats (SWOT) analysis on how the team will perform as it improves the ability to find root causes of issues and key drivers of performances.

1.4 Creating Effective Data Strategies through harnessing AI's Semantic Understanding of Texts

In the same manner as drugs have a shelf-life, data too has a shelf life. Data needs to be accessible and transparent within an organization and should be directing field activities in real-time. The acceleration of improvement in data collection, data visualization, and analytical modeling for large data sets is recently referred to as Big Data technology. It has been made possible thanks to the improvement of AI and machine learning techniques which are known for their ability for mining and interpretation of the vast amount of information in an efficient manner.

The AI tool MUFASA has been developed in an attempt to apply Big Data technology to allow LEO Pharma Canada to take insights from fields and the MI database, and efficiently transform them into data actionable strategies. Concretely, MUFASA extracts topics from massive number of unsolicited inquiries in the form of digital texts such as emails and scripts obtained from phone calls. The following diagrams illustrates this process.

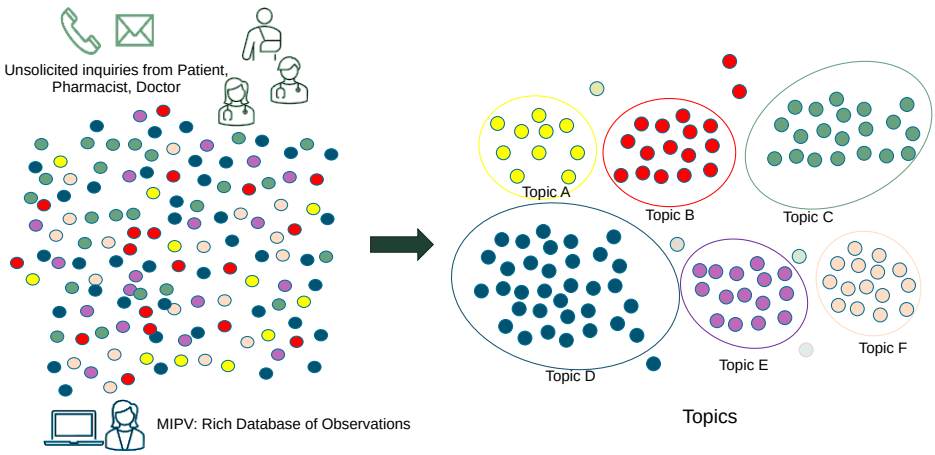

Figure 4: Transform Observations to Topics for Effective Data Strategies.

Traditionally, these inquiry texts are stored in a database maintained by the MI department along with metadata such as time and inquirer information. While such regular data can be easily retrieved and analyzed with clear numerical values and labels, there has been no effective labeling method for the abstract meaning of the questions. Therefore, it was impossible to efficiently analyze large amounts of text data for the inquiries that customers really need and are interested in, which are important factors to consider when making business decisions.

However, recent developments in AI technology in the field of natural language processing have made it possible to quantify the semantic meaning of sentences using an AI model called Sentence Transformer. The MUFASA tool uses that technology to analyze topics and enable fast searches for similar topics.

1.5 Sentence Transformer and Its Two Applications

Sentence Transformer is an AI model that converts words or sentences to vectors in high-dimensional space called embedding space. As Sentence Transformer maps the words or sentences into the vectors as points in the embedding space as shown in Figure 5, it can represent the semantic relationship between words through the distance and angle between vectors.

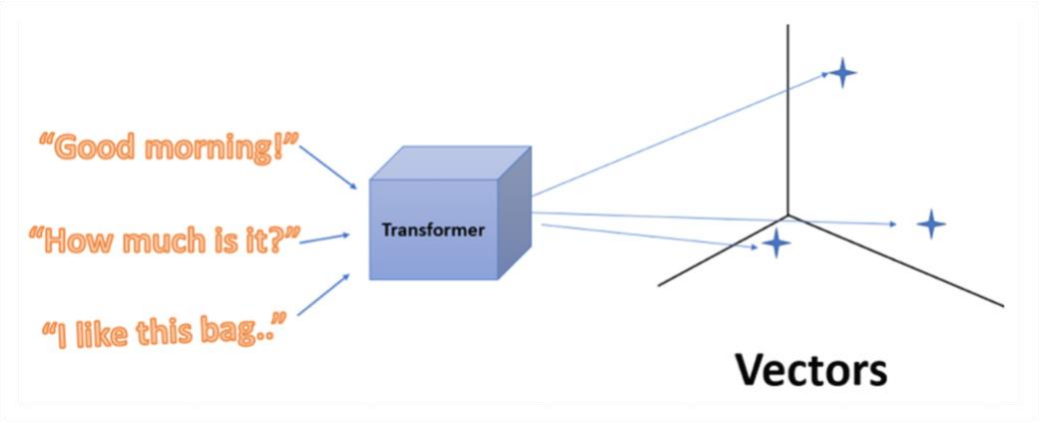

Figure 5: Semantic sentences are put into transformers and plot into high-dimensional space.

In the example of Figure 6, the relationship between the words that describe two genders (man, woman) is located at a certain distance and angles in the embedding space. If we convert the words King and Queen into the same space, they will have the similar distances and angles because King and Queen hold semantically relevant concepts, gender.

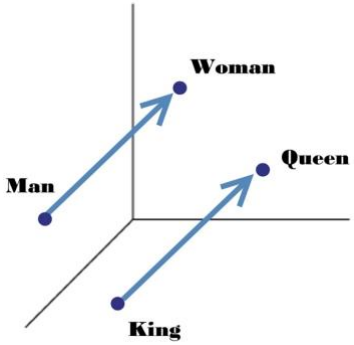

Figure 6: Spatial relationships of the vectors represent semantic relationships of the original sentences.

When all sentences of existing inquiries in the database are transformed to the embedding space as vectors, it can be applied to two actions.

One is to efficiently find the semantically similar sentences to a sentence of interest, which we call a query sentence. It can be achieved by searching other vectors near the vector of the query sentence. Then, by converting those vectors back to the original sentences, we can identify semantically similar sentences. We call this process semantic search, and it will be applied to improve efficiency of MI and MSL case handling.

The second is cluster analysis, which allows us to analyze what kind of topics are present in thousands of sentences in the database. When thousands of sentences are plotted on embedding space as vectors, sentences with similar meanings are naturally placed close together, and form clusters (Figure 7). If the sentences in a cluster are on the same topic, albeit with some minor differences in meaning, then one cluster represents a single topic.

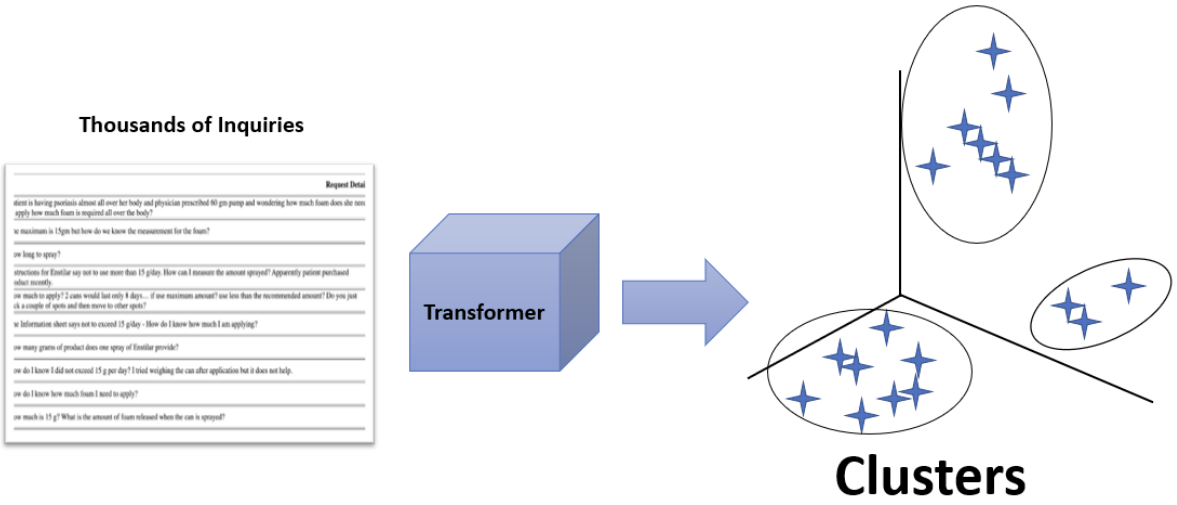

Figure 7: Thousands of inquiries are put through the transformer which then forms clusters of where semantically similar questions are forming clusters.

1.6 Project MUFASA's Objective

Utilizing AI's ability in Big Data cluster analysis with Sentence Transformer, the development and integration of MUFASA aims to achieve two objectives:

1. Improve efficiency for MI and MSL case handling.
2. Deepen insights and strengthen engagement for Medical Affairs.

The first objective is mainly achieved by semantic search, while the second objective can be accomplished by both semantic search and cluster analysis.

3. Materials and Methods

The following section will explain the flow of MUFASA to achieve the two objectives. It will also contain explanation of terminology as well as experiments for testing the validity of the tool. The process flow of MUFASA that is outlined in the Figure 8 consists of four major components: data preparation, text to vectors conversion, semantic search (steps 3-4), and clustering (steps 5-6).

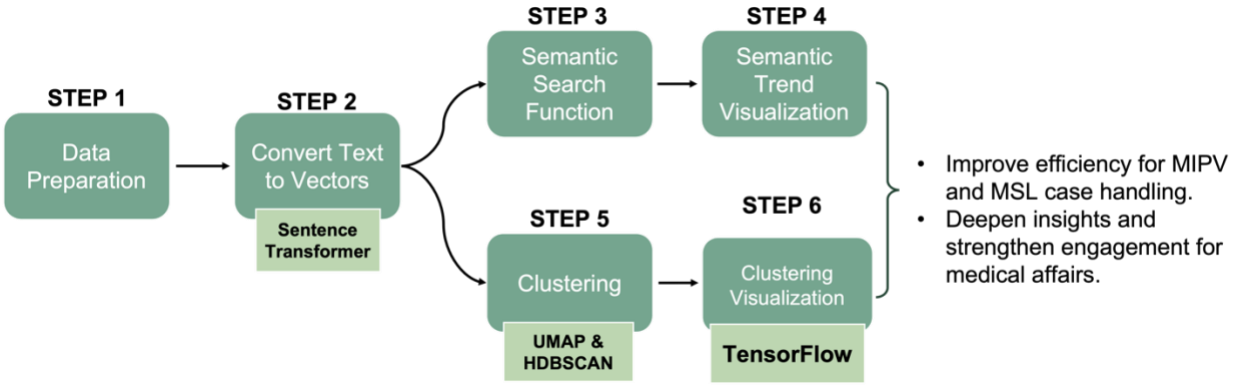

Figure 8: Process flow of MUFASA

Step 1: Data Preparation

The databases owned by the MI department have records of various inquiries to the department for the past 10 years. The inquiries in the database cover all current LEO Pharma marketed products in the year 2022. Each data point includes a case ID, date received, inquiry text, inquirer's contact information, and other various information. All data are exported as a single CSV file which contains 28 columns and 13578 rows as of February 26th, 2022. The file is then processed on the Jupyter Notebook, which is a platform to execute a Python program MUFASA. A Python programming language has a widely used library called Pandas that handles CSV or any form of major table data. The CSV file is loaded by Pandas and then converted into a Pandas DataFrame. Afterward, various data cleaning processes are performed such as renaming columns for simpler names, handling missing or invalid data, and bundling different values that specify the identical entity such as "DrugX" and "DrugX®" into "DrugX".

Step 2: Convert Texts into Vectors

To enable both semantic sentence search and cluster analysis, all sentences from the inquiry database need to be converted to vectors in a high-dimensional vector space by a

Sentence Transformer model. Among multiple Sentence Transformer models that are pretrained and publicly available, the state-of-the-art model *all-mpnet-base-v2* is selected due to its highest performance [4]. All inquiry texts are converted to 768-dimensional vectors by the model. If an inquiry is in French and the English translation is available, the English text will be selected in the conversion process.

Step 3: Semantic Search Function Tool

After the conversions, every vector is located at the embedding space according to its semantic meaning. Therefore, the vectors located nearby have semantic similarities to the original sentences. When the user inputs a search query, the Sentence Transformer will convert that query sentence into a vector and identify other vectors located nearby using k-nearest neighbors' algorithm. The search engine then returns the user-specified number of closest vectors which will be displayed along with the similarity index.

Step 4: Semantic Trend Visualization

MUFASA can visualize the semantically close sentences found by Semantic Search in list view (Figure 10) or chronological line plot (Figure 11). In the list view, the N most semantically similar sentences are displayed with meta information such as Similarity Index and Case ID. On the other hand, the line plot shows what types of contact (HCPs, Specialists, Patients, Pharmacists) and in chronological time order the top N inquires was received.

Step 5: Dimensional Reduction and Clustering

To surface the trends among data points, a large dimensional reduction from 768 to 3 is performed by Uniform Manifold Approximation and Projection (UMAP) algorithm so that the data points can be visualized in 3-dimensional space. Among many dimensional reduction algorithms such as PCA or t-SNE, UMAP has stronger ability to handle massive data size and high dimensionality in fast and scalable way [5].

Given enough vector points are plotted on the 3-dimensional space, a concentrated group of points is referred as a cluster. The attempt of grouping all data points into several clusters is called clustering. The clustering of the MI dataset is performed by HDBSCAN algorithm [6]. Compared to other clustering algorithms such as K-mean, or agglomerative clustering, HDBSCAN is selected as it is density based. This means it can handle clusters with arbitrary shapes and different sizes and densities, which is expected in the datapoints with dimensions reduced from high dimensions such as our dataset. Since defining the group of data as a cluster is somewhat subjective, some parameters needed to be manually selected such as the minimum number of data points to be considered as one cluster. After performing clustering, a new column will be added to the CSV file which will indicate the cluster number beside each case inquiry.

Step 6: Visualization of Clusters

Since the data is in the 3-dimensional space, it is possible to visualize data points with the addition of an assigned color corresponding to their assigned cluster in the previous step. A machine learning library TensorFlow [7] developed by Google provides a visualization tool TensorBoard to view such clusters, which MUFASA uses for the visualization of 3D plotted data.

Figure 9 is a 3D visualization of clusters for SprayX inquiries. Each dot represents a vector converted from an inquiry text. When a user hovers over a dot, the inquiry and metadata can be read.

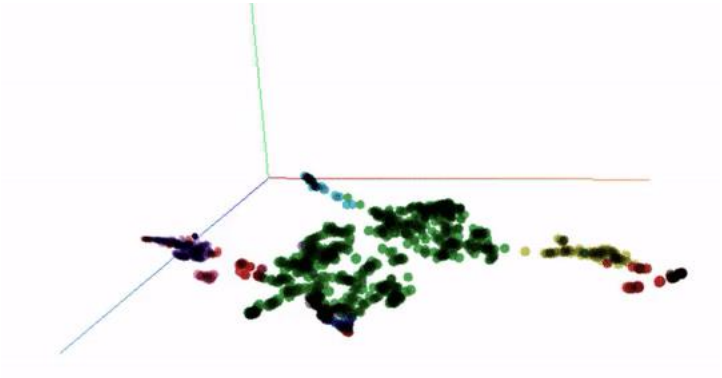

Figure 9: Semantic sentences are plotted in 3-dimensional space.

4. Results

4.1 Improve Efficiency for MI and MSL Case Handling – Semantic Search

Figure 10 is a screenshot of MUFASA semantic search interface in list view executed on Jupyter Notebook. The interface has provided the inquiry followed by the list of semantically similar inquiries that include case ID, date received, inquiry sentence, similarity score, and response summary. It has an ability to filter the results by information in the original data such as time, product of interest, contact type, etc. The user can determine the number of top similar cases that they wish to be returned.

Inquiry on Spray X:

How do I measure Spray X?

Top 5 most similar inquiries in database:

CANA-009180 - 2021-06-09 00:00:00:

Instructions for Spray X say not to use more than 15 g/day. How can I measure the amount sprayed? Apparently patient purchased product recently.

(Score: 0.5364)

Response Summary:

Response Note: 10-Jun-2021: Provided information as per PM. Spraying Spray X for about 2 seconds is enough to cover an area of skin that is about the size of an adult palm. A 2-sec spray dispenses approx 0.5 g of foam.

Reassured patient that we know from our clinical trials that the average amount patients use is 30 g/wk.

CANA-003805 - 2017-03-20 00:00:00:

How can I measure a dose of 15 g?

(Score: 0.5164)

Response Summary:

Response Note: nan

CANA-007324 - 2019-09-25 00:00:00:

How does the patient tell what 5 g is?

(Score: 0.4760)

Figure 10: MUFASA Semantic Search Interface - List View.

For each returned case, MUFASA will provide a similarity index for each sentence. This similarity index will be a value between 0 to 1 (1 being identical). The user will also be able to have an option to adjust the threshold for similarity. However, it is important to note this similarity is not exhaustive and comparable between different types of queries. Additionally, MUFASA is typo tolerant and can handle multiple queries as a batch. Table 1 below shows the selected example of query sentences and top 2 similar inquiries from the past.

Query: "How do I measure SprayX?"
Result 1: *Instruction for SprayX say not to use more than 15g/day. How can I measure the amount with this type of formulation? Apparently patient purchased product recently.*
Result 2: *How can I measure a dose of 15g?*
Query: "Stability of repackaged OintmentY"
Result 1: *How long is Ointment Y good for? Meaning shelf life time and time once opened?*
Result 2: *What is the shelf life of Ointment Y once it has been opened?*
Query: "Big Bumps"
Result 1: *PSP reported patient experienced swelling in the legs*
Result 2: *PV Case report from *** - redness around his injection sites - loks like a wasp bite.*

Table 1: Selected examples of semantic search queries and results of their top 2 similar sentences.

4.2 Deepen Insights and Strengthen Engagement for Medical Affairs – Semantic Search

Figure 11 displays visualization from the semantic search result in line chart view which can be generated seamlessly with the list view. The user manually sets the lower limit of the similarity index along with the inquiry. As a result, MUFASA retrieves all sentences with similarities that exceed the lower limit of similarity index. While the list view allows the user to browse through actual inquiry texts, the line plot view allows the user to visualize how many questions related to a topic were received every month. The line plot function makes it possible to perform various analysis such as the ability to monitor the impact of specific business decisions. In the example below, the inquiry was set to "*Stability of repackaged OintmentY.*"

Analysis:

A total of 24 inquiries which are semantically similar in asking about OintmentY stability after it is opened was identified by MUFASA. Most inquiries are coming from patients (orange) and pharmacists (red). The frequency of inquiries increased in 2019. The timing of increase in frequency of inquiry coming in coincides with the discontinuation of 30g OintmentY tube in mid 2018. Pharmacists appears to be working around the issue of the lack of the 30g tube by repacking the 60g product in other containers but are concerned about the stability of the product.

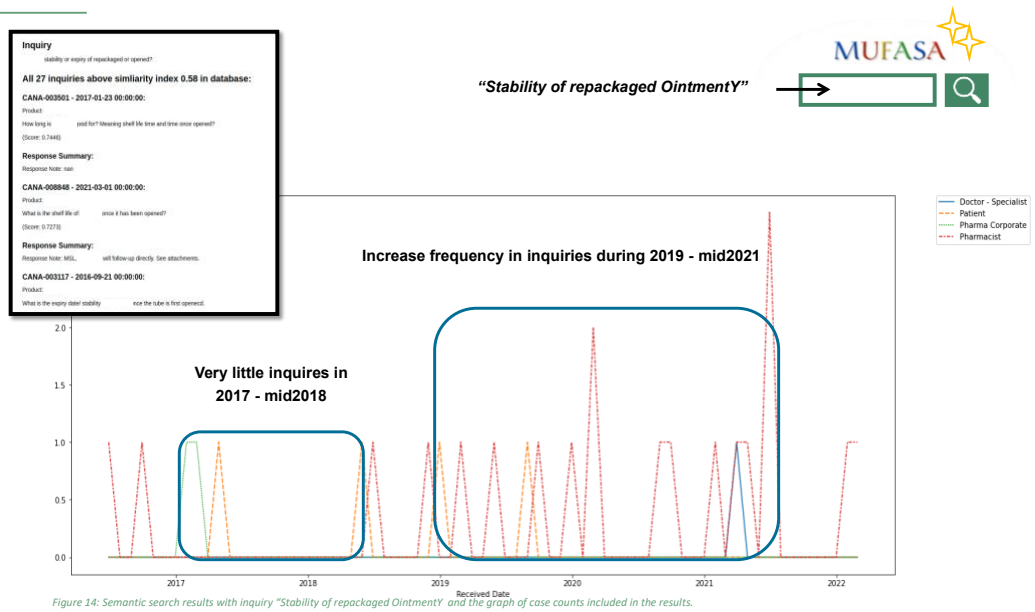

Figure 14: Semantic search results with inquiry "Stability of repackaged OintmentY" and the graph of case counts included in the results.

Figure 11: MUFASA Semantic Search Interface - Line Plot View

4.3 Deepen Insights and Strengthen Engagement for Medical Affairs – Cluster Analysis

To better understand a large set of inquiry data, their themes or major topics need to be identified. HDBSCAN algorithm determined the following number of clusters.

- SprayX Clusters: 9
- OintmentX Clusters: 18
- OintmentY Clusters: 33

Figures 12 is a diagram explaining the concept of assigning numbers to the nine SprayX clusters. Datapoints that are not forming any cluster are categorized as cluster -1, and clusters with missing inquiry texts are assigned to the cluster 0 and 1. Rather than identifying themes through navigating the TensorFlow view (Figure 9), each cluster was converted to list view for easier identification of themes and topics. The final full list of clusters for SprayX, OintmentX, and OintmentY with their themes and respective possible explorations for business actions are provided in Tables 1, 2, and 3 of Appendices.

Figure 12: Assigning numbers to the SprayX clusters to identify themes and potential explorations.

4.4 Validating Sentence Transformer's Understanding Ability

Before progressing the development of the MUFASA tool further, it is important to validate if the Sentence Transformer can semantically map the inquiries which uses medical terminology. Two types of analysis were performed. Firstly, the qualitative analysis (Figure 13), and secondly the comparison between AI clustering and clustering through the manual categorization by MI member during case handling (Request Category examples: Dose,

Pregnancy, Stability, Off-label, Adverse Event) (Figure 14).

Figure 13: Clusters formed by inquiries from different type of contact: Doctors, Patients, and Pharmacists.

In Figure 13, SprayX inquiries are visualized in 3D spaces. Using the visualization tool, inquiries from doctors, pharmacists, and patients are circled in blue. The results show that the inquiries from each type of contact tend to aggregate and form clusters separately. Given each type of contact has its own distinctive interest and language use, it validates the feasibility and accuracy of the Sentence Transformer's mapping and UMAP reduction process to organize the semantic meaning of the inquiries.

Next, by comparing the AI-determined clusters (Figure 14-left column-AI determined cluster) with the manually determined classification (Figure 14-right column: MI affiliate determined 'Request Category'), one can observe the similarity in how the colors are separated. Note that because colors assigned to each cluster are randomly determined in each clustering attempt, the comparisons are made based on how the clusters are separated. The results indicate that the AI can semantically categorize the inquiry to a similar level of human interpretation.

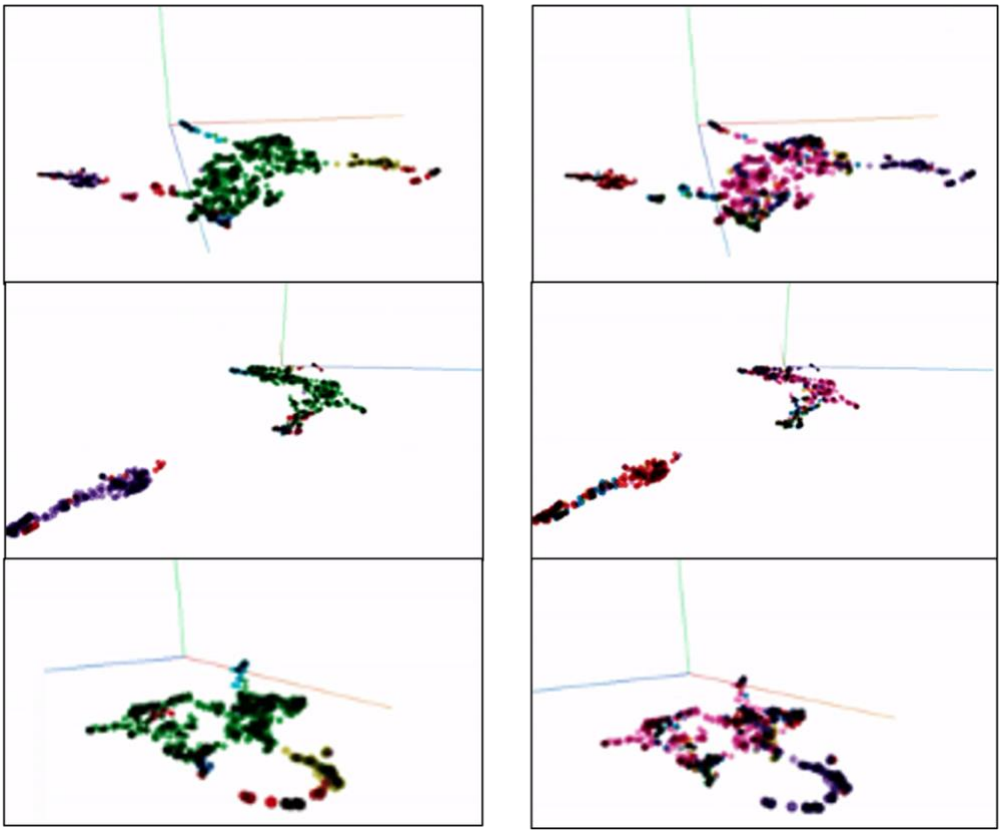

Figure 14: Clusters identified and coloured by HDBSCAN (left) is matching the manually labeled colour (right). This implies that the AI has ability to categorize/assign the clusters as accurate as manual decision.

5. Discussion

5.1 Improve Efficiency for MI and MSL Case Handling – Semantic Search

MUFASA has shown potential to be used by the MI/MSL team in daily use by finding similar cases using semantic search ability. As such, this allows for increase accuracy in finding similar cases even if the description initially provided were vague (Table 1 in section 4.1).

Through daily use, it helps train new MI team members on handling medical inquiries through quickly identifying responses from previous similar cases as new trainers will be able to see how past cases were handled. If a new hire takes twenty minutes per case to communicate with another employee to ask about the past case handling, and there are three cases to ask per day on average, the new hire will save five hours per week by using this semantic search. Moreover, it improves coordination and reduces redundancy in generating new replies as more suitable templates can be prepared that allows new employees to maintain consistency for answers, which also means a reduction in compliance risk.

Overall, MUFASA holds promises to help reduce the workload burden of responding to common inquiries. This will allow the MI team to commit their knowledge to true customer-centricity in handling more specialized inquiries which can meet the changing stakeholder demands and expectations. Rather than allocating resources to answer repetitive questions,

MUFASA allows the medical information team to have more time to focus on understanding the data which in turn would generate actionable insights by allowing the team to proactively offer support and training to other departments when needed.

5.2 Deepen Insights and Strengthen Engagement for Medical Affairs – Semantic Search

MUFASA's semantic search and visualization of known trends allows the team to share a better understanding of the known situations which is useful tool when preparing reports for cross-functional meetings.

With MUFASA, users will be the ones to determine the questions of interest. Rather than relying on predetermined KPIs, which lacks context, insights and trends are now identified through a more natural decision-making process, where unique and appropriate questions that are specific to the situations can be explored. Through the iteration process of asking more questions to MUFASA's search functions, users will be able to better understand the situation and validate speculative hypotheses for any trends identified.

Semantic search with a line plot view also provides great value. Firstly, it can provide monitoring of the volume of workload of certain topics such as the number of inquiries received pertaining to the measurement of foam formulation. Secondly, it can examine markets such as monitoring the number of requests for comparison data with competitors, identifying HCP speciality interests, and tracking new off label uses. Thirdly, it can monitor the success of recent medical education initiatives. For example, the number of research paper requests after a learning activity may serve as a signal for the level of impact the event has made. Lastly it can monitor the impact of business decisions. Figure 11 demonstrates how the effect of discontinuation of a smaller pack size of a product may lead to changed behaviour of pharmacists. Understanding such trends can inform field team members to ask more appropriate questions during their HCP field visits to understand their needs and priorities.

5.3 Deepen Insights and Strengthen Engagement for Medical Affairs – Cluster Analysis

A pharmaceutical company's engagement with HCPs can be further deepened by utilizing MUFASA clustering analysis' ability in identifying topics HCPs would like to learn about.

For the MI department, identifying the most frequently asked questions also allows the team to curate more targeted FAQs sheets for various HCPs, launch a more organized Medical Information website accordingly, and identify questions that can be easily outsourced.

For other Medical Affairs departments, MUFASA's cluster analysis may allow for improved medical education programs. MUFASA can identify new topics for Other Learning Activities (OLAs) and acts as a means of evidence for the justification to support OLAs. Providing justification for OLAs is now required based on recent revisions made to Innovative Medicines Canada (IMC) Code of Ethical Practices. The IMC Code changes came into effect in April of 2022.

Lastly, MUFASA's cluster analysis was also able to identify commercial excellence opportunities such as improving customer satisfaction of the drug through identifying product feedback clusters. Monitoring inquiries pertaining to product inquires can evaluate if there needs to be additional changes to publicly available product information in commercial materials, patient support materials, and Product Monograph (PM).

It is important to note that the purpose of MUFASA's cluster analysis function is not to determine if a problem is considered "significant" through the number of received inquiries but

rather, be used as a tool to find signals that may call for further action. MUFASA can identify problems that may have been unknowingly missed as well as to monitor known issues. As such, there is irrelevance for this paper to analyze if the clusters contain a statistically relevant "n" power of inquiries to be statistically significant.

5.4 Limitation and Future of MUFASA

As a machine learning tool, MUFASA's ability is fundamentally limited by the data available. On a top of that, semantic search results are somewhat affected by the wording and phrasing, and those effects on the results are not clear at this moment. Similarly, what is considered clinically relevant may not be able to be processed by the Sentence Transformer in identifying relevant "similar cases". Medical terminology and dictions are complex and nuanced, but Sentence Transformer models are trained based on usual language setting and not trained for medical situations. Therefore, while there may be an interest in training them from MI case handling to better understand how cases are categorized, there is argument to not train the machine learning model for specific identification of clusters. This is because there is benefit in the lack of training as a lack of training removes the bias for which an individual thinks what a cluster should be.

It is also important to mention that the semantic search does not factor in sentiment analysis. Continual development of MUFASA should explore the integration of AI's ability for sentiment analysis as it is also important for a pharmaceutical company to analyze their social share of voice by sentiment and topic at the same time. While a product may have a high social share of voice (SOV), it is critical for pharmaceutical companies to want to address any negative comments.

On the cluster analysis front, it was observed that one cluster may contain multiple themes. However, the purpose of clustering is to have the AI help the identify themes that may not be easily evident. While clusters are not fully isolated through the clustering process, separating the information from thousands of inquiries to smaller organized clusters of hundred inquiries allows for an easier identification of topics. The purpose of MUFASA ultimately made the process more efficient and effective.

MUFASA is still in its infancy in development and lacks integration with other data ecosystems at LEO Pharma. For example, the addition of geographical data can add tremendous potential to MUFASA. Geographical analysis will provide LEO Pharma a way to measure how broad the reach of a company's scientific message is in comparison to its competitors. When geographical locations are added to other demographic filters, such as gender, age, or occupation to identify important target audiences, this could lead to unique advantages against LEO Pharma's competitors. The future of MUFASA should also explore the addition of other ecosystems that will allow the tool to evaluate large volumes of publications, clinical trials, and text insights from advisory board all in which can help quickly identify, discern any new key topics of interests for HCPs.

6. Conclusion

MUFASA leverages Sentence Transformer and clustering, which are the latest techniques in AI and Big Data technologies. With the growing trend for digitization and

expectations expedited by the Covid-19 pandemic, MUFASA has demonstrated potential to not only improve the efficiency for MI and MSL case handling but also help guide field teams to efficiently generate actionable medical insights and more personalized HCP contents. MUFASA has demonstrated how it can be used to understand the issues and questions that HCPs raise most frequently. The feedback can then be used in the deployment of teams and content creation. Teams will gain the ability to regularly fill knowledge gaps, leading to more productive engagements and better accountability to established processes.

Overall, MUFASA unlocks the understanding of data to the next level between LEO Pharma Canada and other LEO Pharma affiliates. Information has been the key to a better organization and new developments. The more information LEO Pharma Canada has, the more optimally LEO Pharma can organize itself to deliver the best outcomes for patients.

Funding Statement and Acknowledgements

MUFASA project was completed during the residency timeframe of the Industrial Residency program hosted by University of Toronto and sponsored by LEO Pharma Canada between September 2021 till October 2022. The development of this tool, completion of this paper, and many opportunities to present it to the LEO Pharma global team and two international Medical Affairs conferences, “AI in Medical Affairs” and “Medical Affairs Xcellence”, would not have been possible without the exceptional support of my supervisors Laurel Holden and Sabrina Spina, faculty advisor Heather Kertland, colleague Samar Shuhaiber and direct manager Yohan D’Souza. The author is also grateful for the insightful comments offered by the many cross-functional managers at LEO Pharma Canada on the possibilities for application of this tool by other departments. The author would like to thank Izuki Matsuba, Machine Learning Research Engineer at Matterport for the time provided in consultation of latest AI available technology that leads to the development of the tool MUFASA. The generosity and expertise by many of these professionals have improved this project in innumerable ways.

References

[1] Surprise pandemic lessons for Pharma. (2021, September). *IQVIA*. https://www.iqvia.com/-/media/iqvia/pdfs/library/publications/surprise-pandemic-lessons-for-pharma.pdf

[2] Is COVID-19 altering how pharma engages with HCPs? (2020, August). *Accenture*. https://www.accenture.com/ca-en/insights/life-sciences/coronavirus-changing-pharma-hcp-engagement

[3] Medical Affairs: Key imperatives for engaging and educating physicians in a digital world. (2018, May). *McKinsey*. https://www.mckinsey.com/industries/life-sciences/our-

insights/medical-affairs-key-imperatives-for-engaging-and-educating-physicians-in-a-digital-world

[4] Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

[5] McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

[6] McInnes, L., & Healy, J. (2017, November). Accelerated hierarchical density-based clustering. In 2017 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 33-42). IEEE.

[7] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

8. Appendices

8.1 SprayX Cluster Themes

SprayX Cluster Themes		
Cluster	**Themes**	**Potential Explorations**
Cluster -1	Not enough cases to be classified as a cluster	N/A
Cluster 0	Question was missing	N/A
Cluster 1	Question was missing	N/A
Cluster 2	Safety on plane	Med info-Website Potential, Educate Pharmacists
	Allowed on plane/Dangerous goods	
Cluster 3	Measurement of foam (Changes have been added to PM)	# Inquiry Vs Time to see if # of questions decreased (Are the PM changes sufficient)
	Amount/Actuation	
	Concern to not exceed 15gm	
	How much to apply?	
	Confusion on instructions	
Cluster 4	How to apply to Scalp?	Add content to Med Info Website (therein forth described as "Med Info Website")
	Drug holidays	Med Info Website
	Re-starting SprayX	Med Info Website
	Long term efficacy and maintenance	Med Info Website (SprayX-Data) #Inquiry vs time to see if # of questions changes following a) rep reactive handling of HCP requests for long-term/maintenance use; b) PM changes.
	Grease removal/Can't get it out"/"Washed my hair many times"/"How to remove from scalp?"/ "Recommendations are not working"	Sending a follow-up questionnaire if the "Shampoo on dry hair" or general recommendations works after reply.
	Information in the use of SprayX post-4 weeks	Med Info Website (SprayX Data)
	Disposal of SprayX	Educate Pharmacists, Doctors, Med Info Website
	Paul Study: SprayX vs DrugX/ SprayX and DrugY®	Med Info Website (SprayX- Data)

	SprayX Study: SprayX vs DrugZ/ SprayX and DrugZ	Med info Website (SprayX-Data)
	How long to leave application on? / How long does it take to absorb?	Educate Pharmacists (part of counselling).
	Use with Vitamin D	Med Info Website
	Use in adolescence/Use in children	Med Info Website
	Confusion on instructions of "A can lasts 4 days" with their physician's SIG	PM improvements in wording.
	Application on off-label sensitive areas ("Can it be applied on […]". (Areas where PM advises not to use)	Med Info Website
	SprayX vs DrugA/SprayX and DrugA	Med Info Website
	Use in pregnancy/breastfeeding	Med Info Website
	Putting on gloves/socks/occluding area (band-aid) after application from patients/"used under occlusion"	Educate Pharmacists (part of counselling).
	Calcium levels	Med Info Website
	SprayX vs Phototherapy/ SprayX and Phototherapy	Med Info Website
	Staining fabrics	Educate Pharmacists (part of counselling).
	Removal of product. ("Do I wait for it to dry?", "Do I need to wipe off the excess?")	Educate Pharmacists (part of counseling).
	SprayX Trial Data	Med Info Website (SprayX Data)
	Any concerns with the treatment of hair (coloured hair/perm)	Educate Pharmacists (part of counselling), Med-Info Website
	Risks of inhalation	Educate Pharmacists (part of counseling).
	Use of moisturizer/sunscreen	Educate Pharmacists (part of counseling).
	Psoriasis of various kinds (Pustular psoriasis, plaque psoriasis, palmoplantar psoriasis, nail psoriasis)	Med Info Website
	Does it contain "…"	Med Info Website
	Positive feedbacks/Marketing feedback	Data Visualization: Use of MAFUSA
	Hyperpigmentation	Med Info Website
	Photocarcinogenicity/Sun exposure	Educate Pharmacists (part of counselling).
	Is SprayX the same as DrugB®	Med Info Website
	Accidentally got into eyes	Educate Pharmacists (part of counselling).
	Requests for OLA slide decks	Data Visualization: Use of MAFUSA

Cluster 5	Appearance/"why is the foam not foaming?"	Product Monograph improvements in wording. Suggestion: Snow like foam
	Cold feeling after application	Med Info Website
	Product defects	Data Visualization (Use of MAFUSA)
	Application near the eyes	Med Info Website
	What is the environmental impact of improper disposal?	Educate Pharmacists, Doctors, Med Info Website
	Compassionate use	Med Info Website (Instructions/Forms)
Cluster 6	SprayX patient cases program: Use beyond 4 weeks	N/A: Very old cases
Cluster 7	PV Case Reports - Off Label Use	Data Visualization: Use of MAFUSA

Table 2: SprayX cluster themes identified by HDBSCAN algorithm.

8.2 OintmentX Cluster Themes

OintmentX Cluster Themes		
Cluster	Themes	Potential Explorations
Cluster -1	Not enough cases to be identified as a cluster ex: stability once opened (Pharmacists)	NA
Cluster 0	Question was missing	NA
Cluster 1	Question was missing	NA
Cluster 2	Sample requests	Data Visualization: MUFASA (Geographical explorations in future)
Cluster 3	MSDS	Data Visualization: MUFASA
Cluster 4	MSDS (Mainly from pharmacists)	Data Visualization: MUFASA
Cluster 5	Does it contain "…" (Ingredients)	Med Info Website
Cluster 6	PV cases (AE, Off Label)	Data Visualization: MUFASA
Cluster 7	Product defects (cracks to the tube)	Data Visualization: MUFASA
Cluster 8	Safety inside mouth	Med Info Website

	Can it be applied to inside of the noses	Med Info Website
	Use of it more than 14 days	Med Info Website
Cluster 9	MRSA coverage w/OintmentX	Med Info Website
	Requests for OLA slides	Data Visualization: MUFASA
Cluster 10	Availability (Backorder)	N/A
Cluster 11	Availability of pack size of 15g	Compare it with in-use stability
Cluster 12	MSDS requests	Med-Info Website
Cluster 13	Receiving documents	N/A
Cluster 14	Receiving samples	Geographic Visualization
Cluster 15	Why is there a 3-month expiry date? (Patients/Pharmacists)	Med Info Website
Cluster 16	In-use stability after repackaging	Compare it with inquiry for availability of 15g inquiry

Table 3: OintmentX cluster themes identified by HDBSCAN algorithm.

8.3 OintmentY® Cluster Themes

OintmentY Cluster Themes		
Cluster	Themes	Potential Explorations
Cluster -1	Not enough cases to be identified as a cluster: Price is too expensive	N/A
Cluster 0	Question Missing	N/A
Cluster 1	Question Missing	N/A
Cluster 2	Market research- Off label use not caused by atopic dermatitis	N/A
Cluster 3	Question Missing	N/A
Cluster 4	Requests for slide decks	Data Visualization: MUFASA
Cluster 5	Does OintmentY contain "…" (Ingredients)	Med Info Website
Cluster 6	Request for dosing information in children <2 years old	Med Info Website
Cluster 7	Facial flush use with alcohol	Med Info Website, Educate Pharmacists (part of counselling)
Cluster 8	Availability of OintmentY	Date Visualization: MUFASA

		(Compare it with inquiry of stability once opened questions)
Cluster 9	Off label use with Vitiligo	Data Visualization: MUFASA
Cluster 10	Is there a Compassionate Use Program (CUP)?	
Cluster 11	CUP process	Med Info Website (Instructions, Forms)
Cluster 12	Off label use in sensitive area (eyes, face, gluteal cleft)	Data Visualization: MUFASA
Cluster 13	Maximum dose/Maximum duration to use	Med Info Website
Cluster 14	PV-AE Types of Report	Data Visualization: MUFASA
Cluster 15	PV-AE Types of Report with PSP# provided	Data Visualization: MUFASA
Cluster 16	Patient experiencing AE and wanting reimbursement	Data Visualization: MUFASA
Cluster 17	OintmentY use off label- inside mouths	Data Visualization: MUFASA
Cluster 18	Can it be used in sensitive reproductive areas? (genitals)	Med Info Website
Cluster 19	Requests for OintmentY/DrugC Study	Med Info Website #Inquiry vs time to see if # of questions change following a) rep handling of long-term safety/cancer warning message with Trial data b) removal of boxed warning
Cluster 20	Request for Off label use data for Vitiligo	Data Visualization: MUFASA
Cluster 21	Request for information for Off-label uses in other types of areas	Data Visualization: MUFASA
Cluster 22	The use of OintmentY with/without steroids + Why is OintmentY second line?	Med Info Website
Cluster 23	Phototherapy/Sunlight	Med Info Website, Educate Pharmacists (part of counselling)
Cluster 24	Use on eyelids/getting into the eyes	Med Info Website, Educate Pharmacists (part of counselling)
Cluster 25	Malignancy/Cancer risks	Med Info Website Changes in inquiry frequency over time to see if PM change (removal of boxed warning) has had an impact.

Cluster 26	Safety concerns: Renal failure	Med Info Website
Cluster 27	Product Defect: Difficulty getting ointment out of tube.	Product Feedback
Cluster 28	Request on information for the use on children under the age of 2	Med Info Website
Cluster 29	Request on information for the use on children under the age of 2 (provides strength in questions)	Med Info Website
Cluster 30	In-use stability of tube once opened	Data Visualization: MUFASA (Has it been increasing?)
Cluster 31	Doctor recommends putting tube in fridge	Opportunities to educate Doctors/Pharmacists to know that the stability is compromised once put in fridge. Med Info Website should upload information about refrigeration and stability.

Table 4: OitnmentY cluster themes identified by HDBSCAN algorithm.