AI-guided experimental summarization and sentiment analysis of Dockets

Rahul Paul[1], Arya Eskandarian[1], Luis Santana-Quintero[1], Kevin Kho[2], & Harinder Chahal[2]
1-Center for Biologics Evaluation and Research; 2-Office of the Commissioner

Abstract

FDA seeks public input for regulatory decision making via comments submitted to dockets. FDA experts manually extract and review comments before acting. However, with viral campaigns, some dockets, e.g., COVID-19 vaccines, can receive over 100,000 comments, making manual review untenable.

To help FDA staff make docket review less time-intensive and more efficient, we developed an experimental AI-guided tool to auto-extract, de-duplicate, summarize, and conduct sentiment analysis on comments.

Introduction

Problem:

I. FDA receives comments from public to dockets on rules, guidance, advisory committee meetings, etc.
II. Some dockets can have hundreds to thousands of comments
III. Manually preparing each docket for review and action is time consuming.

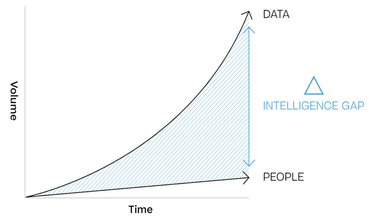

Figure 1. Volume of data with respect of time

Goal: To develop a proof-of-concept prototype that can:
I. Identify and remove duplicated comments.
II. Identify smaller sub-groups of related comments.
III. Understand sentiments of docket comments.
IV. Create summary of long docket comments.
V. To help FDA staff make docket review less time-intensive and more efficient

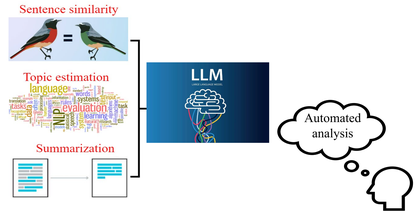

Figure 2. AI-based automated analysis

Materials and Methods

Dataset: Docket no. 742534
No. of comments: 122517

Figure 3. AI-guided pipeline

Our Pipeline steps:

1) Download the docket comments
2) Text embeddings of docket comments by a sentence-transformer
3) Apply K-Means clustering to divide the data into clusters and analyze cluster-intent
4) For every cluster, calculate cosine similarity for sentence similarity
5) Based on a predefined threshold for cosine similarity to remove the similar comments
6) Use a transformer model for experimental summarization
7) Use a transformer model for sentiment assessment

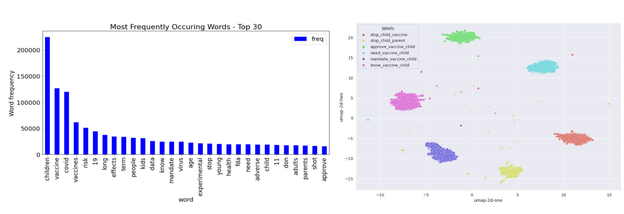

Figure 4. Preliminary data analysis on the docket

Results and Discussion

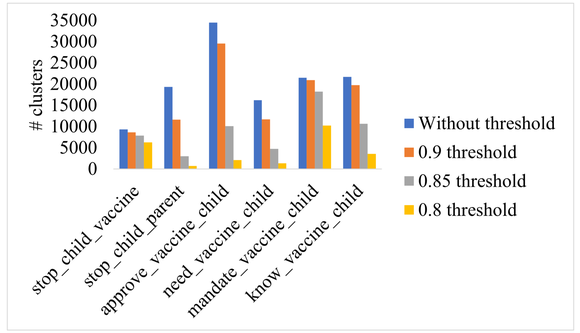

Figure 5. Removal of similar comments with varying thresholds

Summarization: An approach to simply summarize as humans do, which is to not impose the extractive constraint and allow for rephrasing.

101 words

Original sentence: Dear FDA: my comment today is regarding the upcoming decision for children age 5-11 to receive the covid vaccine. I am very much against this age group receiving the vaccine. I believe that children are more at risk from the vaccine than covid. As you know, the vaccine has caused death and severe adverse reactions in all age groups that have received it. We also know that covid is not spread asymptomatically so children do not put others at risk. Please do not approve this vaccine for these children. To do so would not at all be science based. Thank you.

45 words

Summary sentence: FDA is about to decide whether to approve the covid vaccine for children aged 5-11. The vaccine has caused death and severe adverse reactions in all age groups that have received it and is not spread asymptomatically so children do not put others at risk.

Intent analysis (Zero-shot method): Learn a classifier for new classes that are not seen in training

Per the statistical data- children are HIGHLY unlikely to suffer Covid-19 severe reactions. IT IS NOT scientifically determined the effects that a R-DNA can have on growing bodies. Please DO NOT approve the vax for our young children. Take the time to truly obtain the data! Thank you. Negative intent

Dear FDA: my comment today is regarding the upcoming decision for children age 5-11 to receive the covid vaccine. I am very much against this age group receiving the vaccine. I believe that children are more at risk from the vaccine than covid. As you know, the vaccine has caused death and severe adverse reactions in all age groups that have received it. We also know that covid is not spread asymptomatically so children do not put others at risk. Please do not approve this vaccine for these children. To do so would not at all be science based. Thank you. Negative intent

As a med/peds physician and parent to a 9 and 11 year old, I firmly and unequivocally support vaccination for 5-11 year olds. I trust that the data supports the safety and efficacy of the Pfizer mRNA vaccine. Positive intent

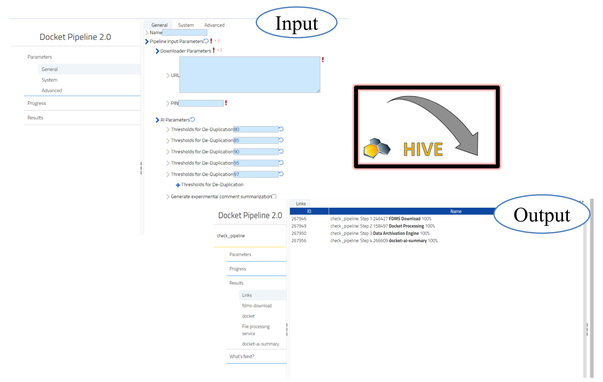

Figure 6. Docket processing API @HIVE

Conclusion

➢Developed an AI-guided experimental workflow to de-duplicate, summarize, and assess sentiment of comments, to accelerate decision making.
➢Our pipeline will facilitate the review process by reducing the number of similar comments as well as generating summary and sentiment.
➢The initial implementation (the prototype) of the AI pipeline is hosted at the HIVE platform, and it is available to registered users for testing through a user-friendly web interface.
➢Additional development and prototype refinement is underway.

Disclaimer: *The information in this presentation represents the opinions of the speaker and does not necessarily represent FDA's position or policy.*