Current Progress:

A web scraper developed by Michael Broad was utilized to extract articles from PubMed. I customized the search terms to exclusively retrieve articles related to ML/AI in focused ultrasound or its related domains. These articles are automatically transferred to an Excel sheet and organized into a tabular format, presenting essential details like titles, authors, abstracts, and publication dates. The scraper enables users to specify a date range for querying articles. The current strategy involves running this scraper on a monthly basis to continuously update the database with the latest articles.

Moreover, I designed a PDF processor that enables users to input any PDF article. The program then provides additional information about the article, encompassing a summary either generated through ChatGPT's API or an analyzer that rates sentence relevance and returns the most important sentences. It also includes classifications (FUS/Non-FUS, Supervised/Unsupervised) generated via a supervised machine learning model created from manually compiled databases. Additionally, a chatbot, along with the API key, employs ChatGPT's API to generate responses to user queries regarding the inputted article. All this information is displayed, with classifications represented as probabilities for each class. Users can subsequently input this data into the database. Presently, approximately forty articles from 2023 fulfill the scraper's search criteria and have been categorized within the database.

Areas for Advancement:

1. The existing scraping methods are confined to PubMed articles, neglecting potentially relevant articles from other journals. Consequently, developing analogous programs to extract data from other centralized databases would accelerate the rate of finding new articles.
2. If we opt for score-based summary generation, enhancing sentence segmentation methods could lead to more consistent and comprehensive sentences.
3. The expansion of classification models remains pivotal to enhance accuracy. Utilizing user feedback could potentially contribute to this improvement.
4. An automated approach to input a predetermined set of questions into the chatbot upon article submission would be useful, eliminating the need for manual input each time.

5. Extending the model's applicability to other classifications, such as the Treatment Cycle or Indication, would further its automation.

Ultimate Objective:

The ultimate aim of this project is to achieve full automation, spanning from scrapers to the integration of data into the final database. To achieve this goal, the following steps need to be automated:

- Automate scraper execution on designated intervals without human intervention.
- Simplify the process of directing scraped articles to the PDF processor, potentially through an automated mechanism.
- Streamline the input of processor results into the final database platforms like Sharepoint or SalesForce.

By automating these tasks, the workflow will be optimized for efficiency and accuracy.