Summarized Report

Approach

This project is designed to retrieve and process research papers from the PubMed database, focusing on extracting specific details and storing them in a CSV file. The approach leverages the PubMed API to fetch papers based on user-defined queries. The retrieved data is then processed to identify key information, ensuring that only papers with at least one non-academic affiliation are included in the final output, while excluding those with solely academic affiliations.

Methodology

The methodology consists of four key steps:

1. **Data Fetching**
   The PubMed API is used to search for research papers matching a user-provided query. Detailed records of the matching papers are retrieved for further processing.
2. **Data Processing**
   From the fetched records, specific details are extracted, including:
   - PubMed ID
   - Title
   - Publication Date
   - Non-academic Authors
   - Company Affiliations
   - Corresponding Author Email

     This step involves identifying authors with non-academic affiliations, such as those linked to companies or other non-academic entities, based on predefined criteria or keywords.
3. **Data Filtering**
   A filtering mechanism is applied to exclude papers where all authors are affiliated exclusively with academic institutions. Only papers with at least one non-academic affiliation are retained for further processing.
4. **Data Storage**
   The processed and filtered data is appended to a user-specified CSV file. The file is structured to ensure readability and accessibility, allowing for easy analysis of the stored information.

Results

The result of this process is a CSV file containing the extracted details of research papers that have at least one non-academic affiliation. Each time the script is executed with a new query, the CSV file is updated with additional relevant papers, creating a cumulative dataset that grows

with each run. This provides an organized and continually expanding resource for tracking papers with non-academic involvement.