# Report: Fetching PubMed Papers

## Approach:

Your approach involves using the BioPython library to interact with the PubMed API and fetch research papers based on a given query. The process is divided into several steps:  
  
1. Setting Up:  
- Import necessary libraries: `Bio.Entrez` for PubMed API interaction and `pandas` for data manipulation.  
- Set your email address to use the PubMed API.  
  
2. Fetching Papers:  
- Define a function `fetch\_pubmed(query: str, max\_results: int = 10)` to search for papers based on the query.  
- Use `Entrez.esearch` to search for papers and retrieve their IDs.  
- Use `Entrez.efetch` to fetch detailed information about the papers using their IDs.  
- Extract relevant details such as PubMed ID, title, publication date, authors, affiliations, and corresponding author email.  
  
3. Saving Results:  
- Define a function `save\_to\_csv(data, filename)` to save the fetched data into a CSV file using `pandas`.  
  
4. Command-Line Interface:  
- Use `argparse` to create a command-line interface for the script.  
- Allow users to pass the search query and filename dynamically.

## Methodology:

1. Search for Papers:  
- Use the PubMed API to search for papers based on the provided query.  
- Retrieve a list of paper IDs.  
  
2. Fetch Paper Details:  
- Fetch detailed information about each paper using the retrieved IDs.  
- Extract and organize relevant details such as PubMed ID, title, publication date, authors, affiliations, and corresponding author email.

3. Identify Non-Academic Authors and Affiliations:  
- Check author affiliations for keywords like 'pharma' or 'biotech' to identify non-academic authors and their company affiliations.  
  
4. Save Results to CSV:  
- Convert the extracted data into a pandas DataFrame.  
- Save the DataFrame to a CSV file.

## Results:

The script successfully fetches and organizes details of research papers based on the provided query. The results include:  
  
- PubMed ID: Unique identifier for each paper.  
- Title: Title of the paper.  
- Publication Date: Date when the paper was published.  
- Non-Academic Authors: List of authors affiliated with non-academic institutions.  
- Company Affiliations: List of affiliations for non-academic authors.  
- Corresponding Author Email: Email address of the corresponding author (if available).  
  
The results are saved to a CSV file, making it easy to analyze and share the data.