



# Project Report: MedQueryPy

## 1 Introduction

MedQueryPy is a **Python package** designed to fetch research papers from **PubMed** and filter them based on author affiliation. It now leverages **AI-powered author classification and research paper summarization** to enhance the accuracy of results.

The goal of this project is to help researchers, professionals, and biotech firms quickly find **non-academic research papers** that are relevant to their field.

---

## 2 Problem Statement

Finding research papers authored by **non-academic professionals** in the **pharmaceutical and biotech industries** is challenging. PubMed provides a vast collection of academic papers, but filtering out non-academic contributions manually is inefficient.

This project aims to:

- ✓ Fetch **PubMed research papers** based on a user's query.
  - ✓ Identify **at least one non-academic author** in the paper.
  - ✓ Use **AI to classify author affiliations** (academic vs. non-academic).
  - ✓ Summarize the paper's abstract using **GPT-4**.
  - ✓ Provide results in an **easy-to-use CSV format**.
  - ✓ Offer a **command-line interface** and a **Python module** for flexibility.
- 

## 3 Approach & Methodology

### Step 1: Fetching Papers from PubMed

- ♦ The program **sends a request to the PubMed API** using a user-provided search query.
- ♦ PubMed returns a list of **unique paper IDs (PubMed IDs)**.

## Step 2: Extracting Paper Details

- ♦ The system retrieves **metadata** for each paper, including:
  - **Title**
  - **Publication Date**
  - **Authors & Affiliations**
  - **Corresponding Author Email**

## Step 3: Identifying Non-Academic Authors with AI

- ♦ Instead of simple keyword filtering (e.g., excluding “University” and “Institute”), we now use **GPT-4** to analyze author affiliations.
- ♦ The AI determines if an author is affiliated with a **non-academic organization** (such as a biotech or pharmaceutical company).

## Step 4: Summarizing Research Papers Using AI

- ♦ The package uses **GPT-4** to generate a **one-line summary** of each research paper’s abstract.
- ♦ This helps users quickly understand the paper’s main findings without reading the full text.

## Step 5: Exporting Results

- ♦ Users can **print results on the console** or **save them to a CSV file** for future reference.
- 

# 4 Implementation

## Project Structure

The project is divided into two main components:

- ✚ `fetcher.py` → Handles PubMed API requests, AI-based filtering, and summarization.
- ✚ `cli.py` → Provides a command-line interface for fetching and saving research papers.

## Installation & Usage

To install MedQueryPy:

```
pip install medquerypy
```

To fetch papers:

```
get-papers-list "COVID-19 vaccines"
```

To save results to a CSV file:

```
get-papers-list "COVID-19 vaccines" -f results.csv
```

---

## 5 Results

- ✓ The tool successfully fetches research papers and **filters** those with **non-academic authors** using AI.
- ✓ It accurately **classifies affiliations** and **summarizes papers** using **GPT-4**.
- ✓ The output is **saved in a CSV file** for easy reference.

PubmedID	Title	Publication Date	Non-academic Author(s)	Company Affiliation(s)	Corresponding Author Email	Summary
12345678	COVID-19 Vaccine Study	2025-01-01	Dr. John Doe	XYZ Biotech	johndoe@xyzbiotech.com	The study explores the effects of COVID-19 vaccines across various age groups.

---

## 6 Conclusion & Future Scope

### What Achieved 🎯

- ✓ Built a fully functional **Python package** with CLI support.
- ✓ Integrated **AI-powered filtering** for non-academic authors.
- ✓ Implemented **GPT-4-based research paper summarization**.
- ✓ Enable easy **CSV export** for research data.

### Next Steps 🚀

- ◆ Improve **author affiliation classification** by training a custom AI model.
  - ◆ Optimize **API request efficiency** for faster processing.
  - ◆ Extend AI features to **summarize multiple papers at once**.
- 

## 7 Final Thoughts

MedQueryPy is a powerful tool that simplifies the process of finding **non-academic research papers** and extracting key insights. By integrating **AI-based filtering and summarization**, it provides a much smarter way to navigate research data.

This tool is useful for **biotech firms, pharmaceutical companies, and research analysts** who need quick access to relevant industry studies.