

Aganitha 2025 Python Take-Home Project

Candidate Name: Kaviya Sri V

Project Title: PubMed Research Paper Finder with Non-Academic Author Detection

OBJECTIVE:

To build a Python-based tool that fetches research papers from PubMed using user-defined queries and filters papers that include at least one **non-academic author affiliated with pharmaceutical or biotech companies**.

TOOLS & TECHNOLOGIES:

- **Python 3.13**
- **Poetry** (for dependency and CLI management)
- **Streamlit** (for LLM-style web UI)
- **Requests, Pandas, LXML, ElementTree**

APPROACH:

1. Queried PubMed using **Entrez eSearch/eFetch APIs**
2. Parsed XML to extract paper metadata, authors, affiliations
3. Filtered out academic affiliations using keyword heuristics
4. Presented results via:
 - Command-line interface (get-papers-list.py)
 - Streamlit-based modern UI (app.py)
5. Saved valid results to CSV with columns: PubMedID, Title, Authors, Affiliation, Email

FEATURES IMPLEMENTED:

- Search from CLI or web UI
- Real-time filtering of non-academic authors
- CSV download functionality
- Debug mode for tracing
- Modular code in fetcher.py for testability
- LLM assistance used responsibly

LLM USAGE:

Used ChatGPT (GPT-4) to:

- Debug environment setup (Poetry, Streamlit)
- Validate filters and error handling

TEST OUTPUT:

Used the query: cancer
Found and saved relevant CSV
Verified only non-academic affiliations included
GUI and CLI both validated

DECLARATION:

I confirm this submission was completed by me, using permitted tools, within the guidelines.
All code is available on GitHub and video demo link is shared.