📄 Project Report: MedQueryPy

1️⃣ Introduction

This report covers the approach, methodology, and results of **MedQueryPy**, a Python package designed to fetch research papers from PubMed and identify those with authors affiliated with pharmaceutical or biotech companies.

2️⃣ Problem Statement

Research papers contain valuable insights, but finding relevant papers from **non-academic sources** is difficult. The goal of this project is to:

- Fetch **PubMed research papers** based on a user's query.
- Identify papers with **at least one non-academic author**.
- Extract key details like **title, publication date, authors, and affiliations**.
- Save results in an **easy-to-use CSV format**.
- Provide a **command-line tool** and a **Python module** for flexibility.

3️⃣ Approach & Methodology

Step 1: Fetching Papers from PubMed

- We use the **PubMed API** to search for research papers based on user input.
- PubMed returns a **list of unique paper IDs (PubMed IDs)**.

Step 2: Extracting Paper Details

- Each paper ID is processed to extract:
  - **Title**
  - **Publication Date**
  - **Authors & Affiliations**
  - **Corresponding Author Email**

Step 3: Identifying Non-Academic Authors

- The program **filters out academic institutions** like "University," "Institute," "Lab."
- Authors affiliated with **biotech/pharma companies** (e.g., "Biotech," "Inc.," "Ltd.") are included.

Step 4: Exporting Results

- Results can be **printed on the console** or **saved to a CSV file** using the CLI.

---

4 Implementation

The project is divided into two key components:

- `fetcher.py` → Handles all API requests & data processing.
- `cli.py` → Provides a command-line interface to run the tool.

Installation & Usage

To install the package:

```
pip install medquerypy
```

To fetch papers:

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

To save results:

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

---

5 Results

- The tool successfully fetches research papers and filters those with **non-academic authors**.
- **Sample output (CSV file):**

PubmedID	Title	Publication Date	Non-academic Author(s)	Company Affiliation(s)	Corresponding Author Email
12345678	COVID-19 Vaccine Study	2025-01-01	Dr. John Doe	XYZ Biotech	johndoe@xyzbiotech.com

6 Conclusion & Future Scope

What We Achieved 🎯

- ✅ Built a working **Python package** with CLI support.
- ✅ Successfully extracted and filtered **non-academic research papers**.
- ✅ Made the tool easy to use with **CSV exports & debugging mode**.

Next Steps 🚀

- Improve **accuracy of company affiliation detection**.
- Optimize **API calls for faster performance**.
- Integrate **AI models** to analyze paper relevance.

7 Final Thoughts

MedQueryPy is a powerful tool that simplifies the process of fetching and analyzing research papers. Whether you're a researcher, biotech firm, or data analyst, this tool makes it easier to extract meaningful insights.