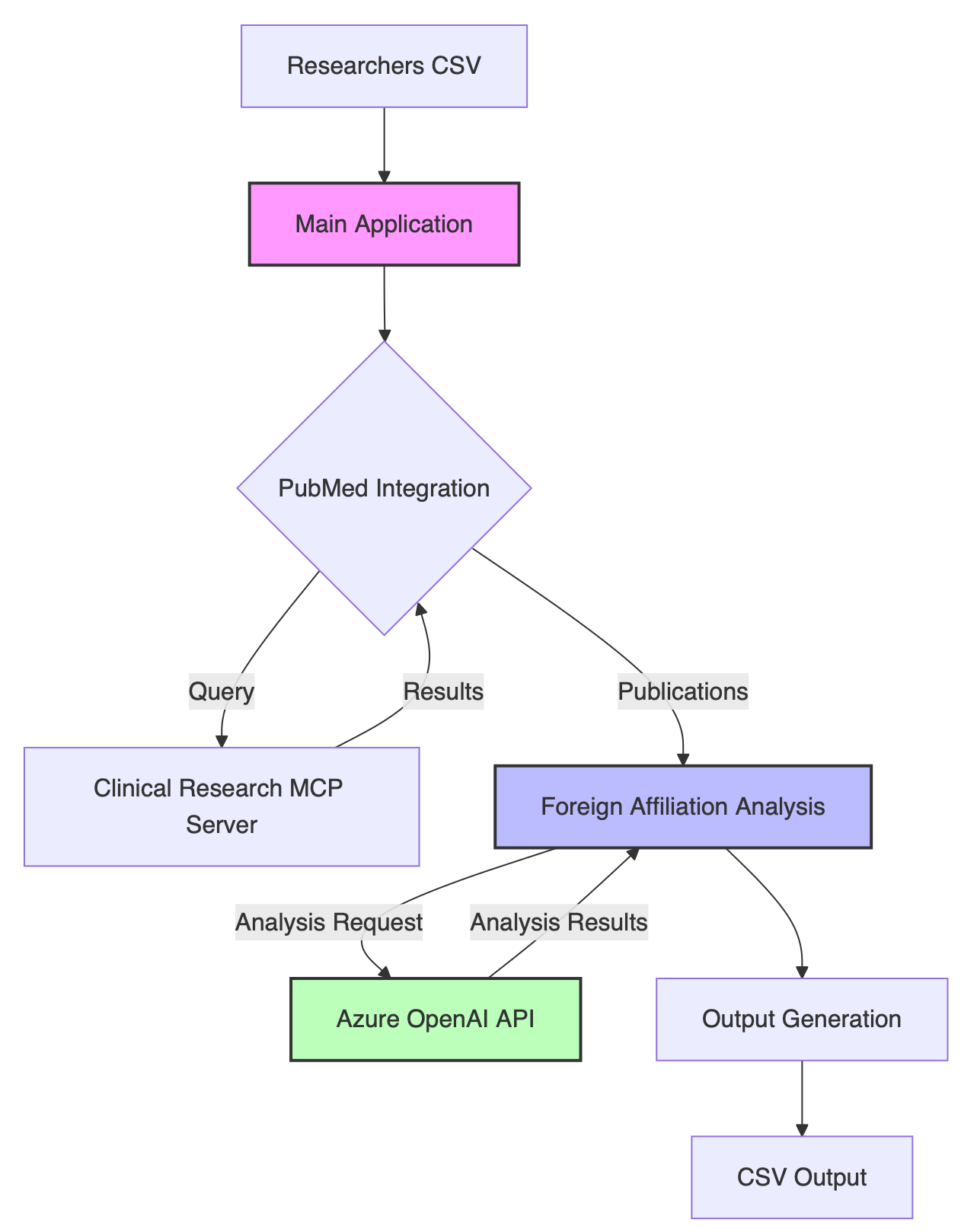
# Foreign Disclosure Analysis Tool for BCH Researchers

This tool analyzes PubMed publications for Boston Children's Hospital (BCH) researchers to identify foreign affiliations and collaborations, with a particular focus on Russia, North Korea, Iran, and China. The analysis results are output to a CSV file with specific fields.

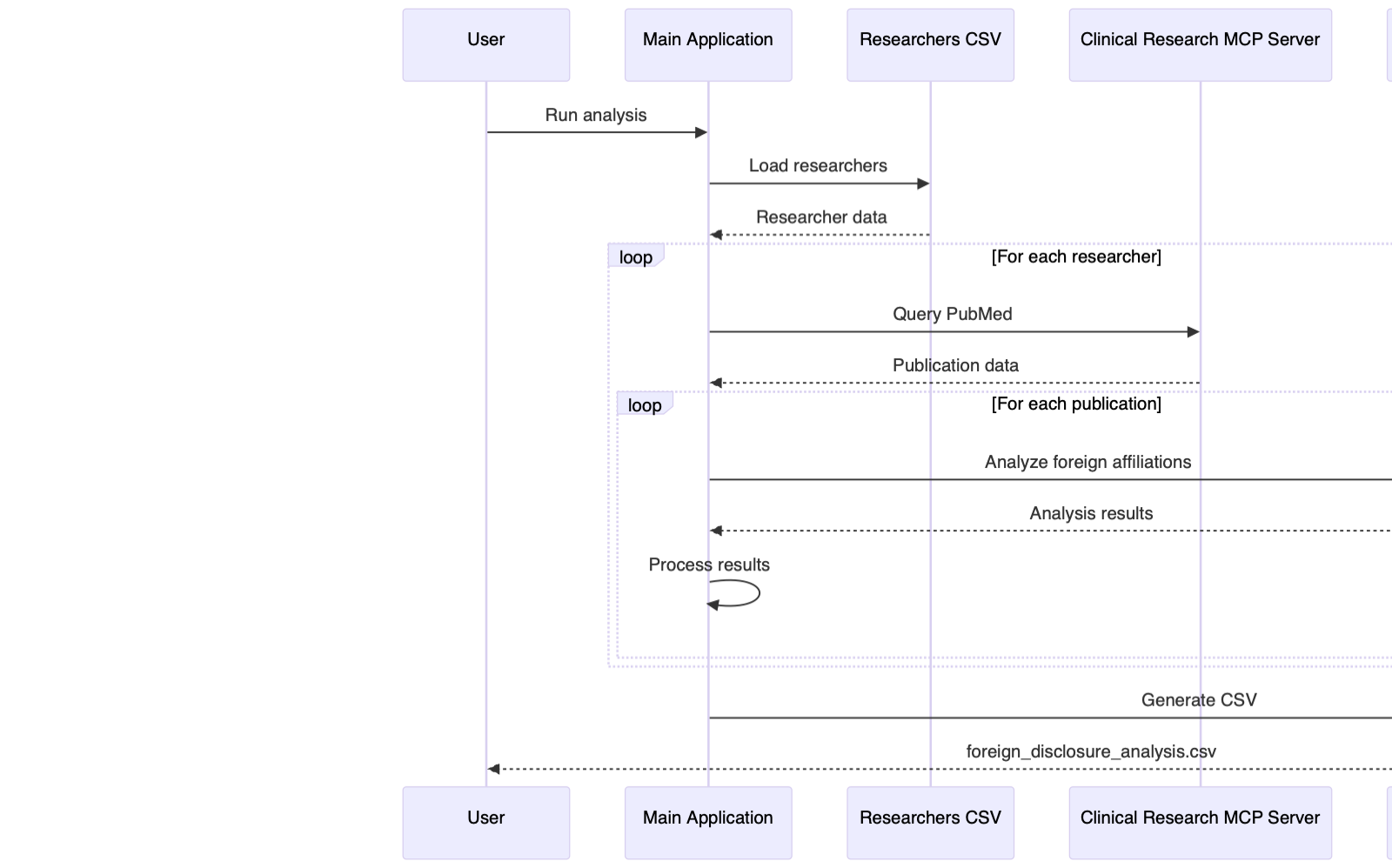
## Overview

The solution uses the Azure OpenAI API and the Clinical Research MCP Server to analyze PubMed publications for foreign disclosure requirements. It processes a list of BCH researchers from a CSV file, queries PubMed for their publications, analyzes the publications for foreign affiliations, and outputs the results to a CSV file.

## System Architecture



## Data Flow



## Features

* Dynamic loading of researcher data from CSV
* PubMed integration via Clinical Research MCP Server
* Foreign affiliation analysis using Azure OpenAI API
* Special flagging for countries of concern (Russia, North Korea, Iran, China)
* Confidence scoring for foreign involvement
* CSV output with detailed information

## Requirements

* Python 3.9+
* Azure OpenAI API access
* Clinical Research MCP Server access

## Installation

1. Clone the repository:

* git clone https://github.com/BCH-IDHA/foreign-disclosure-analysis.git  
  cd foreign-disclosure-analysis

1. Install the required dependencies:

* pip install -r requirements.txt

1. Configure the .env file with your Azure OpenAI API credentials:

* AZURE\_OPENAI\_API\_ENDPOINT=your\_endpoint  
  AZURE\_OPENAI\_API\_KEY=your\_api\_key  
  AZURE\_OPENAI\_API\_VERSION=your\_api\_version  
  AZURE\_OPENAI\_DEPLOYMENT=your\_deployment\_name  
  AZURE\_OPENAI\_MODEL=your\_model\_name

## Usage

1. Prepare your researchers CSV file with the following format:

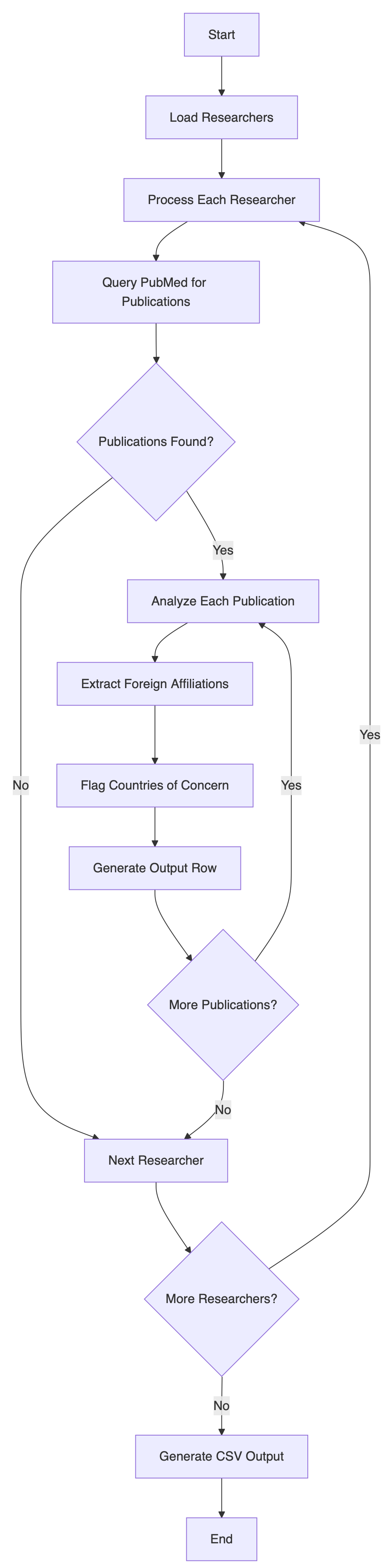
* Researcher last name, Research first name  
  Smith, John  
  Doe, Jane

1. Run the analysis:

* python main.py

1. Review the output in foreign\_disclosure\_analysis.csv

## Analysis Process

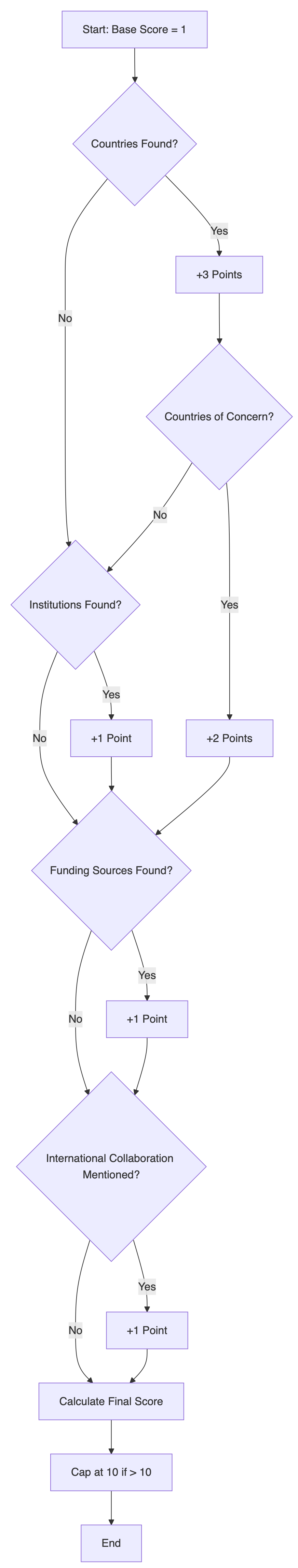


## Output Format

The output CSV file contains the following columns:

* **publication\_name**: The name of the journal where the research was published
* **research\_title**: The title of the research publication
* **author\_name**: The name of the BCH researcher
* **organization\_affiliation**: The organization affiliation (Boston Children's Hospital)
* **countries\_of\_origin**: Countries of origin/association
* **flagged**: Indicates whether the publication has any countries of concern (Yes/No)
* **flagged\_countries**: Lists the specific countries of concern that were flagged
* **confidence\_score**: Confidence score (1-10) regarding foreign involvement
* **funding\_source**: Funding sources for the research

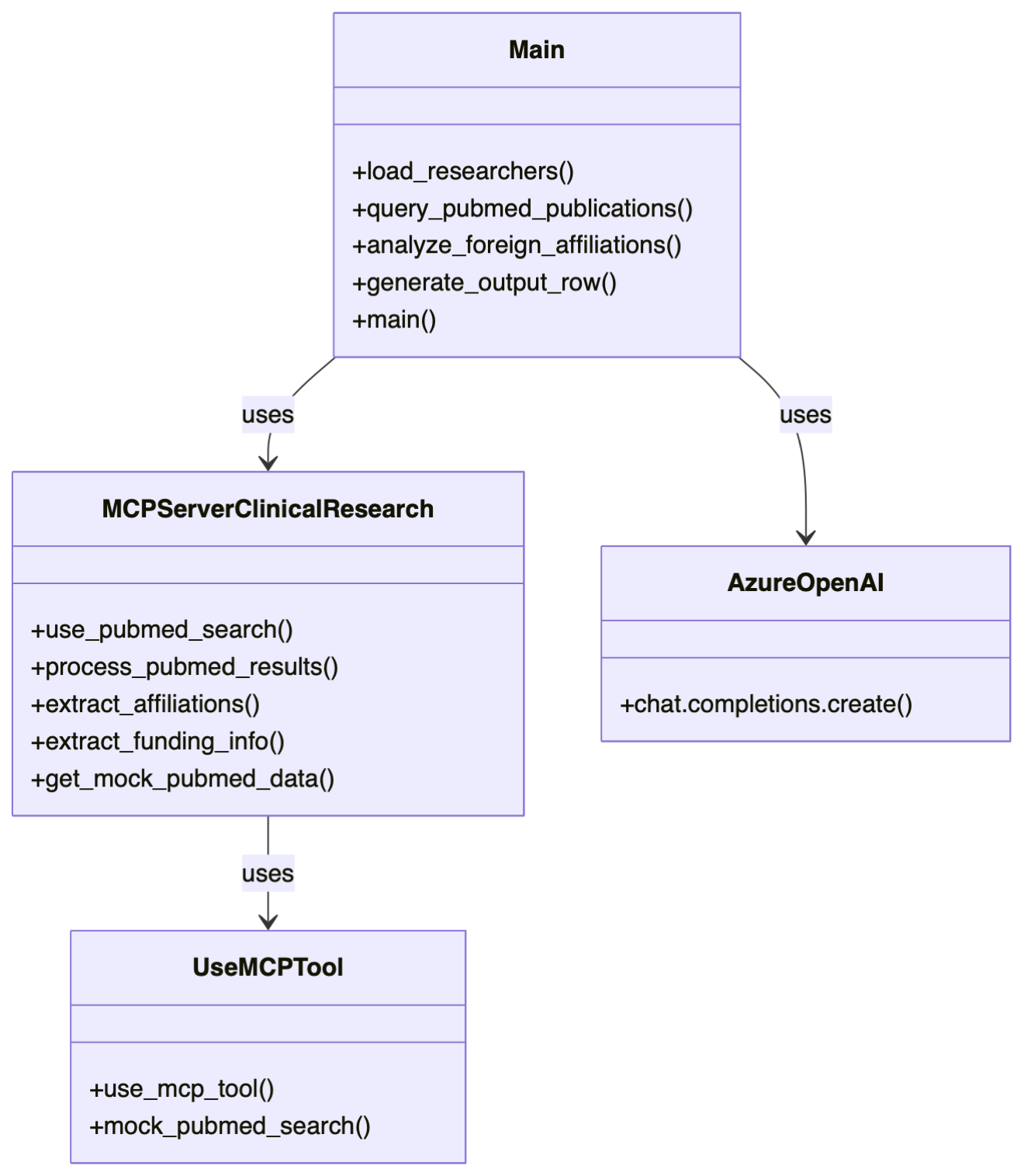
## Confidence Scoring Algorithm



## Project Structure

* main.py: Main script for the analysis
* mcp\_server\_clinical\_research.py: Wrapper for the Clinical Research MCP Server
* use\_mcp\_tool.py: Wrapper for the MCP tool functionality
* requirements.txt: Required Python dependencies
* researchers.csv: Input file containing BCH researcher names
* solution-plan.md: Implementation plan and progress tracking
* README.md: Project documentation

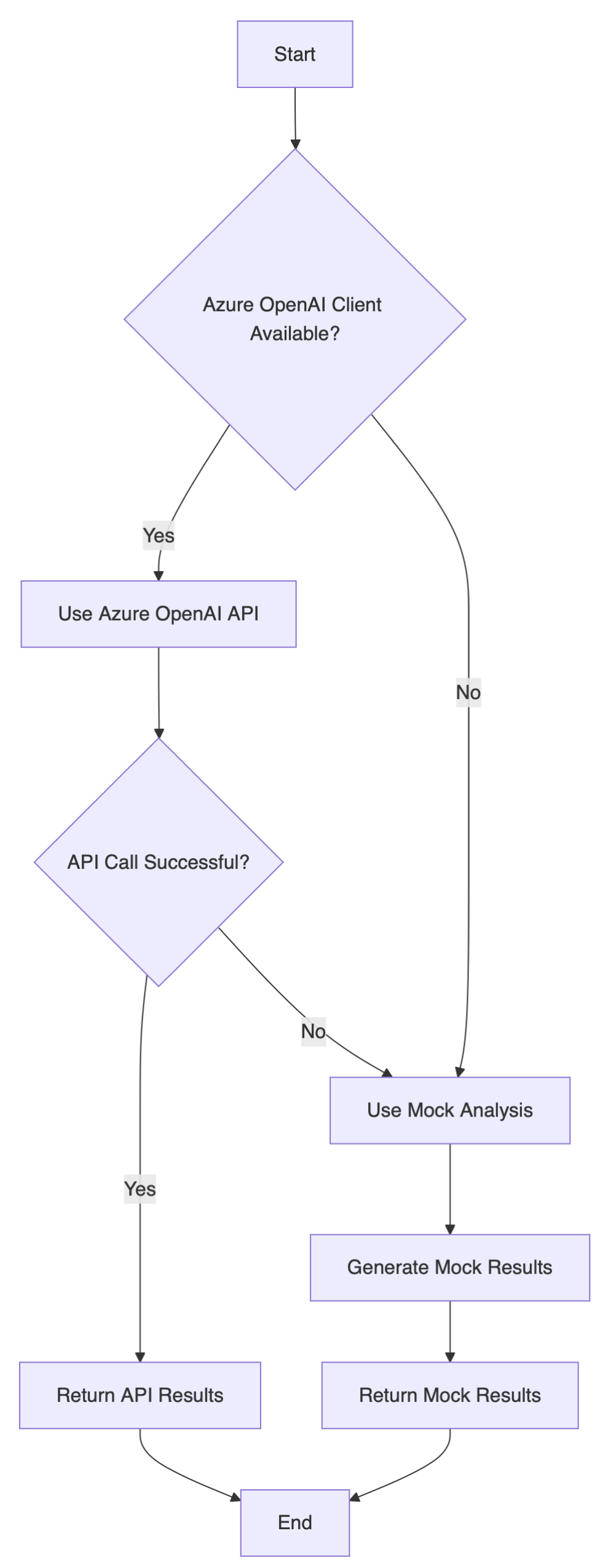
## Component Relationships



## Development

For development and testing purposes, the code includes mock implementations of the PubMed search functionality and Azure OpenAI API. This allows for testing without actual API access.

### Mock Implementation Flow



## License

[Specify your license here]

## Acknowledgments

* Boston Children's Hospital
* Azure OpenAI API
* Clinical Research MCP Server