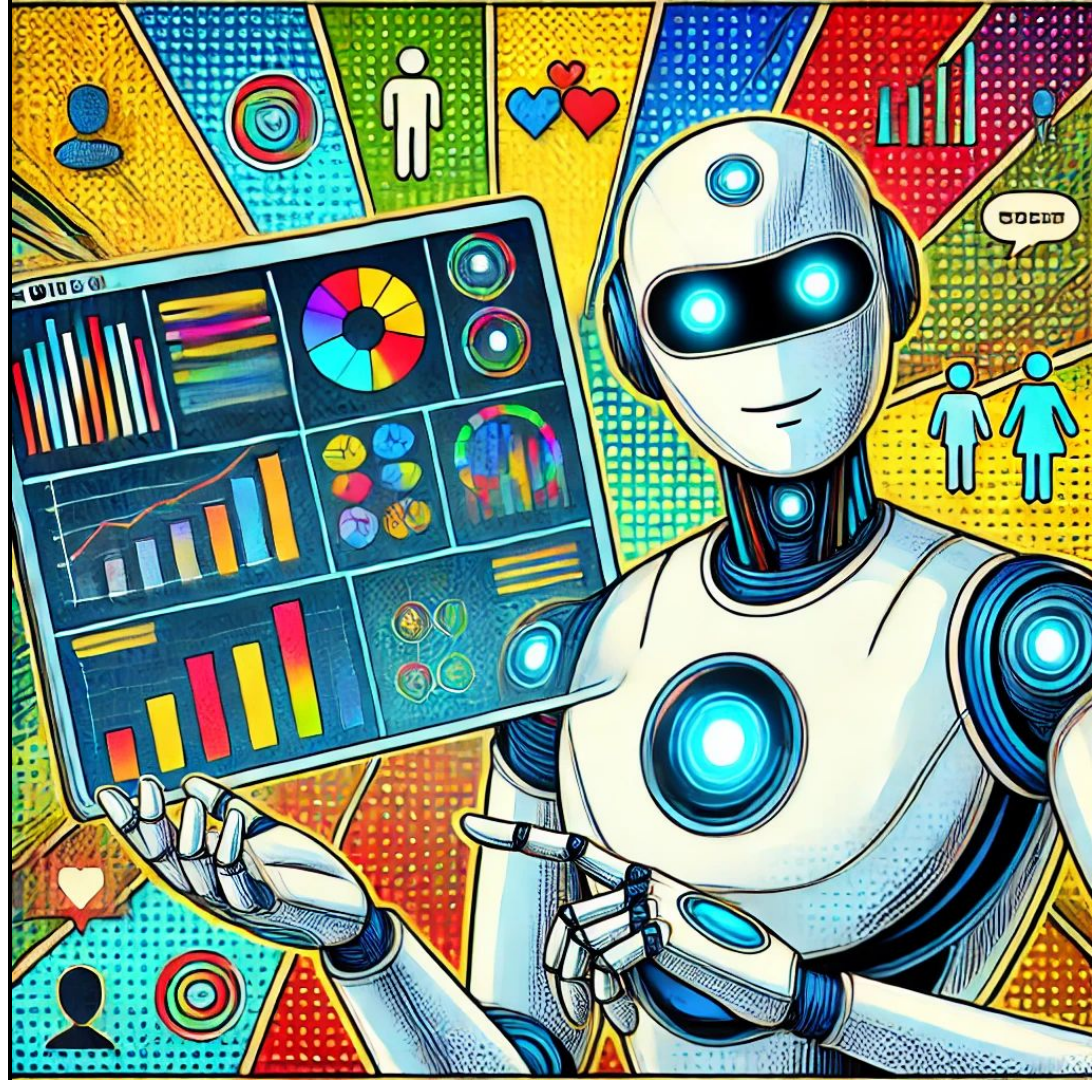


# Project 4

Agentic AI for Census API  
Data Extraction

Michael Minzey

CIS 678  
12/8/2024



# The Challenge with ACS Data Accessibility

- The American Community Survey (ACS) provides critical social and economic data.
- With over 20,000 variables, it is challenging to navigate.
- Accessing ACS data often requires technical skills and API expertise.
- Non-technical researchers struggle to utilize this valuable resource.





# Streamlining Data Access with Agentic AI

- Natural Language Interface
- LLMs for Parsing Intention and Locations
- Dynamic API Construction
- Data Retrieval
- User-Friendly Presentation



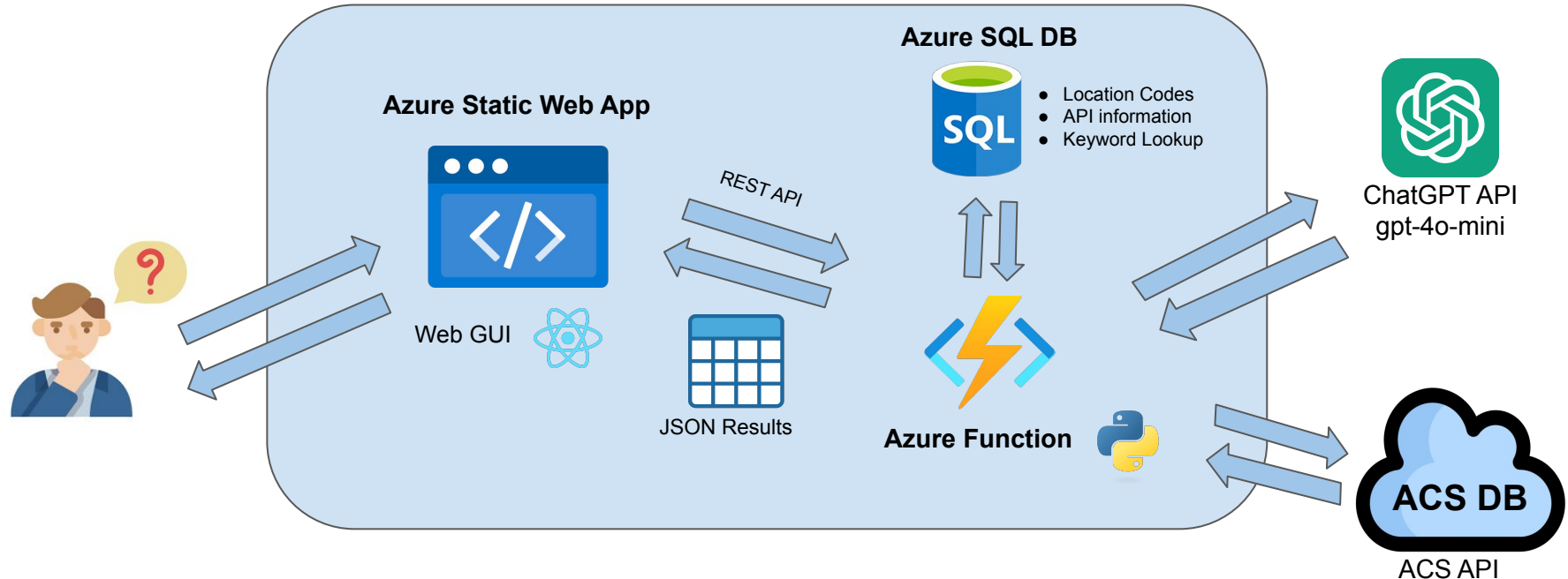
# ACS API: A Gateway to US Socioeconomic Data

- Provides metric data for over 20,000 variables.
  - Example: B08124\_031E
  - Name: Estimate!!Total!!Walked!!Service occupations
- 635 unique “concepts”
  - Example: Means of Transportation to Work by Occupation
- Covers city, county, and state levels.
  - Using a PLACEFP/COUNTYFP and STATEFP code
- Each API call retrieves data for one location based on specified codes.
- Returns data in JSON format for further processing.

**Example - Count of Service Workers Walking to Work in Michigan:**

[https://api.census.gov/data/2022/acs/acs5?get=NAME,B08124\\_031E&for=place:26](https://api.census.gov/data/2022/acs/acs5?get=NAME,B08124_031E&for=place:26)

# Putting the Pieces Together



# Agentic AI Workflow



Get user query



Validate user query (pass/fail/refine)



Parse query for categories, keywords, intention, and locations



Check for valid US locations and get location codes



Get candidate API codes using categories and keywords



Select appropriate API codes based on user intent



Build API calls



Retrieve data



Create results table



Update with friendly metric names and title



Present results to user

## Chat Window

**You:** What is the population of Grand Rapids MI and Chicago Illinois?

**Assistant:** Here are the results for your query.

Enter your question

SUBMIT

## Results

Population Comparison: Grand Rapids, MI vs. Chicago, IL

Metric	Chicago, Illinois	Grand Rapids, MI
Total Population	2,721,914	198,096

# Challenges, Obstacles, and Limitations

- API candidate selection strategy
  - Created 13 categories for the 635 concepts
  - Created a synonym dictionary to assist with matching user and API keywords
- Added a refinement catch for the LLM to improve ambiguous user query
- Recursive LLM strategy for determining final API calls
  - Chunking API candidate lists of 300
- Azure Static Web App APIs have ~45 sec compute window before returning a 500 error
- Significant preprocessing of API metadata
- LLM doesn't always understand the user intention and context when reviewing API results





# Challenges, Obstacles, and Limitations

- Prompt refinement is an art
- Difficult to catch all possible outcomes
  - Unit testing was key
- Using JSON structured output was critical to successful control logic
- Migrating from local dev to cloud hosting was a major hurdle - many technical issues. GitHub CI/CD was very helpful but made debugging slow





# Lessons Learned & Optimizations

- Using LLMs can be a powerful method to add discernment to your project
- Develop stronger ambiguity handling
- Include analytics and visualizations
- Advanced keyword matching
- Export data as CSV



# ACS AI Query Tool

To access the project:

<https://michaelminzey.com/acs-query>

