Day 19 – 23 July 2025 (Wednesday):

Summarization and Data Processing

On this day, the focus shifted from **data extraction** to **processing and summarization** of the collected web content. After the Automated Search Bot successfully retrieved URLs and page content, the next step was to condense the information into concise, meaningful summaries that could be easily interpreted by users.

Using **OpenAI's text summarization model**, I integrated AI into the workflow to automatically process multiple web pages. The agent analyzed textual content from each URL, identified key points, and generated **coherent summaries**. The session included exercises on crafting effective prompts to ensure that summaries were **accurate, concise, and contextually relevant**. Several iterations were conducted to refine output quality, handling cases where content was too verbose or scattered across multiple sections of a page.

The mentor also explained the importance of combining **data extraction and AI summarization** effectively. This step highlighted how automation and AI can reduce manual effort in research and deliver structured insights from unorganized web data. Testing involved cross-checking summaries against original content to verify **accuracy and relevance**, ensuring the bot produces reliable results.

Learning Outcome:

Acquired practical experience in **information condensation using AI models**. Learned how to integrate web data collection with **language model-driven summarization**, forming the basis for efficient and autonomous information retrieval systems.