

Title: Building an AI Agent to Identify Subsidiaries and Associated Domains of a Given Company through Web Browsing

Objective:

The goal of this assignment is to evaluate your ability to design and implement an AI agent capable of **browsing the internet to identify and list subsidiaries and associated domains for a given company name**. This task will assess your skills in data collection, web scraping, natural language processing, and AI-driven information retrieval, all within a constrained timeframe.

Requirements:

1. **Data Collection through Web Browsing:**
 - Design a strategy for the AI agent to browse the internet, focusing on reputable sources such as company websites, business directories, news articles, and other relevant online resources.
 - Implement a web scraping or API-based approach to collect relevant data. Consider the ethical and legal aspects of web scraping.
 - Document the data sources used and the criteria for selecting them.
2. **Information Extraction:**
 - Develop methods to extract relevant information (subsidiaries and associated domains) from the collected data. This could involve techniques LLMs for extracting company names, domain names, and relationships between them.
 - Discuss how you handle challenges like disambiguation of company names, noise in the data, and extraction accuracy.
3. **AI Agent Implementation:**
 - Implement the AI agent with the following capabilities:
 - Accepts a company name as input.
 - Browses the web to find relevant information.
 - Extracts and compiles a list of subsidiaries and associated domains.
 - Ensure that the implementation is functional and can produce results within the 2-3 day timeframe.
4. **Evaluation:**
 - Provide an evaluation of the agent's performance, discussing both its successes and limitations.
 - Suggest potential improvements or next steps if given more time.
5. **Documentation:**
 - Document your approach in a concise report, covering the data collection process, methods used for information extraction, and the AI agent's implementation.
 - Include any assumptions made, challenges faced, and how they were addressed.

Deliverables:

1. **Code:** A well-organized script or Jupyter notebook containing the AI agent's implementation, including any web scraping and NLP code.
2. **Technical Report:** A concise report (2-3 pages) outlining your approach, results, and recommendations.