**1. Objective:**

The design document should detail out overall approach on solving the problem. This could include any class diagrams/sequence diagrams indicating the approach and a high level traceability matrix.

**2. Key Features:**

* Data Collection: Gather sustainability-related data from various sources such as APIs, and pdf files
* Data Analysis: Perform comprehensive analysis on collected data to derive key sustainability metrics using spacy and chat gpt.

**3. Technology Stack:**

* Python: Primary programming language for application development.
* Libraries:
  + Pandas: Data manipulation and analysis.
  + pyPDF2: Reading pdf files
  + Flask: Web framework for building the application's backend.
  + AzureOpenAI, OpenAI : to connect to Azure Open AI resource
  + Spacy: NLP program to remove the unwanted text
  + Bs4: Beautiful soup to get the text from html
  + Requests: to call the url and get the html
  + Selenium: to do webcrawling and fetch data.
* Tools:
  + Git: Version control for collaborative development.
  + VS Code: Development and deployment.
  + Jupyter Notebook: Interactive development and prototyping.
  + Azure: For deploying code to app service and deploying the model to azure ai service

**4. System Architecture:**

* Backend: Flask will handle incoming requests, perform data processing, analysis, and generate responses.
* Data Sources: APIs, databases, pdf files, or web scraping techniques can be utilized to collect sustainability data.
* Analysis Engine: Python scripts will perform data analysis and calculate sustainability metrics.

**5. Data Flow:**

1. Flask receives the request and forwards it to the appropriate backend function.
2. Backend retrieves data from specified sources (APIs, webcrawlers, etc.) or performs analysis on existing data.
3. Data is processed and analyzed using Python scripts.
4. Analysis results are formatted and sent back to user.

**6. Deployment:**

* Deploy the application on cloud platforms to Azure for scalability and reliability.
* Configure monitoring and logging to track application performance and identify issues promptly.