Introduction:
You will be given a set of taxonomies (2, 2_2). For each set you will have to perform 12 tasks that will focus on the usability of the tool: Ease of use, Efficiency, and Effectiveness. And, when you have finished, you will have to answer a questionnaire.

Taxonomies are used to structure information. They consist of a root node, inner nodes (categories) and leaves (characteristics). Even if taxonomies are famous because of biology, taxonomies are used in many other fields such as Software Engineering to keep an organization.
Taxonomies can be constructed from scratch, merged, or expanded using expert knowledge. Using an LLM as the "expert" can facilitate these tasks.

PLEASE FOLLOW EVERY STEP, DO NOT SKIP ANY STEPS OR ELSE YOU WILL GET ERRORS
ONCE YOU MOVE TO A NEW AGENT YOU CANNOT GO BACK

If you encounter any of the following issues, restart the whole application:

Known bugs:
- **Going from Agent 1 The Aggregator back to Agent 2 The Preparator results in an error**
- **In Agent 1 The Aggregator, during the taxonomy construction, clicking the elements in the visualization may result in infinite reloading or the interface disappearing completely**
- **In Agent 1 The Aggregator, categories and characteristics yet to be placed may disappear. If that happens just add them manually.**

Tasks for Taxonomy 2 (Technical)

Agent 2:

1. Data preparation
- 1a. In the welcome page, read the quick start guide and click *Start Pipeline*.
- 1b. In the current *Settings* page click on *Refresh Models* and connect to the large language model (LLM) we downloaded through LM Studio *smollm2-360m-instruct* (make sure LM Studio is running in the background). Check that the embeddings are detected by clicking *Test Embedding Model* and click on *Test Connection* to see if the LLM responds.
- 1c. Click on *Proceed to Data Upload*.
- 1d. In *Data Input* click on *CSV Import* upload the CSV files: *taxonomy2.csv* and *taxonomy2_1.csv* containing the elements

of Taxonomy 1. Remember to click on *Click here to Import the uploaded CSV Files.*

1e. Click *Proceed to Analysis.*

2. In *Similarity Analysis* we will clean up duplicate and redundant elements. We perform this in two steps, first we process categories, and then characteristics.
Calculating embeddings transform the words in the taxonomy into mathematical vectors that can be plotted. Elements that are similar semantically will make groups called "clusters". These clusters are detected by an algorithm called HDBSCAN, the parameters of which can be edited in the *Adjust clustering parameters* box.

2a. Select Categories Only and click on Calculate Embeddings.

2b. Use the *Interactive Heatmap*, *Similarity Matrix*, and *3D Visualization* to see overlapping terms in the taxonomy.

2c. Do you agree with the results of the automatic clustering that you can see in the *Automatically Detected Clusters*?

3. In the original taxonomy there are some categories that are redundant. Identify them through the *interactive heatmap* without using the LLM,

3a. Modify the taxonomy categories so the redundant taxonomy categories are no longer redundant (click on *Select Cluster* to start). Here you may delete one of the characteristics that is a synonym or rename one of them. You decide.

4. In the original taxonomy there are some characteristics that are synonyms and thus redundant.

4a. Try to identify these characteristics yourself through the *interactive heatmap*.

4b. Use the LLM's suggestions to solve the redundancy problem (click on *Get LLM Suggestions* and they will appear below the page). Use your own intuition to either accept or reject the LLM's reasoning. In any case, either yourself manually or via the use of the LLM, resolve the problem with redundant characteristics (e.g. remove, change name, etc).

Once you have identified and fixed all duplicate and redundant elements, we have now a series of loose categories and characteristics that will be used in the next agent to construct the new structure.

4c. Click on *Continue to Review* to see a summary of the changes this first agent has done to the original data, then proceed to Agent 1 by clicking *Transfer Data to Agent 1*.

Agent 1:

5. Some steps need to be performed to prepare the data that we have corrected. We have to calculate embeddings, cluster them again.

5a. In *Step 1: Cluster Categories*, click on *Compute Embeddings* and click on *Update Clustering Parameters and Run* to cluster all categories.

5b. Click on the button *Next: Cluster Characteristics and* click on *Update Clustering Parameters and Run* to now cluster all characteristics.

5c. Use the *2D* and *3D visualization* again to see the output generated. When finished click on *Next: Setup LLM & RAG*.

6. We have to set up the LLM again.

6a. In *Step 3: Setup LLM & RAG* click on *connect to LM Studio* and select the LLM we have used before and click on *Set Model*.

6b. Under *RAG Configuration* upload the original taxonomy CSV files. Upload *taxonomy2.csv* and click *Add Taxonomy*, when done do the same for *taxonomy2_1.csv*.

6c. When the files have been added, click on *Next: Build Taxonomy*.

7. In the *Interactive Taxonomy Builder* use the results of the similarity analysis which you will find in the *Unassigned Categories* to assemble the new taxonomy (click on *add to root* to add directly to the root or click on select and the panels below to place manually). Use your general knowledge, and the AI assistant.

7a. Start with the Categories. One by one, place them where you think best belong. If in doubt, ask the AI for suggestions. Start by using the *Manual Placement* box.

7b. When you have placed all Categories, do the same for all Characteristics. If in doubt, ask the AI for suggestions.

7c. Add the following new characteristics: Add the following new characteristics: "Quantum Computer", "Roomba", and "TPU".

If you are missing a category or characteristic, you may have accidentally deleted it, you can always add it back manually using the *"Add New Element"* panel.

8. Under *Advanced Options* ask the LLM for general advice on the structure, and when finished click on *Transfer Data to Agent 0*.

Agent 0:

9. Use the visualization tool to see the constructed taxonomy. Play with the different visualization options. Do you see any issues with the structure of the taxonomy?
10. Calculate the *Robustness* and *Conciseness* of the taxonomy using the buttons that are provided.
    - 10a. What is the Robustness value?
    - 10b. How many intruders do exist?
    - 10c. What is the conciseness value?
    - 10d. At which level the conciseness gets affected the most?
11. Take a moment to inspect the obtained results, then generate an *Automatic LLM Review*. Does the LLM display results that match those you obtained in step 10 above?
12. Use the *Chat* function to talk to the LLM. Expose your opinion on the constructed taxonomy. What does the LLM say? There is no right or wrong opinion here. The goal is to engage with the LLM and to see how helpful It can be.

Take a print-screen of the constructed taxonomy so you can discuss it with your team

Taxonomy 2

- Hardware
  - CPUs
  - GPUs
  - Smartphones
  - Tablets
  - Storage Systems
- AI Systems
  - Machine Learning
  - Computer Vision
  - Large Language Models
- Robotics
  - Industrial Robots* (Redundant)
  - Manufacturing Robots* (Redundant)
  - Autonomous Vehicles
  - Kitchen Appliances* (Unrelated to Robotics)

Taxonomy 2.1

Software Development (root)

- Development Platforms
  - Web Development
  - Mobile Development
  - Desktop Applications
  - Game Development
- Programming Languages (DUPLICATE #1)
  - Python
  - JavaScript
  - JS* (synonym for JavaScript)
  - Java
    - HTML* (appears in both Programming Languages)
- Programming Languages (DUPLICATE #2)
  - HTML (appears in both Programming Languages)
  - JS (appears in both Programming Languages)
- Software Architecture
  - Databases
  - APIs
  - Cloud Services
  - Microservices