

EcoValid

Final Requirements, Design, Implementation/Testing &
Installation/Delivery

Dr. DK Xu, Dr. Kai Lan

Cole Sanders
Annabella Jackmore
Sid Kaju
Esha Mascarenhas
Madison Torres

North Carolina State University
Department of Computer Science

4/30/2025

Executive Summary

Author: Annabella Jackmore

Reviewer/Editor: Madison Torres

Dr. DK Xu, an Assistant Professor in the CS Department at NC State, and Dr. Kai Lan, an Assistant Professor of Sustainability Science and Engineering in the Department of Forest Biomaterials, are collaborating on developing an intelligent validation platform for environmental data. Their research focuses on using AI-driven solutions to enhance environmental data management, improving large-scale data validation and analysis efficiency and accuracy. Improving this process could accelerate research output, enable better-informed policy decisions, and strengthen sustainability initiatives.

Environmental scientists currently rely on OpenLCA, a life cycle assessment (LCA) software, to manually calculate the environmental impact of individual processes. This approach is time-consuming, inefficient, and prone to human error. This requires significant effort to analyze multiple scenarios, limiting the ability to quickly and accurately extract insights from large datasets. There is no efficient, user-friendly method for querying environmental impact data across different industries and processes without running individual calculations. Dr. Xu and Dr. Lan seek a scalable AI-powered solution that streamlines data retrieval, analysis, and validation for scientists working in sustainability and environmental research.

Our team has developed an AI-powered chatbot that enables scientists to query environmental impact data using natural language rather than manually running calculations in OpenLCA. The chatbot leverages LlamaIndex to convert natural language questions into optimized queries, retrieving relevant results from a vectorized environmental database that stores precomputed environmental impact data. The system significantly reduces the time required to obtain insights, making environmental impact assessments more efficient, accessible, and scalable for all users.

We have completed the requirements section and validated our understanding of project goals with sponsors. Additionally, we have finished the high-level design architecture diagram and GUI design for the chatbot. In terms of implementation, we have integrated our GUI design with the backend RAG (Retrieval-Augmented Generation) functionality. We have implemented user data validation, creating reports and summaries of the data, and customizing responses based on roles. The system responds to most environmental queries with no issues.

For testing, we have finished our system test plan with our project goals but are still running into some inaccuracy due to a parameter issue. We currently have 12/20 tests passing, with others very close to the response we desire from the chatbot. We have all 21 of our unit tests passing and have conducted usability testing with 4 participants.

Project Description

Authors: Madison Torres, Esha Mascarenhas, Annabella Jackmore

Reviewers/Editors: Madison Torres, Annabella Jackmore, Cole Sanders

Sponsor Background

Dr. DK Xu, Assistant Professor of Computer Science, Campus Box 8206, Engineering Bldg II, 890 Oval Dr Room 3320, Raleigh, NC 27606, dxu27@ncsu.edu

Dongkuan (DK) Xu is an Assistant Professor in the CS Department at NC State and leads the NCSU Generative Intelligent Computing Lab. His research is fundamentally grounded in advancing Artificial General Intelligence, particularly automated planning, reliable reasoning, and efficient computing. He has been honored with the Microsoft Accelerating Foundation Models Research Award 2024, the NCSU Carla Savage Award 2024, and the Best Paper Award of ICCCN 2023. DK's research has been published repeatedly in top conferences and journals in AI, NLP, CV, Hardware, and other fields. He serves as the Column Editor for the ACM SIGAI Newsletter and chaired the Workshop on Deep Learning-Hardware Co-Design for Generative AI Acceleration at DAC 2024, co-chaired the Workshop on Dataset Distillation for Computer Vision at CVPR 2024, the Workshop on Deep Learning-Hardware Co-Design for AI Acceleration at AAAI 2023, and the Resource Efficient Learning for Knowledge Discovery Workshops at KDD 2023 and KDD 2024. In addition, Dr. Xu has served as the panelist, area chair, (senior) PC member, and regular reviewer for NSF programs and over forty major conferences and journals. Dr. Xu has extensive research experience in the industry. He has been collaborating with Google DeepMind, Microsoft Research, Moffett AI, and NEC Labs America, and holds multiple US patents. Dr. Xu's long-term research goal is to liberate AI productivity and democratize its application to serve a broader range of populations and real-world applications, equally, sustainably, and responsibly. More information can be found at <https://dongkuanx27.github.io/>

Dr. Kai Lan, Assistant Professor of Sustainability Science and Engineering, Biltmore Hall, 2820 Faucette Dr, Raleigh, NC 27607, klan2@ncsu.edu

Dr. Kai Lan is an Assistant Professor of Sustainability Science and Engineering at the Department of Forest Biomaterials at North Carolina State University. He received his B.S. in Mechanical Engineering from Shanghai Jiao Tong University, China in 2014, M.S. in Mechanical Engineering from University of Michigan in 2016, and Ph.D. in Forest Biomaterials from North Carolina State University in 2020. Prior to joining NCSU, he was a postdoctoral associate at Yale University (2021-2023). Dr. Lan leads the Sustainability Analysis and Innovation Lab (SAIL). SAIL focuses on developing innovative methods and integrated modeling frameworks to assess and advance our industrial systems and the built environment

related to sustainable bioenergy and biomaterials. SAIL uses interdisciplinary approaches, including industrial ecology, sustainable engineering, and data analysis, to address sustainability issues, especially in the context of climate change mitigation and circular economy. Dr. Lan has published 30+ peer-reviewed articles reported in leading journals such as Science, Joule, PNAS, One Earth, Chemical Reviews, EST., and others.

Problem Description

Currently, environmental scientists rely on OpenLCA, a life cycle assessment (LCA) software, to manually calculate the environmental impact of individual processes. This approach is slow, labor-intensive, and prone to human error, limiting the ability to scale analysis for complex scenarios. Large datasets that are being used in the database ecoInvent, hosted on the OpenLCA software may take days or weeks to analyze manually and manual calculations introduce inconsistencies and potential data loss. Users need a scalable solution to process and validate environmental data in real time, providing fast and reliable insights to support research and policy decisions. Additionally, the current manual approach creates a barrier for non-technical users such as policymakers. These users run into challenges related to data accessibility and understanding complex environmental datasets without technical training or support.

EcoValid is an innovative project that addresses the critical challenge of validating and analyzing vast amounts of environmental data. To overcome previous limitations, EcoValid leverages cutting-edge artificial intelligence technologies, specifically retrieval-augmented generation (RAG) and large language models (LLMs), to create an efficient, accurate, and scalable system for environmental data validation.

Proposed Solution & Project Goals/Benefits

Our team has developed an AI-powered chatbot that enables environmental scientists to efficiently query environmental impact data, rather than manually running calculations in OpenLCA. The chatbot integrates a vectorized database containing structured data on environmental effects of mining, energy generation, and other industrial activities, indexed for rapid retrieval. Using LlamaIndex, the system translates natural language queries into optimized queries to vector the DB, ensuring accurate and relevant responses.

The proposed solution has been developed as a web-based application, allowing users to interact with a comprehensive database of environmental processes, inputs, and outputs. This system will enable automated data validation, provide advanced insights, and offer scalability far beyond current manual methods. Users are able to query and validate environmental data such as greenhouse gas emissions or carbon footprints, as well as perform complex validations like multi-step life cycle assessments.

By implementing this AI-driven approach, EcoValid significantly reduces validation time from days or weeks to minutes or hours, potentially decreases error rates by 50% or more compared to manual methods, and achieves near 100% consistency in applying validation criteria. The system's ability to handle hundreds or thousands of validation queries simultaneously will greatly enhance the capacity for environmental data analysis and decision-making. This project represents a transformative step in environmental data management, promising to accelerate research, improve policy-making, and ultimately contribute to more effective environmental stewardship.

Major Goals

1. Enhance Data Accessibility – Provided scientists with an intuitive chat interface to retrieve complex environmental data without requiring SQL expertise.
2. Optimize Retrieval Speed – Utilized LlamaIndex and efficient indexing techniques to reduce query response time and enhance user experience.
3. Increase User Usability – Delivered a user-friendly interface tailored to scientists' workflows, making data retrieval faster and easier than manual database searches.
4. Support Decision-Making – Enabled scientists to analyze environmental impact trends and make data-driven recommendations.

Deliverables

- A fully functional chatbot that interprets user queries, translates them into vector queries, and retrieves relevant environmental impact data.
- A vectorized database with environmental impact data, structured for fast, accurate retrieval.
- A user interface for seamless interaction with the chatbot.
- Documentation and guidelines for system usage, ensuring adoption and scalability.

Project Completion Goals

By the end of the semester, we have:

- Developed and deployed a working chatbot capable of accurately retrieving environmental data based on user queries.
- Allowed users to upload data in chat format to the chatbot for validation.
- Implemented chatbot functionality to create detailed reports of environmental information.
- Implemented chatbot functionality to create validation summaries based on user uploaded information.
- Customized chatbot responses based on user roles.
- Implemented LlamaIndex-based indexing to optimize data retrieval speed.

- Conducted usability testing to refine system performance.
- Provided comprehensive documentation for future enhancements.

Benefits to Stakeholders

- **Sponsor Benefits:** The chatbot enhances data accessibility and streamlines research workflows, allowing scientists to focus on analysis rather than data extraction. This also allows non-technical individuals to access the environmental data without barriers due to knowledge of software or data interpretation challenges.
- **User Benefits:** Environmental scientists gain instant access to relevant data, reducing manual query time and improving research efficiency. Non-technical users have access to an abundant amount of environmental data to help with different decisions such as policy making.
- **Community Benefits:** By accelerating environmental research, the system contributes to informed policy-making, sustainable energy solutions, and effective environmental regulation.

This project directly addresses the problem of inefficient data access in environmental science, delivering a tool that improves accuracy, efficiency, and decision-making for researchers, policymakers, and other possible individuals.

Resources Needed

Author: Annabella Jackmore

Reviewer/Editor: Madison Torres, Cole Sanders

Table 1: Resources Required for EcoValid

Resource Name	Purpose	Status	Version	Licensing Information
openLCA	Software tool to store environmental information	Have access	2.4.0	MPL 2.0
Streamlit	Framework to use for frontend Chatbot implementation	Have access	1.41.0	Apache License 2.0
OpenAI	Using OpenAI	Have access	1.60.2	MIT License

	key to use for our Chatbot implementation			
Lllamaindex	Used to vectorize the data and perform vector search	Have access	0.12.20	MIT License
FastAPI	Used as an API framework to make POST API calls from Chatbot.py to Server.py.	Have access	0.115.12	MIT License

Table 1 contains details about the software resources needed to implement EcoValid.

Risks & Risk Mitigation

Author: Esha Mascarenhas

Reviewer/Editor(s): Madison Torres, Annabella Jackmore

One of the key risks in our system is **incorrect embeddings and vector retrieval**, which can lead to irrelevant or inaccurate responses. This issue arises due to poorly tuned embeddings or inefficiencies in the vector search process using FAISS or ChromaDB. This can lead to compromised quality of insights provided to users and reduce trust in the system. To mitigate this, we combine **vector search with traditional keyword-based retrieval**, ensuring that if embeddings fail to retrieve the most relevant results, keyword matching provides a fallback for improved accuracy. Another critical risk is **the exposure of API keys for OpenAI or other external services**, since our team relies on OpenAI and other external APIs, if these keys are exposed it could lead to security vulnerabilities and unauthorized access. To prevent this, we **store API keys securely using .env files and environment variables**, ensuring they are never hardcoded in the codebase or exposed in public repositories. These mitigation strategies enhance both the accuracy and security of our system, ensuring a more robust and reliable implementation.

Development Methodology

Author: Esha Mascarenhas

Reviewer/Editor(s): Madison Torres, Annabella Jackmore

Our team follows a structured iterative development process that allows for continuous feedback, incremental improvements, and systematic testing. This approach is particularly suitable for EcoValid due to the complexity of integrating AI-based query processing with structured environmental datasets, which requires careful refinement and validation at each stage. Our team's iterations were decided at the beginning of the semester and agreed upon with our sponsors to follow the specified plan. Our sponsors are actively involved throughout the development cycle, monitoring progress and giving feedback, and agree with our iterative development approach.

Development Process

In our iterative development cycle approach, we focus on building and refining the chatbot. Each iteration lasts approximately 2-4 weeks and includes the following key activities:

- Planning and Design - At the start of each iteration, project goals are defined and looked over and expected outcomes are based on sponsor feedback and our progress as a team.
- Implementation - Code is written following best practices that we learned in previous classes and from our own experience, including consistent formatting, modular design, and clear documentation.
- Testing - We look at how to apply different testing to verify the correctness of new features and prevent unintended breaks.
- Review and Feedback - Completed features are demonstrated to the sponsors for evaluation and feedback is incorporated in each iteration.

Approach Reasoning

This approach is ideal for this project and best for our team due to a multitude of reasons. First, AI-based query processing involves complex machine learning models that require fine-tuning and gradual improvement. Also, retrieval and validation of environmental data is based on real-time adjustments to the embeddings and search algorithms. Continuous feedback from our sponsors is key to ensuring that the chatbot meets the functional and usability requirements. This feedback is always changing, making iterative improvements the best to handle this. This method allows our team to address technical challenges incrementally and reduce the risk of system failure due to poor embeddings or data retrieval issues.

Development Practices

- Version Control
 - We use Github for version control, creating feature branches and pull requests for all changes.
 - Code is reviewed through Github by a different person before merging to ensure consistency and catch potential bugs.

- Task Management
 - Our team works to discuss what tasks we are working on at the start of each class time to transfer knowledge and see where we can offer assistance. This helps for check-ins and sponsor meetings to ensure we are all on the same page.
- Meetings and Communication
 - Our team hosts a check-in with the teaching staff every Wednesday to discuss progress, blockers, and next steps. We also discuss this amongst ourselves at the beginning of every class.
 - We meet with the sponsors every Thursday to demonstrate completed features, ask questions, and collect feedback. The facilitator of each meeting gives a progress report on the team's behalf to the sponsors at the beginning of each sponsor meeting.
- Testing
 - Our team follows a test-driven development where we go through usability testing as each feature is implemented.
 - We follow these features by adding in tests to our System Test Plan and writing unit tests alongside.

System Requirements

Authors: Cole Sanders, Annabella Jackmore

Reviewer/Editor: Madison Torres

Overall View

EcoValid is a web application for querying and validating environmental data. Environmental scientists and policymakers often struggle with the complexity of analyzing large datasets manually using existing life cycle assessment tools like OpenLCA. This manual process is time consuming and prone to error, making it difficult to retrieve accurate insights and validate environmental impact data. By combining advanced tools like RAGs and LLMs, we are enabling technical users such as researchers and non-technical users such as policymakers to easily and efficiently access detailed and accurate environmental impact datasets. This data consists of a list of processes such as combusting hard coal in China or growing rice in the US and their environmental impacts as measured by specific metrics such as smog production, global warming potential, eutrophication, etc.

The components of the environmental data used to build the dataset mentioned above include precompiled process data in JSON format. The user can input their own data for validation by typing in the data they would like to have checked.

The use cases of the project include validation of data against trusted sources, creating validation reports to help with user understanding and decision-making, retrieving information from

complex queries, and optional visual tools to help users understand the results such as maps, flow charts, and flow diagrams. An example of a query would be a user asking the chatbot what the environmental impact of producing 1 kWh of electricity in Costa Rica is. An example of validating the data would be the user manually inputting the environmental impacts of a process and asking the chatbot to validate it. The chatbot would then cross-check that information with the information in the database and produce a report. Validation reports consist of an overview of the data to check that everything is accurate and there are no inconsistencies.

EcoValid allows users to choose a role that will customize their interaction with the application. For example, if the user has selected a policymaker as their role, their answer will contain an explanation of any technical terms in the data as well as the units used. Additionally, it will provide context to describe why this data is important for policymakers. However, if they select a researcher role, then the answer will contain a brief and to the point response using technical terms and units with no further explanation. It also has a secure back-end that handles the query processing and validation through the LLMs. Lastly, we have implemented a scalable database that is linked to trusted environmental datasets and included an export option with the validation results created for users to share this data outside the application. Users can access all the functionality of the project by simply interacting with a chatbot and asking for what they need. The users can even open multiple chats and delete chats if they wish to organize their chats for different uses. Our prototype will streamline the process of automating querying and validating environmental data, improving its accessibility for a multitude of users.

When a user accesses the system, they can ask about the environmental impact of different processes and the results will be returned to them through the chatbot. They can also enter their datasets that the system can validate against the data that already exists in the data provided by Dr. Lan and Dr. Xu.

Glossary

LLM - A large language model (LLM) is a type of artificial intelligence (AI) that excels at processing, understanding, and generating human language. LLMs are useful for analyzing, summarizing, and creating content across many industries. Published on July 1, 2024. [1]

RAG - Retrieval-augmented generation (RAG) is an AI framework for improving the quality of LLM-generated responses by grounding the model on external sources of knowledge to supplement the LLM's internal representation of information. [2]

JSON - JSON (JavaScript Object Notation) is a text-based, human-readable data interchange format used to exchange data between web clients and web servers. [3]

LCA - LCA measures the human health and the environment impacts associated with a product. It examines each stage in the life cycle of a product, including raw material extraction, manufacturing, transportation, use and maintenance, and disposal. [4]

- [1] “What Is a Large Language Model (LLM)?” *SAP*, 1 July 2024, [www.sap.com/resources/what-is-large-language-model#:~:text=A%20large%20language%20model%20\(LLM\)%20is%20a%20type%20of%20artificial,Published%20on%20July%201%2C%202024.](https://www.sap.com/resources/what-is-large-language-model#:~:text=A%20large%20language%20model%20(LLM)%20is%20a%20type%20of%20artificial,Published%20on%20July%201%2C%202024.)
- [2] Martineau, Kim. “What Is Retrieval-Augmented Generation (Rag)?” *IBM Research*, IBM, 22 Aug. 2023, [research.ibm.com/blog/retrieval-augmented-generation-RAG#:~:text=Retrieval%2Daugmented%20generation%20\(RAG\),LLM's%20internal%20representation%20of%20information.](https://research.ibm.com/blog/retrieval-augmented-generation-RAG#:~:text=Retrieval%2Daugmented%20generation%20(RAG),LLM's%20internal%20representation%20of%20information.)
- [3] Gillis, Alexander S. “What Is JSON? - Definition from Techtarget.Com.” *TheServerSide.Com*, TheServerSide.com, 26 Aug. 2022, www.theserverside.com/definition/JSON-Javascript-Object-Notation.

Functional Requirements

FR1 System Interaction

- FR 1.1. The system shall answer user queries about the environmental data written in English about the data. (Some examples: What is the environmental impact of producing 1 kWh of electricity in Costa Rica? What is the global warming potential of generating electricity in China by combusting hard coal? What is the average acidification potential of generating 1 kWh globally at medium voltage?)
- FR 1.2. The system shall be able to perform automated validation checks against predefined criteria and trusted sources. (Trusted sources include the data provided by a domain expert Dr. Lan)
- FR 1.3. The system shall generate validation summaries. (A list of all environmental impacts entered by the user along with the corresponding environmental impact data found in the database)
- FR 1.4. The system shall generate detailed reports. (A list of all the environmental impacts associated with a process at a given location)
- FR 1.5. The system shall add context to the answer to the question the user asks. (Explain the units of measurement, the location of the process, and the name of the process)
- FR 1.6 The system shall document the thinking trajectory for the thinking process of the Chatbot and display this trajectory to the user.
- FR 1.7 The system shall have the option for the user to export chat history in a PDF.
- FR 1.8 The system shall provide visualization reports.

FR2 Roles

- FR 2.1. Each user shall be able to choose one of the following roles: Researcher or Policymaker.
- FR 2.2. The chat shall refresh if the user switches roles in the middle of a session

- FR 2.3 The chatbot shall answer users in the Researcher role with domain specific technical vocabulary without further explanation. (Example Response: “The GWP for the process named "mango production | mango | Cutoff, U" in the location RoW (Rest of the World) is 0.1274592400374663 kg CO2 eq per kilogram of mango produced.”)
- FR 2.4 The chatbot shall answer users in the policymaker role. In doing so, it shall explain any technical terms, explain the units used in the data, and add context to describe why the result is important to policymakers. (Example Response Snippet: “These values reflect the amount of carbon dioxide released during mango production, reported as kilograms of CO2 equivalent, which is the standard unit for measuring greenhouse gases' impact on global warming. This data can help policymakers evaluate the climate impacts of agricultural activities and make informed decisions about sustainable practices and policies in different regions.”)

Non-Functional Requirements

- **NFR1 Data**
 - NFR 1.1. The system shall ensure data protection regulations and environmental standards.
- **NFR2 User Access**
 - NFR 1.2. The system shall allow multiple users to access the system simultaneously.
- **NFR3 Prompting**
 - NFR 1.3. The system shall recognize when a question needs an answer from our environmental data

Constraints

4.4 Constraints

- C1. The system shall use a LiteLLM API key as that is the key through which Dr. Xu and Dr. Lan are providing funding to run the system.
- C2. The system shall use the data provided by Dr. Xu and Dr. Lan to build the queries because it has been validated.
- C3. The system shall expect user queries in English since the data is stored in English.

Design

Author(s): , Cole Sanders, Esha Mascarenhas

Reviewer(s)/Editor(s): Sid Kaju

High-Level Design

The **EcoValid** chatbot follows a structured pipeline to process user queries and validate environmental data (see Figure 1). The frontend, implemented in [Chatbot.py](#), uses Streamlit for the chatbot UI and CSS for custom styling. Users interact with this interface by typing a query,

which is then sent to the LLM to determine whether it is an environmental, an environmental data validation, or general query. If the query is general, the LLM generates a response directly. If it is an environmental query or validation query, the message is forwarded to the backend via an API request handled by `Server.py`, which uses FastAPI, a Python-based API framework. The backend then calls `Embeddings.py` to query a vector database using LlamaIndex, a semantic search that retrieves the top 10 most relevant process documents. These results are returned to `Chatbot.py`. `Chatbot.py` sends off the results to the LLM to select the best option which is returned to `Chatbot.py`. `Chatbot.py` then matches that process with detailed data from `TextData.txt`. The selected process information is then translated into human-readable text by the LLM and delivered back to the user. The semantic search and answer refinement process are used to ensure accurate responses and fulfill FR 1.1 and FR 1.2. Context is added by the LLM to match responses to user roles and fulfill FR 2.3 and FR 2.4.

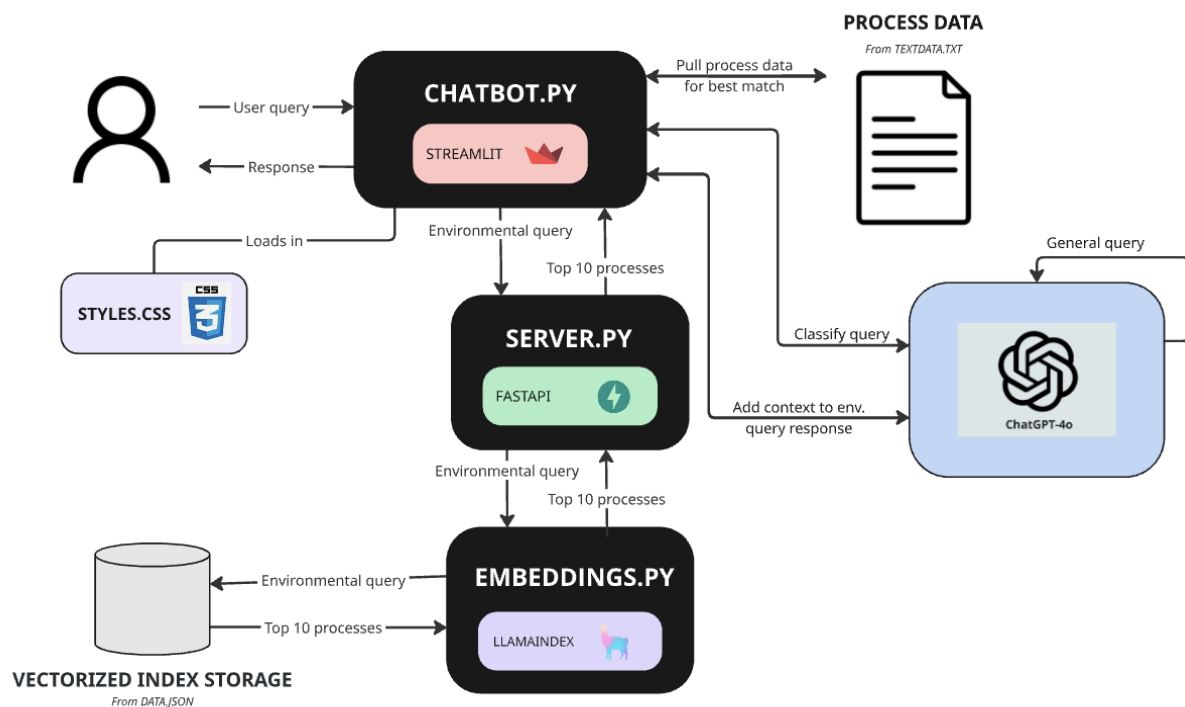


Figure 1: High-Level Architecture

Low-Level Design

Dataflow

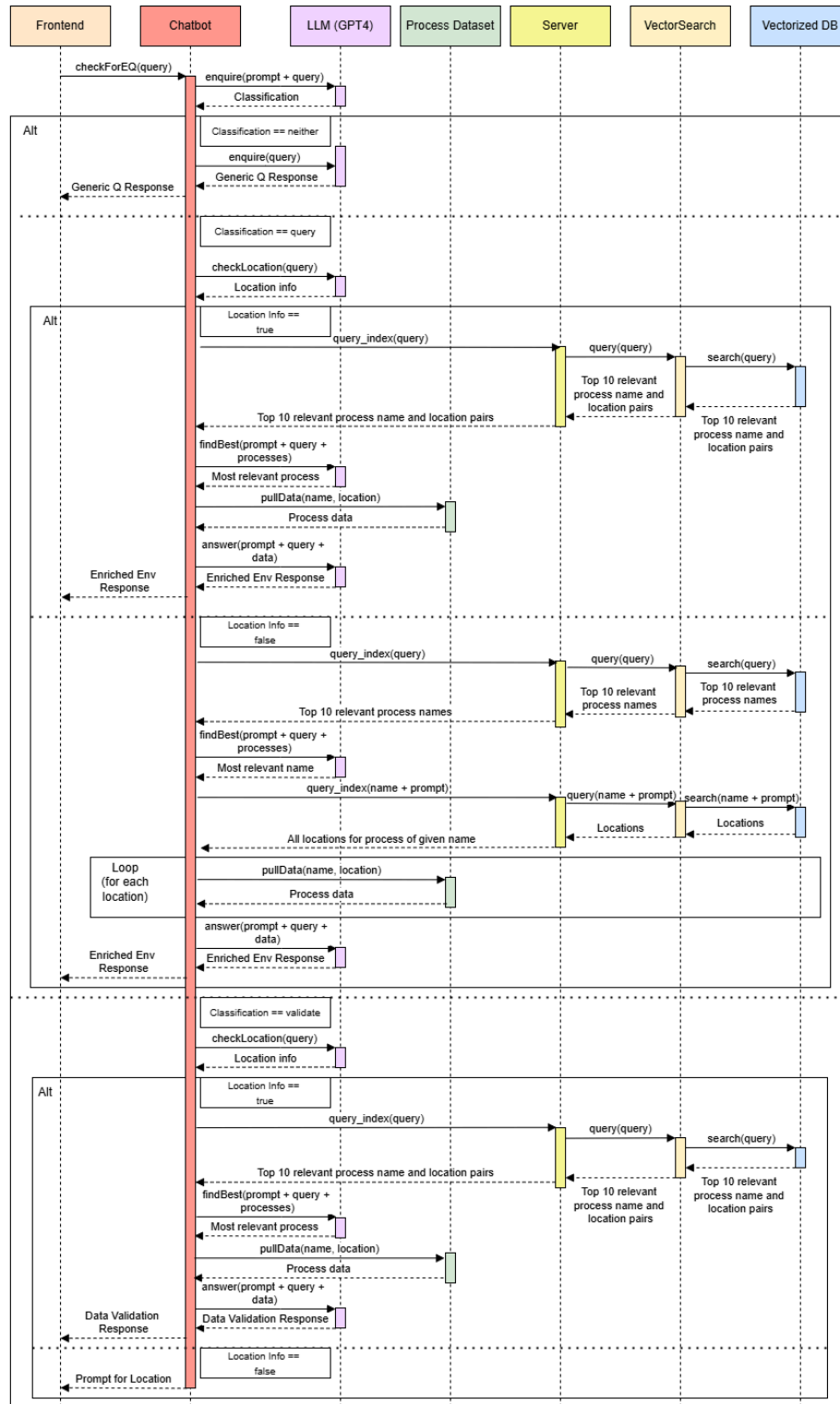


Figure 2: Sequence Diagram of Main Dataflows

A user query is input through the frontend. This is retrieved by Chatbot.py and converted into a string. This string along with a prompt asking to classify the user input is sent to the LLM for a

response. A response is returned in the form of a string as “query”, “validate”, or “neither”. If the response from the LLM is “neither”, the query is a non-environmental input and is sent to the LLM for a generic response which is returned to the frontend.

If the response is “query” then it is an environmental data query. The user input is then sent to the LLM again to ask if there is a location specified. If there is, the user query is sent to the Server.py REST API in JSON format which then calls the vector search function in Embedding.py passing in a string. Embedding.py encodes the string, searches the vector database for similar vectors, decodes the similar vectors into a string, and returns that string to Server.py. Server.py then returns the response to Chatbot.py’s API call in JSON format including the string returned from Embedding.py. This string contains a list of the top ten most relevant process and location pairs as determined by vector search. This string is then sent to the LLM along with the original user query and a prompt asking the LLM to choose the best fit process and location pair to answer the question. The answer will be returned in a string, and that string will be used to search the process dataset in txt format to find the process with the matching name and location to extract its data. Once the data is extracted as a string, the original user query, the extracted data, and a prompt will be sent to the LLM and the final answer will be returned as a string and output to the user. If there is no location specified, a similar process occurs. However, instead of searching for process name location pairs in the vector database, only process names are searched for. Once the best fit name is determined by the LLM, a second prompt is sent off to the vector database asking for all the locations associated with the best fit process name, using all the same data formats as mentioned above describing the first call to the vector database. Once all the locations are returned, the data for that process at each location is extracted from the txt process dataset. The data is combined, and sent off to the LLM along with the original user query, and a prompt. The answer returned by the LLM is output to the user.

If the response is “validate” then it is an environmental data validation question. The dataflows and formats have already been described above, however the implementation for validation questions uses different prompts. The user input is then sent to the LLM to check if a location is specified. If so, the user query is sent to the Server.py REST API in JSON format which calls vector search in Embedding.py passing in a string. Embedding.py encodes the string, searches the vector database for similar vectors, decodes the similar vectors into strings, and returns them to Server.py. Server.py then returns the response to Chatbot.py’s API call in JSON format including the strings returned from Embedding.py. This string contains a list of the top ten most relevant process and location pairs as determined by vector search. This string is then sent to the LLM with the original user query and a prompt asking the LLM to choose the best fit. The answer is returned as a string, and that string will be used to search the process dataset in txt format to find the process with the matching name and location to extract its data. Once the data is extracted as a string, the original user query, the extracted data, and a prompt will be sent to the LLM and the final answer will be returned as a string and output to the user. If there is no

specified location, a string prompting the user to ask their validation question with a specific location will be output to the user.

APIs

Function: Chatbot.getProcesses(prompt:String)

Route: POST http://localhost:8000/namequery

Data details: Current user environmental query to return top ten matching process name and location pairs or just names

```
{'query': 'what is the carbon intensity of landfilling 1 kg mixed waste plastics in Switzerland?'}
```

<Response [200]> for OK and the answer to the query should be returned. The answer payload to the above query is below:

```
{'response': 'Processes: Process name: treatment of waste plastic, mixture, sanitary landfill | waste plastic, mixture. Location: CH Process name: market for waste plastic, mixture | waste plastic, mixture. Location: CH Process name: treatment of waste plastic, mixture, sanitary landfill | electricity, medium voltage. Location: CH Process name: treatment of waste polystyrene, sanitary landfill | waste polystyrene. Location: CH Process name: market for residues, MSWI[F]-WWT-SLF, waste plastic, mixture | residues, MSWI[F]-WWT-SLF, waste plastic, mixture. Location: CH Process name: market for residues, MSWI[F], waste plastic, mixture | residues, MSWI[F], waste plastic, mixture. Location: CH Process name: treatment of decarbonising waste, residual material landfill | decarbonising waste. Location: CH Process name: market for waste plastic, mixture | waste plastic, mixture. Location: FR Process name: market for residues, MSWI[F]-WWT-SLF, waste plastic plaster | residues, MSWI[F]-WWT-SLF, waste plastic plaster. Location: CH Process name: treatment of waste polystyrene, sanitary landfill | electricity, medium voltage. Location: CH'}
```

Function: Chatbot.getLocations(prompt:String)

Route: POST http://localhost:8000/locationquery

Data details: Best fit process name plus prompt engineering to return all locations for given process.

{'query': 'Provide all the locations for each instance of the following process. List them in the format "Location1,Location2,etc." Process name: market for inorganic phosphorus fertiliser, as P2O5 | inorganic phosphorus fertiliser, as P2O5.'}

<Response [200]> for OK and the answer to the query should be returned. The answer payload to the above query is below:

{'response': 'BA,IN,US,RoW.'}

Classes

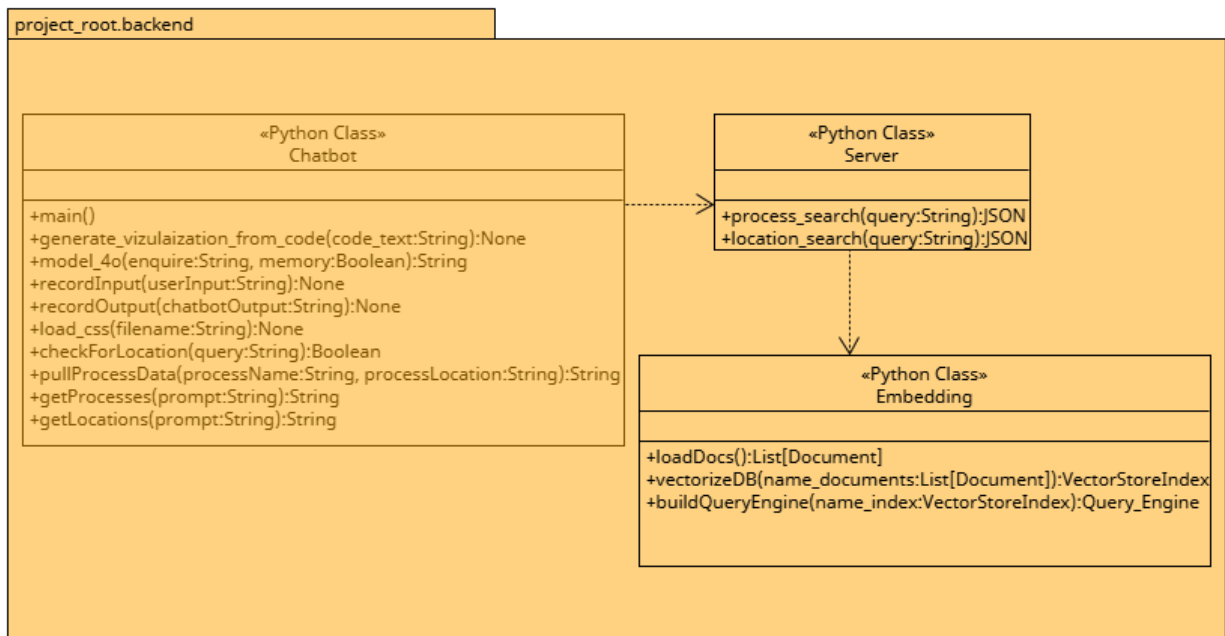


Figure 3: Class Diagram

Chatbot.py:

This is the main body of code. It uses the Streamlit library to create the chatbot frontend, update the frontend, retrieve user queries from the frontend, push answers to the frontend, and store chat history. From here calls are made to the LLM using a LiteLLM API key and the LiteLLM library. Calls are also made to Server.py using the Requests library. It also has the functionality to load the process dataset txt file and extract data.

Embedding.py:

This class uses the LLamaIndex library to vectorize data stored in JSON format and to perform vector search on the databases it has vectorized.

Server.py

This class creates a REST API using FastAPI to handle user queries by calling vector search from Embedding.py.

DataCleaner.py

This class converts environmental process data from a JSON file into a readable text format. It extracts key fields like process name, ID, location, impact metrics, and output values, then writes each entry as a formatted sentence to a text file.

DataExtraction.py

This class enriches a master environmental dataset by pulling output information from individual process files. It adds standardized output details, such as amount and unit, to each process entry, preparing the data for more consistent use in analysis or display.

NameExtractor.py

This class extracts simplified process names and locations from a formatted text file. It outputs a cleaner summary of each process to a new file, making the data easier to browse or use for labeling and role-based filtering.

Data

The data used in this process is a series of processes and information about them in two JSON formats. The dataset consists of over 23,000 entries. The data is structured, listing a process such as grape production or transporting cargo by rail and the associated environment impacts of that process as measured by fossil fuel depletion, respiratory effects, global warming, carcinogenics, smog, acidification, non-carcinogenics, ozone depletion, eutrophication, and excitotoxicity. The units of all the impact per unit of process is also recorded. For example, geothermal power production has a global warming potential of 10 MJ surplus per kWh. An example entry is as follows:

Process in format 1:

```
{
  "name": "market for leachate, SLF, waste glass | leachate, SLF, waste glass | Cutoff, U",
  "id": "00ca40a7-f952-3e81-aca8-97a3d71a1ffc",
  "location": "GLO",
  "impact_results": {
    "Fossil fuel depletion": {
      "Amount": 0.2591615617351554,
      "Unit": "MJ surplus"
    }
  },
}
```

```

"Respiratory effects": {
  "Amount": 0.00028351683127839,
  "Unit": "kg PM2.5 eq"
},
"Global warming": {
  "Amount": 0.24882826264673022,
  "Unit": "kg CO2 eq"
},
"Carcinogenics": {
  "Amount": 3.3864243791365285e-07,
  "Unit": "CTUh"
},
"Smog": {
  "Amount": 0.015025065883468525,
  "Unit": "kg O3 eq"
},
"Acidification": {
  "Amount": 0.0009401520887998984,
  "Unit": "kg SO2 eq"
},
"Non carcinogenics": {
  "Amount": 1.040197603143794e-07,
  "Unit": "CTUh"
},
"Ozone depletion": {
  "Amount": 3.517171040651846e-09,
  "Unit": "kg CFC-11 eq"
},
"Eutrophication": {
  "Amount": 0.0016290990589930396,
  "Unit": "kg N eq"
},
"Ecotoxicity": {
  "Amount": 8.468621418917154,
  "Unit": "CTUe"
}
}

```

Process in format 2:

```

{
  "@type": "Process",
  "@id": "00ca40a7-f952-3e81-aca8-97a3d71a1ffc",
  "name": "market for leachate, SLF, waste glass | leachate, SLF, waste glass | Cutoff, U",
  "description": "ecoQuery:
https://ecoquery.ecoinvent.org/3.10/cutoff/dataset/30335/documentation
This is a market activity, representing a treatment mix. In the case of products needing treatment, market mixes are supplied by the activities treating the product in the geography defined by the market, and they supply the activities needing to treat the product, as they have generated it as a by-product in the Undefined processes (present as a negative input in system models). Transport to the treating facility or losses are also accounted in this type of markets, when relevant. This is the market for leachate, SLF, waste glass in Global geography. Transport from producers to consumers of this product in the geography covered by the market is included.
The activity starts with the input from the treatment processes.
A market for a material for treatment (wastes or recyclables) ends with the provision of treatment, i.e. a mix of treatment providers, to activities that generate the reference product as a waste/by-product.",
  "category": "E:Water supply; sewerage, waste management and remediation activities/37:Sewerage/370:Sewerage/3700:Sewerage",
  "version": "00.00.000",
  "tags": ["ConvertedDataset", "Sector: Waste Treatment \u0026 Recycling"],
  "processType": "UNIT_PROCESS",
  "isInfrastructureProcess": false,
  "location": {"@type": "Location", "@id": "56bca136-90bb-3a77-9abb-7ce558af711e", "name": "Global", "category": "Global"},
  "processDocumentation": {
    "validFrom": "2010-01-01",
    "validUntil": "2023-12-31",
    "geographyDescription": "The disposal mix is country-specific for Global geography.",
    "technologyDescription": "The dataset supplies a disposal mix (technology mix) of leachate, SLF, waste glass to the respective waste producing activities.",
  },
  "dataGenerator": {"@type": "Actor", "@id": "e98986e8-51ee-468d-b4de-00f989a6dc7f", "name": "Nikolia Stoikou"},
  "dataDocumentor": {"@type": "Actor", "@id": "e98986e8-51ee-468d-b4de-00f989a6dc7f", "name": "Nikolia Stoikou"},
  "creationDate": "2023-08-30T11:41:47Z",
  "isCopyrightProtected": true
},
  "lastInternalId": 19,
  "exchangeDqSystem": {"@type": "DQSystem", "@id": "e7ac7cf6-5457-453e-99f9-d889826fffe8", "name": "ecoinvent Data Quality System"},

```

```

"parameters":[],
"exchanges":[
  {
    "@type":"Exchange", "isAvoidedProduct":false, "isInput":false,
    "amount":3.38794222782851E-7, "description":"calculated based on the properties of the
leachate as they are shown in the respective treatment dataset. Globally, 58.73% losses are
considered (Doka G., 2021. A model for composition-specific life cycle inventories of
regionalised wastewater fates, Separate publication, Doka LCA, Zurich, Switzerland.
http://www.doka.ch/publications.htm)",
    "internalId":13,
    "location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Glo
bal","category":"Global"},
    "flow":{"@type":"Flow","@id":"b3ebdcc3-c588-4997-95d2-9785b26b34e1","name":"Lead
II","category":"Elementary flows/Emission to water/surface
water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},
    "unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffcf8f","name":"kg
SWU"},
    "flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","n
ame":"Mass","category":"Technical flow properties","refUnit":"kg"}
  },
  { "@type":"Exchange", "isAvoidedProduct":false, "isInput":false, "amount":1.0,
    "description":"", "internalId":2,
    "location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Glo
bal","category":"Global"},
    "flow":{"@type":"Flow","@id":"cf36d681-a62b-59a3-aa65-d01f1fb84053","name":"leachate,
SLF, waste glass","category":"E:Water supply; sewerage, waste management and remediation
activities/37:Sewerage/370:Sewerage/3700:Sewerage","flowType":"WASTE_FLOW","refUnit":
"m3"},
    "unit":{"@type":"Unit","@id":"1c3a9695-398d-4b1f-b07e-a8715b610f70","name":"m3"},
    "flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-22da-a746-0800200c9a66","n
ame":"Volume","category":"Technical flow properties","refUnit":"m3"},
    "defaultProvider":{"@type":"Process","@id":"8d207689-2391-3174-8959-81a6eb94ccbf","nam
e":"treatment of leachate, SLF, waste glass, wastewater treatment | leachate, SLF, waste glass |
Cutoff, U","category":"E:Water supply; sewerage, waste management and remediation
activities/37:Sewerage/370:Sewerage/3700:Sewerage","processType":"UNIT_PROCESS","flow
Type":"WASTE_FLOW","location":"GLO"}}, {"@type":"Exchange","isAvoidedProduct":false,"
isInput":true,"amount":0.1723,"description":"calculated based on Doka G., 2021. A model for
composition-specific life cycle inventories of regionalised wastewater fates, Separate
publication, Zurich, Switzerland.
http://www.doka.ch/publications.htm","costValue":0.0078029501,"currency":{"@type":"Currenc

```

y", "@id": "7de74e65-ff97-404f-bf57-5d2a9134fb43", "name": "Euro", "refUnit": "EUR"}, {"internalId": 3, "location": {"@type": "Location", "@id": "56bca136-90bb-3a77-9abb-7ce558af711e", "name": "Global", "category": "Global"}, "flow": {"@type": "Flow", "@id": "16fca750-330e-5aba-98f9-c87b3639c934", "name": "sewer grid, rural", "category": "F:Construction/42:Civil engineering/422:Construction of utility projects/4220:Construction of utility projects", "flowType": "PRODUCT_FLOW", "refUnit": "m3"}, "unit": {"@type": "Unit", "@id": "1c3a9695-398d-4b1f-b07e-a8715b610f70", "name": "m3"}, "flowProperty": {"@type": "FlowProperty", "@id": "93a60a56-a3c8-22da-a746-0800200c9a66", "name": "Volume", "category": "Technical flow properties", "refUnit": "m3"}, "defaultProvider": {"@type": "Process", "@id": "7b981ccb-d192-333b-8d44-25dc11a41ede", "name": "market for sewer grid, rural | sewer grid, rural | Cutoff, U", "category": "F:Construction/42:Civil engineering/422:Construction of utility projects/4220:Construction of utility projects", "processType": "UNIT_PROCESS", "flowType": "PRODUCT_FLOW", "location": "GLO"}, {"@type": "Exchange", "isAvoidedProduct": false, "isInput": false, "amount": 2.11236139285149E-6, "description": "calculated based on TOC amount as it is described in Doka G. (2003) Life Cycle Inventories of Waste Treatment Services. ecoinvent report No. 13, Swiss Centre for Life Cycle Inventories, Dübendorf, December 2003. Available at: https://www.doka.ch/13_III_Landfills.pdf", "internalId": 11, "location": {"@type": "Location", "@id": "56bca136-90bb-3a77-9abb-7ce558af711e", "name": "Global", "category": "Global"}, "flow": {"@type": "Flow", "@id": "960c0f37-f34c-4fc1-b77c-22d8b35fd8d5", "name": "DOC, Dissolved Organic Carbon", "category": "Elementary flows/Emission to water/surface water", "flowType": "ELEMENTARY_FLOW", "refUnit": "kg"}, "unit": {"@type": "Unit", "@id": "95dd50e9-c184-4412-9afc-9764b1ffcf8f", "name": "kg SWU"}, "flowProperty": {"@type": "FlowProperty", "@id": "93a60a56-a3c8-11da-a746-0800200b9a66", "name": "Mass", "category": "Technical flow properties", "refUnit": "kg"}, {"@type": "Exchange", "isAvoidedProduct": false, "isInput": false, "amount": 5.14860558035204E-6, "description": "calculated based on the properties of the leachate as they are shown in the respective treatment dataset. Globally, 58.73% losses are considered (Doka G., 2021. A model for composition-specific life cycle inventories of regionalised wastewater fates, Separate publication, Doka LCA, Zurich, Switzerland. <http://www.doka.ch/publications.htm>"), "internalId": 9, "location": {"@type": "Location", "@id": "56bca136-90bb-3a77-9abb-7ce558af711e", "name": "Global", "category": "Global"}, "flow": {"@type": "Flow", "@id": "e34d3da4-a3d5-41be-84b5-458afe32c990", "name": "Chromium III", "category": "Elementary flows/Emission to water/surface water", "flowType": "ELEMENTARY_FLOW", "refUnit": "kg"}, "unit": {"@type": "Unit", "@id": "95dd50e9-c184-4412-9afc-9764b1ffcf8f", "name": "kg SWU"}, "flowProperty": {"@type": "FlowProperty", "@id": "93a60a56-a3c8-11da-a746-0800200b9a66", "name": "Mass", "category": "Technical flow properties", "refUnit": "kg"}, {"@type": "Exchange", "isAvoidedProduct": false, "isInput": false, "am

amount":5.46950716886606E-7,"description":"calculated based on TOC amount as it is described in Doka G. (2003) Life Cycle Inventories of Waste Treatment Services. ecoinvent report No. 13, Swiss Centre for Life Cycle Inventories, Dübendorf, December 2003. Available at: https://www.doka.ch/13_III_Landfills.pdf","internalId":5,"location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Global","category":"Global"},"flow":{"@type":"Flow","@id":"70d467b6-115e-43c5-add2-441de9411348","name":"BOD5, Biological Oxygen Demand","category":"Elementary flows/Emission to water/surface water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},"unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffc8f","name":"kg SWU"},"flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","name":"Mass","category":"Technical flow properties","refUnit":"kg"}},{ "@type":"Exchange","isAvoidedProduct":false,"isInput":false,"amount":5.70134944433439E-8,"description":"calculated based on the properties of the leachate as they are shown in the respective treatment dataset. Globally, 58.73% losses are considered (Doka G., 2021. A model for composition-specific life cycle inventories of regionalised wastewater fates, Separate publication, Doka LCA, Zurich, Switzerland. <http://www.doka.ch/publications.htm>"),"internalId":10,"location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Global","category":"Global"},"flow":{"@type":"Flow","@id":"6d9550e2-e670-44c1-bad8-c0c4975ffc7","name":"Copper ion","category":"Elementary flows/Emission to water/surface water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},"unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffc8f","name":"kg SWU"},"flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","name":"Mass","category":"Technical flow properties","refUnit":"kg"}},{ "@type":"Exchange","isAvoidedProduct":false,"isInput":false,"amount":0.0267826162975192,"description":"calculated based on the properties of the leachate as they are shown in the respective treatment dataset. Globally, 58.73% losses are considered (Doka G., 2021. A model for composition-specific life cycle inventories of regionalised wastewater fates, Separate publication, Doka LCA, Zurich, Switzerland. <http://www.doka.ch/publications.htm>"),"internalId":16,"location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Global","category":"Global"},"flow":{"@type":"Flow","@id":"fc2371dc-5bff-41f6-a155-697fbf727b56","name":"Silicon","category":"Elementary flows/Emission to water/surface water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},"unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffc8f","name":"kg SWU"},"flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","name":"Mass","category":"Technical flow properties","refUnit":"kg"}},{ "@type":"Exchange","isAvoidedProduct":false,"isInput":false,"amount":2.94685133222682E-7,"description":"calculated based on the properties of the leachate as they are shown in the respective treatment dataset. Globally, 58.73% losses are considered (Doka

G., 2021. A model for composition-specific life cycle inventories of regionalised wastewater fates, Separate publication, Doka LCA, Zurich, Switzerland.

```
http://www.doka.ch/publications.htm)","internalId":12,"location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Global","category":"Global"},"flow":{"@type":"Flow","@id":"4d07c245-a030-5399-a74e-aaac30014ec3","name":"Hydrogen","category":"Elementary flows/Emission to water/surface water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},"unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffcf8f","name":"kg SWU"},"flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","name":"Mass","category":"Technical flow properties","refUnit":"kg"}},{ "@type":"Exchange","isAvoidedProduct":false,"isInput":false,"amount":2.31039527343132E-6,"description":"calculated based on TOC amount as it is described in Doka G. (2003) Life Cycle Inventories of Waste Treatment Services. ecoinvent report No. 13, Swiss Centre for Life Cycle Inventories, Dübendorf, December 2003. Available at: https://www.doka.ch/13_III_Landfills.pdf","internalId":6,"location":{"@type":"Location","@id":"56bca136-90bb-3a77-9abb-7ce558af711e","name":"Global","category":"Global"},"flow":{"@type":"Flow","@id":"fc0b5c85-3b49-42c2-a3fd-db7e57b696e3","name":"COD, Chemical Oxygen Demand","category":"Elementary flows/Emission to water/surface water","flowType":"ELEMENTARY_FLOW","refUnit":"kg"},"unit":{"@type":"Unit","@id":"95dd50e9-c184-4412-9afc-9764b1ffcf8f","name":"kg SWU"},"flowProperty":{"@type":"FlowProperty","@id":"93a60a56-a3c8-11da-a746-0800200b9a66","name":"Mass","category":"Technical flow properties","refUnit":"kg"}}, "socialAspects":[], "allocationFactors":[] }
```

Output data was extracted from processes in format 2 and added to process information in format 1. This was done by matching the two processes based on process ID, matching the name between the two “|” in the process name field with the flow name of an exchange in format 2 data, and then extracting the amount and units of that exchange. That output information was added to the format 1 data in the following format:

```
"process_output": {
  "Amount": 1.0,
  "Unit": "m3"
}
```

From there format 1 data was converted into strings to improve the accuracy of vector search. This is held in the TextData.txt file. The above process would be in the format:

Process name: market for leachate, SLF, waste glass | leachate, SLF, waste glass | Cutoff, U.
Process ID: 00ca40a7-f952-3e81-aca8-97a3d71a1ffc. Location: GLO. Fossil fuel depletion

impact: 0.2591615617351554 MJ surplus. Respiratory effects impact: 0.00028351683127839 kg PM2.5 eq. Global warming impact: 0.24882826264673022 kg CO2 eq. Carcinogenics impact: 3.3864243791365285e-07 CTUh. Smog impact: 0.015025065883468525 kg O3 eq. Acidification impact: 0.0009401520887998984 kg SO2 eq. Non carcinogenics impact: 1.040197603143794e-07 CTUh. Ozone depletion impact: 3.517171040651846e-09 kg CFC-11 eq. Eutrophication impact: 0.0016290990589930396 kg N eq. Ecotoxicity impact: 8.468621418917154 CTUe. Process output: 1.0 m3.

Each line in the text file contains the data for one process. The name and location fields were extracted to form a second dataset called NameData.txt. The above process in this data file would be shown as:

Process name: market for leachate, SLF, waste glass | leachate, SLF, waste glass | Cutoff, U.
Location: GLO.

The NameData.txt file was used to create the vector database because it leaves out any information not relevant to semantically identifying a process. TextData.txt is loaded into Chatbot.py and used to store the full impact data for retrieval once the name and location of the process relevant to the user query has been identified by vector search and the LLM.

The storage of the vectorized database is as follows:

Embedding.cpython-39.pyc in project_root/backed/_pycache_

default_vector_store.JSON, docstore.JSON, graph_store.JSON, image_vector_store.JSON, index_store.JSON in project_root/backend/nameStorage

The vector database can be accessed through the REST API created by Server.py which uses the vector search functionality from Embedding.py.

LLM

The **LLM layer** is the core intelligence behind the EcoValid chatbot, responsible for processing user queries, retrieving relevant environmental data, and generating meaningful responses. It operates within a **Retrieval-Augmented Generation (RAG) framework**, ensuring that responses are grounded in accurate and contextually relevant data rather than relying solely on the model's internal knowledge. When a user submits a query, the **embedding model** converts it into a vector representation, which is then matched against a **vector database** containing pre-indexed environmental datasets. The most relevant documents are retrieved and passed to the

LLM as context, enabling it to generate informed responses. Additionally, if the user submits environmental impact data for validation, the LLM compares it against **trusted sources** and generates a validation report, summarizing discrepancies and providing insights. The **LLM also refines outputs** through structured prompting, ensuring clarity and conciseness in responses. This combination of retrieval-based search and generative AI allows EcoValid to deliver fact-based, dynamic, and user-friendly interactions.

The OpenAI LLM used is accessed through a third party provider called LiteLLM. Calls to the LLM are made using the completion function from LiteLLM's python library.

FastAPI

FastAPI is a python compatible web framework that is used to build APIs. An API was necessary to achieve NFR2, allowing access to multiple users simultaneously. Thus, FastAPI was used to build an API connection that would allow multiple users to perform vector search on the same vector database simultaneously.

Streamlit

Streamlit is a python library that streamlines the process of chatbot creation. It was used as a basis for our chatbot implementation. It allowed us to create a generic chatbot frontend that could retrieve input from users into the backend and push text from the backend out to the frontend. We used it as a basis for our frontend, building customizations on top of it for our particular uses. It is also what we use to read text input from the users to the backend and output the query responses to the frontend.

Prompt Engineering

Prompt engineering is used throughout Chatbot.py to utilize the full functionality of the LLM. It is used to check if a user query is an environmental query or an environmental validation question or neither, return the process names and location of the process associated with environmental queries, choose the best fit process to the query, generate visualizations, and answer user questions using process data with added context. It is essential to retrieving accurate environmental data to correctly answer users' environmental and environmental data validation queries laid out in FR1.1 and FR1.2. It further contributes to achieving FR 1.3, FR 1.4, FR 1.5, FR 1.8, FR 2.3, and FR 2.4 by customizing the responses from the LLM and vectorized database.

For the call from the chatbot to the LLM to classify the user input-

"Based on the following input, classify the user's intent as one of the following categories: "query" if the user is asking a technical question about environmental impact data. "validate" if the user is requesting validation of environmental impact data they have provided. "neither" if

the input does not fit into either category. Provide only the category name as the answer. Provide the correct answer for each location. Input:” + userInput

For the call to the LLM to check if a user query identifies a location:

“Does this query specify a single region? (Global counts as a region) Answer "yes" or "no".

Query: ” + userInput

For the call to the vector database identify the best fit process name and location-

“Given this query, return just the verbatim text (process name and location) of the process that best matches the query. Query:” + userInput

For the call to the vector database to identify the best fit process name-

“Given this query, return just the verbatim text of the process that best matches the query.

Query:” + userInput

For the call to the vector database to identify the locations of a process-

“Provide all the locations for each instance of the following process. List them in the format

"Location1,Location2,etc." Process Name: “ + processName

For the call to the LLM to answer the user query with retrieved data and add context for a policymaker-

"You are answering a question about environmental impact using real data. Use the retrieved information below to craft your response, ensuring accuracy. Retrieved Data:" + retrievedData + "User query:" + userInput + "Answer the user query factually without making assumptions. Craft your answer to a policymaker audience. Do not show equations. Show the process name and location in your answer."

For call to the LLM to answer user query with retrieved data and add context for a researcher-

"You are answering a question about environmental impact using real data. Use the retrieved information below to craft your response, ensuring accuracy. Retrieved Data:" + retrievedData + "User query:" + userInput + "Answer the user query factually without making assumptions. Craft your answer to a researcher audience. Do not show equations. Show the process name and location in your answer."

For the call to the LLM to generate the code for visuals:

“Given the following question and answer, return Python code using ONLY matplotlib to generate a simple visualization to support the answer. Use plt.subplots(), and ensure the chart includes a title and axis labels. Return ONLY the code in a Python code block, and DO NOT include explanations or markdown. And please double check the code to make sure there are no

errors. This code is going to be run automatically. Question:" + userInput + "Answer: {str}" + finalResponse

For call to the LLM to validate data with retrieved data and add context for a policymaker-

"You are validating user data about environmental impact using real data. Use this process data to check the accuracy of the user data. Retrieved Data: " + str + "User data:" + prompt +

"Validate user data factually without making assumptions. Craft your answer to a policymaker audience. Do not use Equations. List the process name and location in the answer."

For call to the LLM to validate data with retrieved data and add context for a researcher-

"You are validating user data about environmental impact using real data. Use this process data to check the accuracy of the user data. Retrieved Data: " + str + "User data:" + prompt +

"Validate user data factually without making assumptions. Craft your answer to a researcher audience. Do not use Equations. List the process name and location in the answer."

GUI Design

Current design:

When users launch the application, the GUI design will open to a landing page. This page will describe what the project is, why the user should use it, and give usage tips. This landing page then directs users to the chatbot page. They are prompted by a menu in the top left to select their role as either "Researcher" or "Policymaker." This way we can customize answers to users in those two roles as laid out in our requirements. After they select their role a bar on the bottom of the screen pops up that allows them to enter queries into the chatbot. A chatbot format was chosen because it fits well with the requirement of answering the users environmental queries and will be a familiar and easy to navigate format for most users. To allow for data validation, users can also manually enter data into the query field and ask for it to be validated against our trusted environmental database. If a user decides to change their role, all previous messages will be cleared off the screen. When the user clicks "Export Chat" a button named "Download PDF" appears under it to allow them to download the chat as a PDF.

Launch Page

Copy 1

Welcome to EcoValid

Your AI-powered assistant for understanding environmental impact data.

What is EcoValid?

EcoValid helps researchers and policy makers explore complex environmental data using simple operations.

Why use it?

- Improved data insights from experts
- Fast, data-backed answers
- Tailored results for your role
- Simple chat interface

Try asking:

- "What is the environmental impact of fossil electricity in China?"
- "How does wind energy compare to solar energy for global warming?"

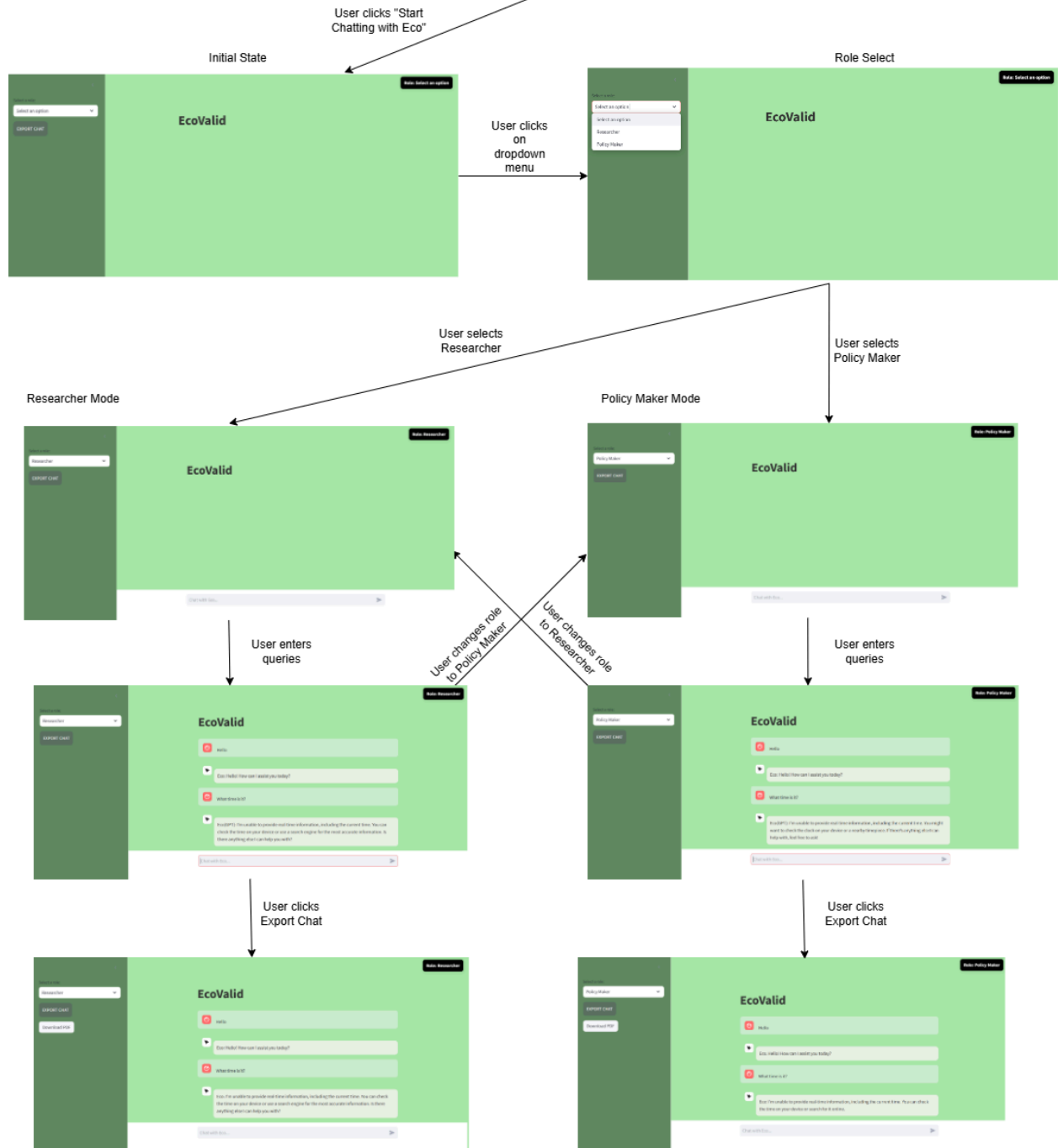


Figure 4: Current GUI

Implementation

Author: Esha Mascarenhas

Reviewer/Editor(s): Cole Sanders, Annabella Jackmore

At the time of **Final Project Report (IPR)** submission, the **EcoValid chatbot** implementation consists of multiple key Python modules handling different functionalities.

The **frontend**, implemented in Chatbot.py, provides the user interface using **Streamlit**. It includes a **chatbot interface**. The frontend sends user queries to the **backend server (Server.py)** via **REST API calls**.

The **backend**, built with **FastAPI**, processes incoming requests and manages interactions with the **LLM**, the **vector database**, and the **validation module**.

The Server.py file defines the API endpoints, including ones for **handling user queries**, **retrieving relevant information**, and **processing validation checks**.

The **LLM layer** plays a crucial role in processing user input. The embeddings.py file is responsible for **embedding environmental data** from data.json using **FAISS/ChromaDB**. This allows the system to **retrieve semantically similar environmental impact information** when a user submits a query. Once relevant data is retrieved, the **LLM refines the response** before sending it back to the frontend. Additionally, if the user inputs data for validation, the chatbot routes the data through **LlamaIndex**, which **converts unstructured files into structured JSON format** for further processing.

Iteration Definition

Author: Sid Kaju

Reviewer/Editor: Cole Sanders, [Esha Mascarenhas](#)

Iteration 1 (02/03/2025 - 02/17/2025)

- We worked on creating frontend wireframes, requirements, and a system test plan.
- We created a naive chatbot that can answer any question that the user asks based on the knowledge from OpenAI's gpt4-0 model. This includes the frontend for the chatbot and the backend that uses the OpenAI api key.
- We added the option to select a role for a chat session. [FR 3.1]
- The chat refreshes if the user changes roles. [FR 3.2]

Iteration 2 (02/17/2025 - 03/8/2025)

- We worked on getting our chatbot to answer specific questions based on our environmental data using a Retrieval-Augmented Generation (RAG) pipeline. [FR 2.1]
- The system can recognize when a question needs our environmental data to be answered or if it's just needs a regular answer [NFR 4.1]

Iteration 3 (03/17/2025 - 03/31/2025)

- The chatbot is able to generate detailed reports related to the answer to the question that the user asks [FR 2.3]
- The system is able to add context to the answers to help the user understand the answer better. [FR 2.6]
- The user is able to enter data to check if it's valid based on our dataset. [FR1.2]
- The user receives a validation report when data is entered into the chatbot [FR1.2]

Iteration 4 (04/01/2025 - 04/17/2025)

- The full validated data is integrated into the chatbot [FR 1.5].
- System tests that cover the entire functionality of EcoValid are written.
- Wrote unit tests that cover the entire functionality of EcoValid. The unit tests reached our goals of 70% statement coverage, 60% branch coverage and 80% method coverage.

Security Considerations

Author: Sid Kaju

Reviewer/Editor: Madison Torres

Our sponsors decided to keep security features outside the scope of the project. They made it clear that our focus should be on the functionality. Since this is a new project, the end goal of how this application is going to be used was not entirely defined. This includes details like the kinds of users using the project, whether it will be available to the public or not, and whether or not more teams would work on it before it is completely ready to be launched.

Throughout our implementation, we, as a team, still attempted to add basic security wherever we felt like it was needed. We have discussed what we have implemented down below in some of the paragraphs, but most of it is focused on what could and needs to be implemented if the application gets launched to the public. We also mentioned how these security considerations can be integrated into our current project.

Confidentiality: We are protecting the API keys by using them as environment variables, ensuring they are never hardcoded. They are excluded in our codebase through a .gitignore file to reduce risk of exposure. It is important that the user cannot access this and use our API key, driving up costs. The data is also stored as vectors instead of plain text. For future consideration, there could be encrypted storage of the API keys and other sensitive data relating to the project or implementing an automatic key rotation in case of leak.

Integrity: It is critical to ensure that the environmental process data is accurate, especially since the tool will be used by policymakers to make strategic decisions. All the environmental impact data has been compiled by our sponsors, Dr Xu. and Dr. Lan, from trustworthy sources. Nobody has the ability to modify or corrupt the data. To further prevent this from happening, future teams can limit who has write access to the data to the development team and the sponsors. Also, they can continue to use our version control on Github to track any changes made and rollback in case of errors made. For future consideration, an automated monitoring system of some sort can be implemented to look for any inconsistencies in the dataset.

Availability: The chatbot needs to be available and responsive at all times for all users. Risks include API rate limits, losing connection to the data through network outages, and excessive user load. We are using the paid plan for our LLM to increase the API rate limit and handle large user loads. Additionally, we are using an older version of the LLM that offers more cost effective API calls to further support large user loads. Future teams could work on limiting the amount a single user can spend with our API key. This means they wouldn't be allowed to spam messages or ask irrelevant questions. Our team has looked into the framework Flask-Limiter that can help impose these rates and prevent spamming. In the future, teams could cache frequently retrieved data. This would also help if OpenAI ever goes down. Libraries such as Flask-limiter can be used to enforce rate limits. Human detection tools such as CAPTCHA to avoid any bots should be considered.

Identification and Authentication: Our system does not have any login functionality or user authentication features. All the users have the same level of access and functionality offered to them. This may make the system vulnerable to bots and data scraping attacks. In the future, other teams can implement login functionality, which can also expand the functionality to allow for a more personal experience with enhanced privacy. This functionality can be done through Flask or Django easily for secure user verification. Implementing multi-factor authentication can be done for added security.

Accountability: A potential threat is if any user includes sensitive information within their query, which goes through to OpenAI. Future teams could work on adding a disclaimer to all users that will remind them not to include any sensitive data. They can also implement logging within each session to monitor suspicious activity without storing sensitive data. Storing these

logs with encryption and limiting access will prevent tampering and having session-based tracking can identify any suspicious patterns. Also, if there is a future login functionality, real-time monitoring for unusual activity such as repeated failed attempts can be implemented.

Privacy: The chatbot does not collect any info from the user or the user queries. However, since the chatbot queries are processed by OpenAI, the user data is temporarily exposed to OpenAI's systems. If logs are implemented in the future, the software can store the logs and make sure they do not contain any sensitive information. The user queries are being sent to OpenAI, so the user needs to be careful while constructing queries. Disclaimers to let the user know that the content of the queries goes to OpenAI and are being processed externally can be shown. Future teams can anonymize user data before sending it to OpenAI to prevent exposure of any sensitive data and configure the system to delete logs after a specified period of time. In the future, an option to allow users to opt out of any sort of logging done can be offered to keep their own privacy.

Project Folder Structure

Author: Esha Mascarenhas

Reviewer/Editor: Annabella Jackmore

The project is organized into three main sections: backend, frontend, and tests.

Within the backend folder, there are several key Python files and subfolders. Notable files include Chatbot.py (the main chatbot logic and interface), Server.py (which manages API interactions), and Embedding.py (responsible for handling vector embeddings and retrieval). The backend also contains utility scripts such as DataCleaner.py, DataExtraction.py, and NameExtractor.py, as well as essential data files like Data.json, NameData.txt, and TextData.txt. Subfolders like Impact_Data, Process_Data/processes, and nameStorage organize supporting datasets and resources.

The frontend folder contains a components subfolder (for assets such as bird.jpg) and a pages subfolder (for frontend configuration like package.json).

The tests folder contains unit tests for validating individual components of the backend, including modules for API endpoints, classification logic, embedding functions, mock model responses, and edge case handling. These tests ensure the robustness and reliability of the system by maintaining high test coverage and supporting automated verification during development.

At the project root, you'll find standard files including .gitignore (to specify ignored files for version control) and README (which provides an overview and instructions for the project).

This structure separates backend logic, data, and frontend assets for maintainability and clarity.

This structured organization ensures that different aspects of the project are well-separated and easy to navigate.

Project Configuration/Settings

Authors: Esha Mascarenhas, Cole Sanders

Reviewer/Editor: Annabella Jackmore

The project configuration is managed through environment variables and required dependencies that must be installed for the system to function properly. The .env file is used to store sensitive information such as API keys, authentication credentials, and other configurable parameters. This ensures security by preventing hardcoded secrets in the codebase. A backup of this file, .env.backup, is also maintained.

To set up the project, a new user must install the required dependencies. The system relies on several Python libraries, which can be installed using pip. A requirements.txt file should ideally be included to allow for easy installation by running `pip install -r requirements.txt`.

Based on the project structure and implementation files, the requirements.txt file should contain the following dependencies, which are required for the system to function properly:

```
streamlit
fastapi
uvicorn
openai
llama-index
faiss-cpu
python-dotenv
pydantic
numpy
pandas
json5
os
requests
litellm
reportlab
re
io
base64
matplotlib
```

To set up a local .env file to connect to the LLM, first create a .env file then add the following:

```
API_KEY = sk-RFtD2ZNA2JFex5hf-5Pnw
```

Next download the large file, default__vector_store.json and add it to the following folder:

```
project_root/backend/nameStorage
```

To run the project, users must use two terminals. In one terminal, run Server.py with the following command:

```
python project_root/backend/Server.py
```

In the second terminal navigate to the project_root directory. Then run the following command:

```
python -m streamlit run backend/Chatbot.py
```

This will launch the application in your browser. Additionally you can navigate to the following address:

```
http://192.168.86.31:8501
```

Testing

Authors: Madison Torres, Cole Sanders, Annabella Jackmore

Reviewer/Editor: Sid Kaju

Overall View

The EcoValid system integrates AI-based query processing with environmental impact datasets. Given a heavy reliance on Large Language Models(LLMs), data ingestion, and API interactions, we employed different layers of testing to ensure functionality, performance, and reliability. Our testing strategy includes unit testing, acceptance testing, and usability testing to ensure a robust, accurate, and user-friendly system.

Types of Testing Used

Unit testing was conducted to validate individual components of the system. This included testing data ingestion and the vector indexing process, ensuring that environmental data was properly parsed and stored for retrieval. We also tested the logic behind query processing and

classification, verifying that environmental queries were correctly identified and routed. The correctness of API endpoints was assessed to ensure they responded reliably and accurately. We also simulated OpenAI API interactions and handled different types of edge cases to make sure that the system could handle all types of queries, including ones that we did not have the data for and empty ones.

Acceptance testing focused on the functionality from an end-user perspective. We verified that user roles were clearly defined and selectable, and that role-based interactions worked as intended. The system was evaluated to confirm that environmental queries returned relevant, accurate impact data, while general queries were correctly handled by GPT-4o. Specifically, testing whether environmental queries were routed to the RAG-based model that uses vector search, while general questions were handled by the GPT-4o model.

Lastly, usability testing assessed the intuitiveness and clarity of the Streamlit-based user interface. We ensured that users across different roles could easily navigate the interface, input queries, and interpret system responses. The UI was evaluated for its ability to easily switch between roles and its effectiveness in adapting its responses based on the selected role, including showing different environmental data or recommendations depending on the user's perspective.

Testing Environment:

- Operation System: Windows 10/11
- Programming Language: Python 3.13
- Frameworks:
 - FastAPI (API)
 - LlamaIndex (Data Ingestion)
 - pytest (Unit Testing)
 - Streamlit (UI)

The project uses a structured dataset of over 23,000 entries, stored in AllProcessData.json, detailing the environmental impacts of various processes such as energy production and waste treatment. Data exists in two formats, merged using process IDs, with results written to TextData.txt for full impact descriptions and NameData.txt for simplified name-location pairs. These files power a vector database built from NameData.txt, enabling semantic search through files like default_vector_store.json and index_store.json in the nameStorage directory. The database is accessed via a REST API in Server.py, allowing the chatbot to retrieve accurate environmental data during queries.

To ensure tests run independently of external services, we used pytest-mock and Python's built-in unittest.mock library to simulate API calls and database interactions. This approach allowed us to test system logic and error handling without relying on live OpenAI responses or external data sources, improving test reliability and execution speed.

Test Setup Instructions:

To set up the testing environment:

1. Ensure that the following files:
default__vector_store.json
docstore.json
graph_store.json
image__vector_store.json
index_store.json
are in the following places for proper indexing:
2025Spring-Team24-Xu-Lan/project_root/backend/nameStorage
2025Spring-Team24-Xu-Lan/project_root/backend/storage
2025Spring-Team24-Xu-Lan/project_root/storage
2025Spring-Team24-Xu-Lan/storage
2. Navigate to the project root directory with the following command: `cd project_root`
3. Run unit tests with the following command: `pytest tests`
4. (Optional) To get coverage reports, run the unit tests with the following command: `pytest tests --cov=backend --cov-report=term-missing --cov-report=html --cov-branch`

Unit Testing

Unit Testing Tools

- pytest - Used for primary unit testing framework.
- pytest-cov - Used to generate code coverage reports.
- pytest-mock - Mocks API responses and function calls.

Tested Components

Test File	Functionality Covered
test_embedding.py	Tests data ingestion and vector index creation.
test_query_engine.py	Validates query processing.
test_api.py	Ensures FastAPI handles queries correctly.
test_classification.py	Confirms query classification logic

	(environmental vs general).
test_mocking.py	Mocks OpenAI API calls to prevent real charges.
test_edge_cases.py	Tests edge cases that arise such as invalid queries and empty inputs.
test_chatbotUI.py	Tests the Streamlit based UI export chat, functions.

Coverage Goals and Expectations

Statement Coverage: $\geq 70\%$

Branch Coverage: $\geq 60\%$

Method Coverage: $\geq 80\%$

Our coverage goals are designed to ensure that a majority of the codebase is exercised during testing, making it easier to detect regressions and logic errors early. Statement coverage of at least 70% ensures that most of the lines of code are executed at least once, reducing the risk of untested functionality. Branch coverage of at least 60% is appropriate given that a lot of the query logic has conditionals. The project relies on complex decision-making logic, and this ensures that both the expected and edge-case scenarios are verified. Method coverage of 80% is chosen to confirm that almost all defined functions undergo testing, guaranteeing that core functionalities are reliable. These goals strike a balance between feasibility and thoroughness, ensuring that critical system functions are adequately tested and behavior is confirmed.

To verify our coverage, we run the following command:

```
pytest backend/tests --cov=backend--cov-report=term-missing
```

This generates the following coverage report to show which parts of code were executed during testing:

```
----- coverage: platform win32, python 3.13.1-final-0 -----
```

```
Name                               Stmts  Miss  Branch  BrPart  Cover  Missing
```

```
-----
```

backend\Chatbot.py	239	162	64	4	30%	68->63, 134-135, 137->143, 148-469
backend\DataCleaner.py	19	19	4	0	0%	1-31
backend\DataExtraction.py	22	22	6	0	0%	3-46
backend\Embedding.py	33	0	8	1	98%	37->35
backend\NameExtractor.py	18	18	4	0	0%	1-34
backend\Server.py	27	3	4	1	87%	11-13, 42
backend__init__.py	0	0	0	0	100%	

TOTAL	358	224	90	6	35%	

Excluded Components

Certain components are not tested in the unit tests due to the following limitations:

1. Chatbot.py - While there are some components tested in this class, the entirety of the class is not able to be tested. This is due to the fact that the structure of our UI runs certain methods in an if-else statement that is defined inside at runtime of the function, and not at import. This makes the functions only exist after Streamlit runs Chatboy.py and pytest only imports the file once at the very beginning.
2. DataCleaner.py, DataExtraction.py, and NameExtractor.py - These three classes served as helper utilities to transform and prepare the environmental process data used by the main application. Since these scripts were used primarily for one-time data formatting and cleanup prior to deployment, they operate independently of the chatbot's runtime logic. As such, they were not included in the unit tests because they perform straightforward file I/O and data parsing without exposing reusable logic or contributing directly to application functionality.
3. Third-Party Libraries - LlamaIndex and OpenAI models are mocked rather than directly tested to save cost. Mocking ensures that unit tests can run without reliance on external APIs, but usability testing helps validate whether responses generated by these services are relevant and interpretable for end-users.

Usability testing will help our team evaluate these excluded components by allowing real users to interact with the system and identify potential usability gaps. This ensures that even components not covered by the automated tests, such as the deployment scripts, undergo

thorough evaluation through hands-on testing. Also with usability tests, we can undergo thorough evaluation with newly tested components such as the UI. By collecting user feedback, we can refine the UI, improve role-based interactions, and detect overlooked issues that might not be apparent in unit testing. Additionally, usability testing helps validate the effectiveness of UI-based error handling and accessibility for different user roles.

Testing Summary

The successful results of all unit tests indicate that the project is both functionally complete and technically sound. Every major component, ranging from the chatbot logic and API endpoints to classification and embedding modules, has been thoroughly tested, with all tests passing. This confirms that the system behaves as expected across a wide range of scenarios, including standard use cases and edge conditions. The passing test suite provides strong evidence of code correctness, stability, and integration consistency, and it reflects a high level of confidence in the system's reliability. Overall, the test results validate that the project is ready for deployment or further user evaluation. However, while unit tests verify the correctness of individual components, they cannot fully evaluate the accuracy or appropriateness of the chatbot's generated responses. The following acceptance tests will ensure that the user experience aligns with project goals and that the model delivers helpful, relevant information in a conversational context. Overall, the passing unit tests confirm backend integrity, while system testing ensures practical effectiveness.

Acceptance Testing

Acceptance Test Plan: Team 24

Annabella Jackmore, Sid Kaju, Esha Mascarenhas, Cole Sanders, Madison Torres

Test Plan Executed: 4/24/2025

Acceptance testing ensures that the system functions as intended and meets user expectations. Our team has developed a test plan covering various aspects of the system, from role selection to chatbot functionality. Below is the Acceptance Test Plan for Team 24:

System Function

This section of the Test Plan works on testing the system's functionality, as in what we expect the system to do. This functionality includes role selection for the user and dealing with valid and invalid cases of these, as well as switching roles during chats.

Test ID	Description	Expected Results	Actual Results
1. testRoleSelectionResearcher	<u>Preconditions:</u> None <ol style="list-style-type: none"> 1. User navigates to the chatbot UI at: http://localhost:8501/ (will update with official link if deployed). 2. User selects the role “Researcher” from the side panel drop-down menu. 3. User will close the side panel. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Researcher”. • The message panel at the bottom of the page opens for the user to give input to the chatbot. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Researcher”. • The message panel at the bottom of the page opens for the user to give input to the chatbot.
2. testRoleSelectionPolicyMaker	<u>Preconditions:</u> None <ol style="list-style-type: none"> 1. User navigates to the chatbot UI at: http://localhost:8501/ (will update with official link if deployed). 2. User selects the role “Policymaker” from the side panel drop-down menu. 3. User will close the side panel. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Policymaker” • The message panel at the bottom of the page opens for the user to give input to the chatbot. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Policymaker” • The message panel at the bottom of the page opens for the user to give input to the chatbot.
3. testInvalidRoleSelection	<u>Preconditions:</u> None <ol style="list-style-type: none"> 1. User navigates to the chatbot UI at: http://localhost:8501/ (will update with official link if deployed). 2. User clicks on the drop-down menu, and backspace all of the text “Select an option”. 3. User types in “Waiter” into the box and presses enter. 	<ul style="list-style-type: none"> • The drop-down menu will display “No results”. • The user will not be able to select this role and is not able to move forward with the role selection. 	<ul style="list-style-type: none"> • The drop-down menu will display “No results”. • The user will not be able to select this role and is not able to move forward with the role selection.
4. testUnselectRoleDuringChat	<u>Preconditions:</u> Test 1 completed (testRoleSelectionResearcher) and Test 6 completed (testBasicQuery) <ol style="list-style-type: none"> 1. User navigates to the side panel and opens the dropdown menu. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Select an option” • The messages exchanged will 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Select an option” • The messages exchanged will

Test ID	Description	Expected Results	Actual Results
	2. From the drop-down menu, the user selects the option “Select an option”. 3. The user closes the side panel.	disappear and the input box at the bottom disappears, not allowing the user to chat with Eco.	disappear and the input box at the bottom disappears, not allowing the user to chat with Eco.
5. testRoleChange	<u>Preconditions:</u> Test 1 completed (testRoleSelectionResearcher) and Test 6 completed (testBasicQuery) 1. User opens the side panel and opens the drop-down menu. 2. From the drop-down menu, the user selects the role “Policymaker” 3. The user closes the side panel.	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Policymaker” • The messages exchanged will disappear and host a blank slate. 	<ul style="list-style-type: none"> • The Role: icon in the top right corner switched to “Policymaker” • The messages exchanged will disappear and host a blank slate.

Chatbot

This section of the Test Plan works on testing the Chatbot. Since there isn’t a way to completely predict what the chatbot will send back as a response, included in the Expect Results is the content that is required in the messages sent back. The chatbot should return a message containing the content, but how the content is structured is variable. This functionality includes getting environmental information based on user queries, generating validation summaries, user data upload, and generating detailed reports.

Test ID	Description	Expected Results	Actual Results
6. testBasicQuery	<u>Preconditions:</u> Test 1 completed (testRoleSelectionResearcher) 1. User navigates to the chat panel at the bottom of the page. 2. User types “Hello Eco!” 3. User hits the ‘Enter’ key.	The chatbot displays the text “Generating” and then replies with the required following content: - A welcome message.	The chatbot displays the text “Generating” and then replies with: - “Hello! How can I assist you today?”
7.	<u>Preconditions:</u> Test 1 completed	The chatbot displays the	The chatbot displays the

Test ID	Description	Expected Results	Actual Results
testUnquerableChat	(testRoleSelectionResearcher) <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types “ ” 3. User hits the ‘Enter’ key. 	text “Generating” and then replies with the required following content: <ul style="list-style-type: none"> - A prompt to the user for input. 	text “Generating” and then replies with: <ul style="list-style-type: none"> - “Hello! How can I assist you today?”
8. testUserUploadsValidData	<u>Preconditions:</u> Test 1 completed (testRoleSelectionResearcher) <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “Can you validate this data: In Sweden (SE), the GWP is 10.667660808703475 kg CO2 eq per m³ for producing 1 m3 spruce timber” 3. User hits the ‘Enter’ key. 	The data is validated. The response from the chatbot indicates that the data is valid.	The chatbot displays the text “Generating” and then replies with: <ul style="list-style-type: none"> - “The user data claims that the global warming potential (GWP) for producing 1 m³ of spruce timber in Sweden (SE) is 10.667660808703475 kg CO2 eq per m³. According to the retrieved process data, the global warming impact for the process of softwood forestry of spruce under sustainable forest management in the specified location is indeed 10.667660808703475 kg CO2 eq per m³ of output. Hence, the user data is factually

Test ID	Description	Expected Results	Actual Results
			<p>accurate in regard to the global warming potential for this specific process and location.</p> <ul style="list-style-type: none"> - Process name: softwood forestry, spruce, sustainable forest management Location: SE.”
9. testUserUploadsInvalidData	<p><u>Preconditions:</u> Test 1 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “Can you validate this data: the ecotoxicity impact of bus maintenance in Switzerland is 3 CTUe”. 3. User hits the ‘Enter’ key. 	<p>The data is validated. The response from the chatbot indicates that the data is invalid and prompts for reupload.</p>	<p>The chatbot displays the text “Generating” and then replies with:</p> <ul style="list-style-type: none"> - “Upon examination of the retrieved process data for "maintenance, bus" in Switzerland, the ecotoxicity impact is reported as 916,779.92136533 27 Comparative Toxicity Units for ecosystems (CTUe). This represents the quantifiable potential harm to ecosystems from pollutants released during the

Test ID	Description	Expected Results	Actual Results
			<p>maintenance process. In contrast, the user has provided data indicating an ecotoxicity impact value of only 3 CTUe.</p> <p>For the researcher audience, it is crucial to highlight the stark disparity between the retrieved data and the user data. The value from the authoritative source suggests a significantly higher ecotoxicity impact associated with bus maintenance than the user data indicates. Such a marked difference underscores the importance of ensuring the accuracy and reliability of environmental impact assessments.</p> <p>Ultimately, the user data does not align with the retrieved data for ecotoxicity impact. The user data appears to inaccurately underrepresent the ecological consequences of bus maintenance in Switzerland. Researchers</p>

Test ID	Description	Expected Results	Actual Results
			should be aware that the validated data indicate a substantial ecotoxicity burden, as reflected in the process data retrieved.”
10. testPolicyMakerQueryResponse	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionPolicyMaker)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the acidification of bus maintenance in Switzerland and what can be done for this?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - Process: “maintenance, bus” - Acidification: “31.71 kg SO2 eq - Location: “CH” - Information about how to lower this number. 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The acidification impact of bus maintenance in Switzerland is 31.710 kg SO2 eq per the process data for "maintenance, bus maintenance, bus Cutoff, U." <p>To address acidification, strategies may include using maintenance practices that reduce emissions of sulfur dioxide (SO2) and nitrogen oxides (NOx), such as regular engine tune-ups, using low-sulfur fuels, and implementing catalytic converters where applicable. Additionally, improving maintenance protocols to incorporate</p>

Test ID	Description	Expected Results	Actual Results
			more efficient and environmentally friendly technologies could further mitigate the impact.

Expected Results from Sponsors

This section of the Test Plan works on testing with the sponsors' questions and expected results. Since there isn't a way to completely predict what the chatbot will send back as a response, included in the Expect Results is the content that is required in the messages sent back. The chatbot should return a message containing the content, but how the content is structured is variable. This functionality includes getting environmental information based on user queries.

Test ID	Description	Expected Results	Actual Results
11. testSponsorQuery1	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in "What is the life-cycle global warming potential of generating 1 kWh electricity (high voltage) in the USA by average?" 3. User hits the 'Enter' key 	<p>The chatbot displays the text "Generating" and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.45968 kg CO2 eq from process "market group for electricity, high voltage electricity, high voltage Cutoff, U" 	<p>The chatbot displays the text "Generating" and then replies with the following content:</p> <ul style="list-style-type: none"> - " To determine the life-cycle global warming potential (GWP) of generating 1 kWh of electricity (high voltage) in the USA, specifically for the region US-WECC (Western Electricity Coordinating Council), based on the provided data, we primarily have information from

Test ID	Description	Expected Results	Actual Results
			<p>the process, "electricity production, natural gas, conventional power plant electricity, high voltage."</p> <ul style="list-style-type: none">- From this process, the global warming impact is given as 0.598659573154043 kg CO2 equivalent per 3.6 MJ of electricity produced. Since 1 kWh is equivalent to 3.6 MJ, this translates directly to a global warming potential of approximately 0.5987 kg CO2 equivalent per kWh for this specific process.- Process Name: Electricity production, natural gas, conventional power plant Location: US-WECC- Given the lack of specific comprehensive average data for all forms of electricity generation in the region, this figure

Test ID	Description	Expected Results	Actual Results
			<p>should be regarded as the impact associated with electricity produced from natural gas in a conventional power plant within the specified area. For more comprehensive average data across all types of power generation, one would need aggregated data from various power generation sources, including wind, solar, nuclear, and others mentioned, but such data is not provided in the current dataset.”</p> <p>Reason for failure: The system works based on identifying processes that semantically match the question. However it has no reasoning capacity. Here the question is about an average, but the system has no ability to deduce that the average can be found in a process with the “market group for electricity” name. Those two terms did not semantically match.</p>

Test ID	Description	Expected Results	Actual Results
12. testSponsorQuery2	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the life-cycle acidification potential of generating 1 kWh electricity (high voltage) in the USA by average?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.00108 SO2 eq, from process “market group for electricity, high voltage electricity, high voltage Cutoff, U” 	<p>The chatbot displays the text “Generating” and then replies with the following content:</p> <ul style="list-style-type: none"> - “The life-cycle acidification potential of generating electricity using a natural gas combined cycle power plant, under the US-WECC region, is represented by the process "electricity production, natural gas, combined cycle power plant electricity, high voltage Cutoff, U." Based on the provided data, the acidification impact for this process is 0.000248410693269 4882 kg SO2 equivalent for a process output of 3.6 MJ. To convert the acidification impact to a per kWh basis, where 1 kWh is equivalent to 3.6 MJ, this translates directly to 0.000248410693269

Test ID	Description	Expected Results	Actual Results
			<p>4882 kg SO2 eq per kWh.”</p> <p>Reason for failure: The system works based on identifying processes that semantically match the question. However it has no reasoning capacity. Here the question is about an average, but the system has no ability to deduce that the average can be found in a process with the “market group for electricity” name. Those two terms did not semantically match.</p>
13. testSponsorQuery3	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the carbon intensity of generating 1 kWh electricity (high voltage) in the USA by average?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.45968 kg CO2 eq, from process “market group for electricity, high voltage electricity, high voltage Cutoff, U” 	<p>The chatbot displays the text “Generating” and then replies with the following content:</p> <ul style="list-style-type: none"> - “The carbon intensity of generating 1 kWh of electricity (high voltage) in the USA, specifically in the US-WECC region, can be derived from the process data. The relevant process is "electricity, high voltage, production mix electricity, high voltage Cutoff, U" with the

Test ID	Description	Expected Results	Actual Results
			<p>Process ID 33466a5e-7ad7-3508-9a53-af7c4ef9cf78. This process has a global warming impact of 0.376420653478276 kg CO2 eq per 3.6 MJ of energy produced. When converted to kWh (1 kWh is equivalent to 3.6 MJ), the carbon intensity remains the same at approximately 0.376 kg CO2 eq/kWh.”</p> <p>Reason for failure: The system works based on identifying processes that semantically match the question. However it has no reasoning capacity. Here the question is about an average, but the system has no ability to deduce that the average can be found in a process with the “market group for electricity” name. Those two terms did not semantically match.</p>
14. testSponsorQuery4	<u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)	The chatbot displays the text “Generating” and then replies with the required following content:	The chatbot displays the text “Generating” and then replies with the required following content:

Test ID	Description	Expected Results	Actual Results
	<ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the carbon footprint of generating 1 kWh electricity (high voltage) in the USA by average?” 3. User hits the ‘Enter’ key 	<ul style="list-style-type: none"> - 0.45968 kg CO2 eq, from process “market group for electricity, high voltage electricity, high voltage Cutoff, U” 	<ul style="list-style-type: none"> - “The carbon footprint of generating 1 kWh of high voltage electricity in the USA's WECC region can be derived from the process "market for electricity, high voltage electricity, high voltage Cutoff, U," with a global warming impact of approximately 0.38194 kg CO2 equivalent per 3.6 MJ of electricity produced. This process is applicable to the location US-WECC. Researchers can use this data as a reference point for understanding the carbon emissions associated with electricity generation in this specific region.” <p>Reason for failure: The process name successfully matched with the question, but the location matched to a subregion with a similar</p>

Test ID	Description	Expected Results	Actual Results
			name: US-WECC instead of just US.
15. testSponsorQuery5	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of generating 1 kWh electricity (high voltage) via geothermal in the southeastern US?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.07184kg CO2 eq, from process “electricity production, deep geothermal electricity, high voltage Cutoff, U” 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - To determine the global warming potential (GWP) of generating 1 kWh (equivalent to 3.6 MJ) of electricity via deep geothermal sources in the southeastern US (US-SERC region), we rely on the data retrieved from the specific process named "electricity production, deep geothermal electricity, high voltage Cutoff, U." - According to the process data, the global warming impact for producing 3.6 MJ of electricity is quantified as 0.07183816858151341 kg CO2 equivalent. Therefore, when generating 1 kWh of

Test ID	Description	Expected Results	Actual Results
			<p>electricity using deep geothermal methods in the southeastern US, the associated GWP is 0.071838 kg CO2 eq/kWh.</p> <ul style="list-style-type: none"> - Process name: electricity production, deep geothermal electricity, high voltage Cutoff, U.
16. testSponsorQuery6	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the range of GWP of generating 1 kWh electricity via geothermal in the US?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - Should be given a range around 0.1 kg CO2 eq, and a list of processes containing “electricity production, deep geothermal” 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The process data provided pertains to the environmental impacts of electricity production from deep geothermal sources in the US-Western Electricity Coordinating Council (US-WECC) region. From this data, the Global Warming Potential (GWP) of producing 3.6 MJ of

Test ID	Description	Expected Results	Actual Results
			<p>electricity is given as 0.071838168581513 86 kg CO2 eq.</p> <ul style="list-style-type: none"> - Converting 3.6 MJ to kWh, we note that 1 kWh is equivalent to 3.6 MJ. <p>Therefore, the GWP of generating 1 kWh of electricity via deep geothermal in this region is approximately 0.0718 kg CO2 eq.</p> <ul style="list-style-type: none"> - The process name related to this data is "electricity production, deep geothermal electricity, high voltage Cutoff, U."
17. testSponsorQuery7	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of producing 1 m3 spruce timber?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.04343 kg CO2 eq in Sweden, softwood forestry, spruce, sustainable forest management cleft timber, measured as dry mass Cutoff, U_SE - 0.04737 kg CO2 eq in Germany, softwood 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The Global Warming Potential (GWP) of producing 1 m³ of spruce timber under the process "softwood forestry, spruce, sustainable forest management

Test ID	Description	Expected Results	Actual Results
		<p>forestry, spruce, sustainable forest management cleft timber, measured as dry mass Cutoff, U_DE</p> <ul style="list-style-type: none"> - 0.04343 kg CO2 eq in Rest of World, softwood forestry, spruce, sustainable forest management cleft timber, measured as dry mass Cutoff, U_RoW 	<p>sawlog and veneer log, softwood, measured as solid wood under bark Cutoff, U" varies by location.</p> <p>Specifically:</p> <ul style="list-style-type: none"> • In the rest of the world (RoW), the GWP is 10.77206988429 0321 kg CO2 eq per m³ of spruce timber produced. • In Sweden (SE), the GWP is 10.66766080870 3475 kg CO2 eq per m³. • In Germany (DE), the GWP is 12.36047374152 6809 kg CO2 eq per m³. <p>Thus, you would reference the specific location for the most accurate data regarding the GWP impact.</p>
18. testSponsorQuery8	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.22613 kg CO2 eq in India, grape production grape Cutoff, U_IN. 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The Global Warming Potential

Test ID	Description	Expected Results	Actual Results
	<ol style="list-style-type: none"> User types in “What is the GWP of producing 1 kg grape?” User hits the ‘Enter’ key 	<ul style="list-style-type: none"> 0.14175 kg CO2 eq in Rest of World, grape production grape Cutoff, U_RoW 	<p>(GWP) of producing 1 kg of grapes varies depending on the geographic location of the production process. Based on the retrieved data:</p> <ol style="list-style-type: none"> For grape production in the Rest of the World (RoW), the GWP impact is 0.141747137584 44796 kg CO2 eq per 1 kg of grapes. For grape production in India (IN), the GWP impact is 0.226125901355 05758 kg CO2 eq per 1 kg of grapes. <p>The process name associated with the data is "grape production grape Cutoff, U."</p>
19. testSponsorQuery9	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> User navigates to the 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> 0.58627 kg CO2 eq, 	<p>The chatbot displays the text “Generating” and then replies with the following content:</p> <ul style="list-style-type: none"> “The Global

Test ID	Description	Expected Results	Actual Results
	<p>chat panel at the bottom of the page.</p> <p>2. User types in “What is the GWP of producing 1 kg offshore petroleum in Mexico?”</p> <p>3. User hits the ‘Enter’ key</p>	<p>petroleum and gas production, offshore petroleum Cutoff, U_MX</p>	<p>Warming Potential (GWP) of producing 1 kg of offshore petroleum in Mexico, according to the data for the process name "petroleum and gas production, offshore petroleum Cutoff, U," is 0.586267803438689 2 kg CO2 equivalent. This metric provides a measure of the greenhouse gases emitted, expressed in terms of the equivalent mass of carbon dioxide, during the production of offshore petroleum in the specified location.”</p>
20. testSponsorQuery10	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <p>1. User navigates to the chat panel at the bottom of the page.</p> <p>2. User types in “What is the GWP of producing 1 kg</p>	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <p>- 0.58627 kg CO2 eq, petroleum and gas production, offshore petroleum Cutoff, U_MX</p>	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <p>- The global warming potential (GWP) of producing 1 kg of offshore petroleum in Mexico, according to the</p>

Test ID	Description	Expected Results	Actual Results
	<p>offshore gasoline in Mexico?”</p> <p>3. User hits the ‘Enter’ key</p>		<p>retrieved process data for petroleum and gas production (Process ID: 8426fcf8-bc9b-3363-b531-f3e7c697b326), is 0.5862678034386892 kg CO2 equivalent. This value indicates the contribution to global warming from the production of 1 kg of offshore petroleum in this specific context.</p> <ul style="list-style-type: none"> - Process name: petroleum and gas production, offshore petroleum Cutoff, U. -
21. testSponsorQuery11	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of producing 10 kg offshore gasoline in Mexico?” 3. User hits the ‘Enter’ 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 5.8627 kg CO2 eq, petroleum and gas production, offshore petroleum Cutoff, U_MX 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The Global Warming Potential (GWP) for producing 1 kg of offshore petroleum in the specified process is 0.5862678034386892 kg CO2

Test ID	Description	Expected Results	Actual Results
	key		equivalent. Therefore, for producing 10 kg of offshore gasoline, the GWP would be 5.862678034386892 kg CO2 equivalent. The process used for this data is "petroleum and gas production, offshore petroleum Cutoff, U" located in Mexico.
22. testSponsorQuery12	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of producing 10 kg offshore gasoline?” 3. User hits the ‘Enter’ key 	<p>Provides the range from the following list:</p> <p>petroleum and gas production, offshore petroleum Cutoff, U - AE petroleum and gas production, offshore petroleum Cutoff, U - AR petroleum and gas production, offshore petroleum Cutoff, U - AU petroleum and gas production, offshore petroleum Cutoff, U - AZ petroleum and gas production, offshore petroleum Cutoff, U - BR petroleum and gas production, offshore petroleum Cutoff, U - CA petroleum and gas production, offshore petroleum Cutoff, U - CN petroleum and gas production, offshore petroleum Cutoff, U - CO petroleum and gas production, offshore petroleum Cutoff, U - DE petroleum and gas production, offshore petroleum Cutoff, U - EC petroleum and gas production, offshore petroleum Cutoff, U - EG petroleum and gas production, offshore petroleum Cutoff, U - GB petroleum and gas production, offshore petroleum Cutoff, U - ID petroleum and gas production, offshore petroleum Cutoff, U - IN petroleum and gas production, offshore petroleum Cutoff, U - IR petroleum and gas production, offshore petroleum Cutoff, U - IT petroleum and gas production, offshore petroleum Cutoff, U - KW petroleum and gas production, offshore petroleum Cutoff, U - KZ petroleum and gas production, offshore petroleum Cutoff, U - LV petroleum and gas production, offshore petroleum Cutoff, U - MX petroleum and gas production, offshore petroleum Cutoff, U - MY petroleum and gas production, offshore petroleum Cutoff, U - NG petroleum and gas production, offshore petroleum Cutoff, U - NL petroleum and gas production, offshore petroleum Cutoff, U - NO petroleum and gas production, offshore petroleum Cutoff, U - PE petroleum and gas production, offshore petroleum Cutoff, U - PL petroleum and gas production, offshore petroleum Cutoff, U - QA petroleum and gas production, offshore petroleum Cutoff, U - RO petroleum and gas production, offshore petroleum Cutoff, U - RU petroleum and gas production, offshore petroleum Cutoff, U - SA petroleum and gas production, offshore petroleum Cutoff, U - TH petroleum and gas production, offshore petroleum Cutoff, U - TM petroleum and gas production, offshore petroleum Cutoff, U - TR petroleum and gas production, offshore petroleum Cutoff, U - TT petroleum and gas production, offshore petroleum Cutoff, U - US petroleum and gas production, offshore petroleum Cutoff, U - VE</p>	<p>The chatbot displays the text “Generating” and then replies with the following content:</p> <ul style="list-style-type: none"> - (Missing some locations) To determine the Global Warming Potential (GWP) of producing 10 kg of offshore gasoline, we can use the processes from the dataset provided. Here are the GWP impacts for producing 1 kg of offshore petroleum (interpreted as gasoline for this context) from different locations:

Test ID	Description	Expected Results	Actual Results
			<div>1. KW (Kuwait): 0.229315367320 9569 kg CO2 eq. per kg</div> <div>2. GB (Great Britain): 0.202770524082 00666 kg CO2 eq. per kg</div> <div>3. US (United States): 0.387847135847 74594 kg CO2 eq. per kg</div> <div>4. EG (Egypt): 0.563575567700 9318 kg CO2 eq. per kg</div> <div>5. TT (Trinidad and Tobago): 0.415954406287 2464 kg CO2 eq. per kg</div> <div>6. NG (Nigeria): 0.746858600532 1952 kg CO2 eq. per kg</div> <div>For each location, the GWP for 10 kg of offshore gasoline production would be ten times the GWP impact given for 1 kg:</div> <div><ul style="list-style-type: none">Kuwait: 2.293153673209 569 kg CO2 eq. for 10 kg</div>

Test ID	Description	Expected Results	Actual Results
			<ul style="list-style-type: none">● Great Britain: 2.027705240820 0666 kg CO2 eq. for 10 kg● United States: 3.878471358477 4594 kg CO2 eq. for 10 kg● Egypt: 5.635755677009 318 kg CO2 eq. for 10 kg● Trinidad and Tobago: 4.159544062872 464 kg CO2 eq. for 10 kg● Nigeria: 7.468586005321 952 kg CO2 eq. for 10 kg <p>Presenting this data allows researchers to assess the GWP of producing 10 kg of offshore gasoline at different locations, depending on the geographical origin of production.</p> <p>Reason for failure: The vector search function did not accurately identify all the locations for which there is data</p>

Test ID	Description	Expected Results	Actual Results
23. testSponsorQuery13	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none">1. User navigates to the chat panel at the bottom of the page.2. User types in “What is the GWP of producing 1 kg low alloyed steel?”3. User hits the ‘Enter’ key	<p>Asks further if its converter or electric OR lists all the results:</p> <div><div>steel production, converter, low-alloyed steel, low-alloyed Cutoff, U - IN</div><div>steel production, converter, low-alloyed steel, low-alloyed Cutoff, U - RER</div><div>steel production, converter, low-alloyed steel, low-alloyed Cutoff, U - RoW</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - AT</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - CA-QC</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - CH</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - Europe without Switzerland and Austria</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - IN</div><div>steel production, electric, low-alloyed steel, low-alloyed Cutoff, U - RoW</div><div>steel production, low-alloyed, hot rolled steel, low-alloyed, hot rolled Cutoff, U - RER</div><div>steel production, low-alloyed, hot rolled steel, low-alloyed, hot rolled Cutoff, U - RoW</div></div>	<p>The chatbot displays the text “Generating” and then replies with the following content:</p> <ul style="list-style-type: none">- (Missing some results) The Global Warming Potential (GWP) of producing 1 kg of low-alloyed steel varies based on the production process used and the geographical location of the process. From the retrieved data:<ol style="list-style-type: none">1. Process Name: Steel production, converter, low-alloyed Steel, low-alloyed Cutoff, U.<ul style="list-style-type: none">o Location: RoW (Rest of the World)o GWP Impact: 2.3223 761427 01434 kg CO2 eq.

Test ID	Description	Expected Results	Actual Results
			<div>2. Process Name: Steel production, converter, low-alloyed Steel, low-alloyed Cutoff, U.</div> <div><div>○ Location: RER (Rest of Europe an Region)</div><div>○ GWP Impact: 2.1709 372735 67236 kg CO2 eq.</div></div> <div>The GWP figures indicate the CO2 equivalent emissions per kilogram of low-alloyed steel produced through the converter method in different regions. The rest of the world (RoW) process shows slightly higher emissions compared to the Rest of European Region (RER).</div> <div>Reason for failure: The vector search function is</div>

Test ID	Description	Expected Results	Actual Results
			only set up to retrieve the first most relevant process name and the system is not configured to ask the user follow up questions or search for two similar processes at once. This question would need to be split into two in order to work with the system.
24. testSponsorQuery14	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of transporting 1 t goods for 1 km via railway in the US?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - Averagely it should be 0.06103 kg CO2 eq, market for transport, freight train transport, freight train Cutoff, U_US 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The Global Warming Potential (GWP) of transporting 1 ton of goods for 1 kilometer via railway in the US is 0.061026672431983 49 kg CO2 equivalents. This data is specific to the process identified as "transport, freight train, diesel transport, freight train Cutoff, U."
25. testSponsorQuery15	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p>	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p>	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p>

Test ID	Description	Expected Results	Actual Results
	<ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of transporting 10 t goods for 10 km via railway in the US?” 3. User hits the ‘Enter’ key 	<ul style="list-style-type: none"> - Averagely it should be 610.3 kg CO2 eq, market for transport, freight train transport, freight train Cutoff, U_US 	<ul style="list-style-type: none"> - “The Global Warming Potential (GWP) of transporting goods via railway can be determined by analyzing the specific environmental impact data associated with the relevant process. For the transportation of freight using a diesel-powered train in the United States, the process identified is "transport, freight train, diesel transport, freight train Cutoff, U." - For this specific process, the global warming impact is measured at 0.061026672431983 49 kg CO2 eq per tonne-kilometer (t*km). Considering the user query to transport 10 tonnes of goods over a distance of 10 kilometers, this involves a total of 100

Test ID	Description	Expected Results	Actual Results
			<p>tonne-kilometers.</p> <p>Therefore, the total GWP for transporting the specified load is calculated by multiplying the impact per tkm by the total tkm involved:</p> <p>Global Warming Impact for transporting 10 tonnes over 10 km: 6.102667243198349 kg CO2 eq.</p> <p>This result provides an overview of the environmental impact in terms of greenhouse gas emissions for the given scenario based on the process data for transporting freight via railway in the U.S.”</p> <p>Reason for failure: The system works based on identifying processes that semantically match the question. However it has no reasoning capacity. Here the question is about an</p>

Test ID	Description	Expected Results	Actual Results
			average, but the system has no ability to deduce that the average can be found in a process with the “market for transport” name. Those two terms did not semantically match.
26. testSponsorQuery16	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of transporting 1 t goods for 1 km via truck, larger than 32 metric ton, with EURO6 emission standard?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.10243 kg CO2 eq, market for transport, freight, lorry >32 metric ton, EURO6 transport, freight, lorry >32 metric ton, EURO6 Cutoff, U_RoW 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - To determine the Global Warming Potential (GWP) of transporting 1 ton of goods over 1 kilometer using a truck larger than 32 metric tons with the EURO6 emission standard, we refer to the retrieved data on relevant transportation processes. <p>There are two primary geographical contexts provided for this transport process: "Rest of World (RoW)" and "Regionally Specific European Region (RER)." Both contexts reflect very similar GWP impacts:</p> <ol style="list-style-type: none"> 1. RoW Context:

Test ID	Description	Expected Results	Actual Results
			<p>The Global Warming Impact is 0.102434369284 80087 kilograms CO2 equivalent per ton-kilometer.</p> <p>2. RER Context: The Global Warming Impact is 0.101787589795 81948 kilograms CO2 equivalent per ton-kilometer.</p> <p>Thus, for the transportation of goods using a truck larger than 32 metric tons under the EURO6 standard, the GWP ranges from approximately 0.1018 to 0.1024 kg CO2 eq per t*km, depending on the geographical context.</p> <p>For precise environmental impact assessments, researchers could use either value based on the geographical scope of their study.</p> <p>The processes involved in this analysis are:</p> <ul style="list-style-type: none">• Transport, freight, lorry >32 metric ton,

Test ID	Description	Expected Results	Actual Results
			<p>EURO6 Cutoff, U in location RoW and RER, characterized by process IDs 20e0da0a-816e-3637-bede-a9ac4b000dca (RoW) and 7866c8b1-6484-31ef-9b09-b9d957952252 (RER).</p>
27. testSponsorQuery17	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the carbon footprint of 1kg tap water in South Africa?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.00144 kg CO2 eq, market for tap water tap water Cutoff, U_ZA 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - “The carbon footprint of producing 1 kg of tap water in South Africa, through conventional treatment processes, is 0.0005516135057768086 kg CO2 equivalent. This data corresponds to the process identified as "tap water production, conventional treatment tap water Cutoff, U" with Process ID

Test ID	Description	Expected Results	Actual Results
			<p>33dc6e40-acd9-331e-b225-786b814adf01, located in South Africa (ZA).”</p> <p>Reason for failure: Here the question is generic, and the system gave a reasonable response. However it is not the exact response expected in the expected result. The user did not specify a treatment method which led to this ambiguity.</p>
28. testSponsorQuery18	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the carbon intensity of landfilling 1 kg mixed waste plastics in Switzerland?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.08145 kg CO2 eq, treatment of waste plastic, mixture, sanitary landfill waste plastic, mixture Cutoff, U_CH 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - The carbon intensity of landfilling 1 kg of mixed waste plastics in Switzerland, according to the retrieved data, is quantified by the global warming impact, which is 0.08145375721786438 kg CO2 equivalent. This value provides a direct measure of the contribution to climate change associated with the

Test ID	Description	Expected Results	Actual Results
			<p>process of landfilling waste plastics. The data is derived from the process named "treatment of waste plastic, mixture, sanitary landfill waste plastic, mixture Cutoff, U," with the location specified as Switzerland (CH). This impact quantifies the greenhouse gas emissions specifically associated with this waste management practice, reflecting its environmental footprint in terms of carbon emissions.</p>
29. testSponsorQuery19	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the life-cycle GHG emissions of landfilling 1 kg mixed waste plastics in 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 0.08145 kg CO2 eq, treatment of waste plastic, mixture, sanitary landfill waste plastic, mixture Cutoff, U_CH 	<ul style="list-style-type: none"> - The life-cycle greenhouse gas (GHG) emissions associated with the landfilling of 1 kg of mixed waste plastics in Switzerland, according to the provided data, is quantified as 0.081453757217864 38 kg CO2 equivalent. This

Test ID	Description	Expected Results	Actual Results
	<p>Switzerland?”</p> <p>3. User hits the ‘Enter’ key</p>		<p>result is derived from an analysis of the environmental impacts from the process named: "treatment of waste plastic, mixture, sanitary landfill waste plastic, mixture Cutoff, U."</p> <p>This process reflects the emissions resulting from the management of waste plastics disposed of in a sanitary landfill setting within Switzerland, offering insights into the carbon footprint of such treatment.</p>
30. testSponsorQuery20	<p><u>Preconditions:</u> Test 2 completed (testRoleSelectionResearcher)</p> <ol style="list-style-type: none"> 1. User navigates to the chat panel at the bottom of the page. 2. User types in “What is the GWP of building a luxury hotel, global average?” 3. User hits the ‘Enter’ key 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - 2.19961 E8 kg CO2eq, market for building, luxury hotel building, luxury hotel Cutoff, U_GLO 	<p>The chatbot displays the text “Generating” and then replies with the required following content:</p> <ul style="list-style-type: none"> - “The Global Warming Potential (GWP) of the process "market for building, luxury hotel" (Process ID: 4e9f737a-da6a-35d2-ac20-467e21edddf9) located in the global average

Test ID	Description	Expected Results	Actual Results
			(GLO) is 219,961,403.891316 92 kg CO2 eq per item built. This quantifies the impact in terms of carbon dioxide equivalent emissions associated with constructing a luxury hotel.”

Testing Summary

Our current testing confirms that all core functional requirements have been met; however, we continue to observe accuracy issues in chatbot responses. These inconsistencies are largely due to ambiguous process names in the dataset and limitations of the current embedding model. Adding clear, consistent descriptions for each environmental process in the NameData.txt file—or in a supplementary reference file—would reduce ambiguity, support prompt template design, and improve interpretability. Additionally, upgrading to a newer embedding model would enhance semantic matching, especially for edge cases or underrepresented queries, helping ensure more reliable and contextually appropriate chatbot outputs. Ongoing acceptance testing will guide further refinements and help the next team identify and resolve remaining issues.

Other Testing - Usability Testing

Introduction

This usability test was conducted to evaluate the effectiveness and user experience of the chatbot in handling user inquiries about environmental data, generating reports based on user queries, validating user uploaded data, and generating summaries based on user data uploaded. The data returned in responses to users can be very technical in nature and there may be a struggle to interpret such reports or information. Also, with uploading data, users may upload incorrect data and the chatbot must be able to provide clear feedback and solutions. With all the variability in a chatbot, usability testing must be conducted to ensure the chatbot isn't abandoned and that there aren't any friction points. The think-aloud method was chosen to gather real-time insights into user behavior that traditional survey methods might miss and identify areas of improvement. By observing users as they interact with the chatbot, we can pinpoint exactly where and why usability issues arise, allowing for more targeted improvements.

Date Performed	4/24/2025
Researchers	Annabella Jackmore, Sid Kaju, Esha Mascarenhas, Cole Sanders, Madison Torres

Participants

To ensure diverse perspectives, we recruited 5 users, categorized as follows:

5 new participants - These will be new users who have not been exposed to the project and have no prior knowledge. The users will be completely new to the project and will have no involvement in design, development, or earlier testing phases of the project.

This distribution allows us to analyze how inexperienced users navigate the chatbot. This will reveal any potential usability gaps that might not be apparent when testing only a singular user.

Participant Name	Participant Type	Participant #	Facilitator
Sarah Allard	New Participant	Participant 1	Madison Torres
Jennifer Giordano	New Participant	Participant 2	Annabella Jackmore
Minty Pattarapuwadon	New Participant	Participant 3	Cole Sanders
Anish Ramaiya	New Participant	Participant 4	Sid Kaju
Shrey Tanna	New Participant	Participant 5	Esha Mascarenhas

Task Scenario

Each participant will be asked to complete four predefined tasks that represent common use cases for EcoValid:

1. Navigate through the Introduction Page
 - a. Goal: Ensure users can understand the purpose and features of EcoValid before interacting with the tool.

- b. Expected Behavior: Users scroll through and click through the introduction content, grasp the main objectives of EcoValid, and proceed to the next stage without confusion.
 - c. Potential Challenges: Users may skip or skim important explanations if the content is too dense or unengaging, or they may not realize they need to continue to the next step.
- 2. Select a role based on a more high-level response - Researcher Role
 - a. Goal: Evaluate how easily users can identify and select a role aligned with high-level, data-focused queries.
 - b. Expected Behavior: Users recognize the Researcher Role as suited for analytical, data-centric inquiries and select it without hesitation.
 - c. Potential Challenges: Users may be unsure which role best fits their intentions, especially if role descriptions are vague or too similar.
- 3. Ask the chatbot “Eco” an environmental impact question for a specific area based on Researcher Role:

What is the carbon footprint of 1kg of tap water?

- a. Goal: Evaluate how well the chatbot provides straightforward factual information in a high-level response.
 - b. Expected Behavior: The chatbot should quickly provide information about the environmental impact in the specified area that is more high-level for users to understand.
 - c. Potential Challenges: Users may struggle with phrasing the query correctly, or the chatbot might fail to provide location-based results accurately.
- 4. Select a role based on a more low-level response - Policymaker Role
 - a. Goal: Assess whether users can correctly associate policy-relevant concerns with the Policymaker Role and select it accordingly.
 - b. Expected Behavior: Users identify the Policymaker Role as appropriate for applied or decision-making questions and select it.
 - c. Potential Challenges: Users might confuse this with the Researcher Role or overlook the distinction between high-level and low-level information needs.
- 5. Ask the chatbot “Eco” an environmental impact question for a specific area based on Policy Maker Role:

What is the carbon footprint of 1kg of tap water?

- a. Goal: Evaluate how well the chatbot provides straightforward factual information in a low-level response.
 - b. Expected Behavior: The chatbot should quickly provide information about the environmental impact in the specified area that is more high-level for users to understand.
 - c. Potential Challenges: Users may struggle with phrasing the query correctly, or the chatbot might fail to provide location-based results accurately.

Procedure

Each participant went through the following structured testing process:

1. Introduction & Instructions
 - a. Users will be briefed on the purpose of the test and asked to “think aloud” while interacting with the chatbot.
 - b. They will be instructed to verbalize their thoughts, expectations, and any confusion they experience.
2. Task Execution
 - a. Users will attempt each of the four tasks while continuously narrating their thought process.
 - b. If any difficulties are encountered, users will be encouraged to express what they found confusing or frustrating.
3. Observer Role:
 - a. A facilitator will be observing each session, which will consist of one of the researchers. They will be taking detailed notes on:
 - i. Points where users hesitated or got stuck.
 - ii. Areas where users expressed confusion.
 - iii. Situations where users had to rephrase or retry their queries.
 - iv. Areas where users expressed praise.
4. Post-Test Interview:
 - a. After users complete the tasks listed in the Task Scenario section, users will be asked these follow-up questions:
 - i. “What did you find most intuitive about the chatbot?”
 - ii. “Was there anything that frustrated you?”
 - iii. How would you compare this chatbot to others you have used?”

Testing Script

**Italics read by Moderator*

Section 1: Introduction

Hi, thank you for joining us today! We're testing a prototype called EcoValid, which helps users explore the environmental impact of different processes. You'll be interacting with a demo interface, asking questions, and reviewing data and visualizations. Through this, you will be helping us test the design. Please feel free to think out loud and let us know what you're thinking as you go. Ready to begin?

Section 2: Pre-Test Questions

Have you used any tools that track sustainability or environmental impact before?

On a scale of 1–5, how comfortable are you using data dashboards or chatbots?

Have you ever used the software openLCA to get environmental data?

Section 3: Tasks

Task 1 - Navigate through Introduction Page

This is EcoValid, Please take a moment to look over and navigate through the introduction page. Tell me what you notice, and what you think this tool is about.

(After the participant has gone through the Introduction Page)

Did you find this page helpful?

Was it hard to navigate through this page?

Task 2 - Select a Researcher Role

Now, imagine you're a scientist at the EPA who wants more information about the carbon footprint of water. Select the appropriate role to get started.

Task 3 - Start a Query with the Researcher Role

As a scientist, you are curious about the effects of switching urban areas to tap water from well water. You are wanting to know how 1 kg of tap water would affect the carbon footprint in different areas. Use the interface to ask Eco about this and start a conversation about it.

Task 4 - Switch Role to Policymaker

Now, imagine that you are a politician in the United States, who is making new regulation on tap water expansion. Select the appropriate role to get started.

Task 5 - Start a Query with the Policymaker Role

As a politician, you are curious about how switching to tap water in your city would affect the carbon footprint by just 1 kg by comparing it to different areas that have implemented it. Use the interface to ask Eco about this and start a conversation about it.

(After the participant has gone through the second query)

Did you notice a clear difference between the two responses?

Did you feel that you understood a certain response better than the other?

Section 4: Post-Task Questions (after each task)

Was that task easy or difficult?

What part (if any) confused you?

On a scale from 1–5, how satisfied were you with this part of the experience?

Section 5: Final Questions

What did you like most about using EcoValid?

What would you improve or change?

Would you consider using a tool like this in your daily life or work?

Data Collection

1. Observer Notes
 - a. Observers documented user reactions in real time, identifying any issues that arise.
 - i. For example, if multiple users struggle with choosing a role, this would be flagged as a design issue.
2. Completion Time Analysis
 - a. Observers will document how much time taken to complete each task.
 - b. Longer completion times indicate potential usability issues.
3. User Frustration Analysis
 - a. If there are any frustrations expressed (i.e. sighing, excessive retyping, direct complaints), this will be noted.
 - b. Specific chatbot responses that triggered frustration were analyzed for improvement.

Observations

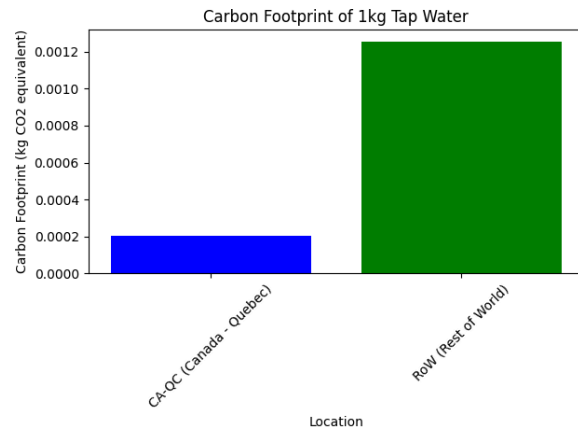
Question/Action	Observed - Participant # 1
Pre-Test Questions	
Have you used any tools that track sustainability or environmental impact before?	No
On a scale of 1–5, how comfortable are you using data dashboards or chatbots?	4

Have you ever used the software openLCA to get environmental data?	No
Task 1 - Navigate Through Introduction Page	
Tell me what you notice, and what you think this tool is about.	<ul style="list-style-type: none"> - Noticed the cute icon and commented on the page setup looking “well-structured” - Tool description: “The tool is about asking questions about the environment to a chatbot that is assumed to be pulled from an environmental database.”
Did you find this page helpful?	Yes
Was it hard to navigate through this page?	No
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 2 - Select a Researcher Role	
Selected proper role?	Yes, Researcher
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 3 - Start a Query with Researcher Role	
Query submitted:	“Summary of carbon footprint of 1 kg of tap water.”
Output based on query:	“Eco(GPT): The carbon footprint of 1 kg of tap water is relatively low compared to other commodities and activities, mainly because

	<p>the processes involved are less energy-intensive. The footprint primarily stems from the energy required for water treatment and distribution. Here's a summary of key factors contributing to the carbon footprint of tap water:</p> <p>Water Treatment:</p> <p>Energy Use: The treatment of water involves processes like filtration, disinfection, and sometimes desalination, all of which require energy. The extent of treatment depends on the source and the quality of the water.</p> <p>Chemical Use: Chemicals used in the treatment process, such as chlorine for disinfection, contribute indirectly through their production and transportation.</p> <p>Distribution:</p> <p>Energy is required to pump water through infrastructure to end consumers. The distance and elevation changes can impact energy usage.</p> <p>Maintenance of the infrastructure also adds a minor footprint, stemming from repair operations and materials.</p> <p>Infrastructure:</p> <p>The construction and maintenance of water infrastructure (pipes, treatment plants) have an embodied carbon footprint. However, this is spread over many years of use.</p> <p>Regional Variations:</p> <p>The carbon footprint of tap water can vary significantly depending on the region due to differences in energy sources (renewable vs. fossil fuels) and the efficiency of the water systems.</p> <p>Overall, the carbon footprint per kilogram (or</p>
--	---

	liter) of tap water is generally very small, often estimated to be under 0.1 grams of CO2 equivalent per liter. This makes tap water a more environmentally friendly option compared to bottled water, which has a much higher footprint due to packaging and transportation. Always consider contextual variations and technological advancements that could slightly shift these figures.”
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	3
Task 4 - Select a Policymaker Role	
Selected proper role?	Yes, Policymaker
Was that task easy or difficult?	Easy
What part (if any) confused you?	No
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 5 - Start a Query with Policymaker Role	
Query submitted:	“What is the carbon footprint of 1kg of tap water?”
Output based on query:	“Eco(RAG): The carbon footprint of tap water refers to the total amount of carbon dioxide (CO2) emissions associated with the production and distribution of 1 kilogram of tap water. This is an important environmental metric as it helps policy makers and environmental advocates understand the impact of water supply systems on global warming. Here, we provide two examples from different locations to illustrate this

	<p>concept:</p> <p>Process Name: Market for tap water tap water Cutoff, U.</p> <p>Location: CA-QC (Canada - Quebec)</p> <p>Carbon Footprint: 0.000205674 kg CO2 equivalent.</p> <p>This value represents the contribution to global warming by releasing this amount of CO2 during the production and distribution process for 1 kg of tap water in Quebec.</p> <p>Process Name: Market for tap water tap water Cutoff, U.</p> <p>Location: RoW (Rest of World)</p> <p>Carbon Footprint: 0.001255279 kg CO2 equivalent.</p> <p>This is the carbon footprint associated with global water supply infrastructure other than specific regions, highlighting a different impact compared to Quebec, reflecting variations in local processes and efficiencies.</p> <p>For policy makers, it is crucial to understand that a lower carbon footprint indicates a more environmentally friendly process, reducing the potential contribution to global climate change. By comparing these figures, we can better evaluate the efficiency and environmental impact of regional water supply systems and strive for improvements that could further lower emissions.</p> <p>Visualization:</p>
--	---



Thinking Trejectory:

LLM input: Based on the following input, classify the user's intent as one of the following categories: "query" if the user is asking a technical question about environmental impact data. "validate" if the user is requesting validation of environmental impact data they have provided. "neither" if the input does not fit into either category. Provide only the category name as the answer. Provide the correct answer for each location. Input: What is the carbon footprint of 1kg of tap water?

LLM response: query

RAG input: What is the carbon footprint of 1kg of tap water?

RAG output: Processes: Process name: market for tap water | tap water. Location: CA-QC Process name: market for tap water | tap water. Location: RoW Process name: market for tap water | tap water. Location: CO Process name: market for tap water | tap water. Location: Europe Process name: tap water production, direct filtration treatment | tap water. Location: CO Process name: tap water production, direct filtration treatment | tap water. Location: CA-QC Process name: market for tap water | tap water. Location: BR Process name: tap water production, artificial

	<p>recharged wells tap water. Location: CA-QC Process name: market for tap water tap water. Location: ZA Process name: tap water production, conventional treatment tap water. Location: CA-QC LLM input: Given this query, return just the verbatim text of the process that best matches the query, minus its location. Query:What is the carbon footprint of 1kg of tap water? Processes: Process name: market for tap water tap water. Location: CA-QC Process name: market for tap water tap water. Location: RoW Process name: market for tap water tap water. Location: CO Process name: market for tap water tap water. Location: Europe Process name: tap water production, direct filtration treatment tap water. Location: CO Process name: tap water production, direct filtration treatment tap water. Location: CA-QC Process name: market for tap water tap water. Location: BR Process name: tap water production, artificial recharged wells tap water. Location: CA-QC Process name: market for tap water tap water. Location: ZA Process name: tap water production, conventional treatment tap water. Location: CA-QC LLM output: Processes: Process name: market for tap water tap water. Location: CA-QC Process name: market for tap water tap water. Location: RoW Process name: market for tap water tap water. Location: CO Process name: market for tap water tap water. Location: Europe Process name: tap water production, direct filtration treatment tap water. Location: CO Process name: tap water production, direct filtration treatment tap water. Location: CA-QC Process name: market for tap water tap water. Location: BR Process name: tap water production, artificial</p>
--	---

	<p>recharged wells tap water. Location: CA-QC Process name: market for tap water tap water. Location: ZA Process name: tap water production, conventional treatment tap water. Location: CA-QC RAG input: Provide all the locations for each instance of the following process.List them in the format "Location1,Location2,etc." Process Name: Processes: Process name: market for tap water tap water. Location: CA-QC Process name: market for tap water tap water. Location: RoW Process name: market for tap water tap water. Location: CO Process name: market for tap water tap water. Location: Europe Process name: tap water production, direct filtration treatment tap water. Location: CO Process name: tap water production, direct filtration treatment tap water. Location: CA-QC Process name: market for tap water tap water. Location: BR Process name: tap water production, artificial recharged wells tap water. Location: CA-QC Process name: market for tap water tap water. Location: ZA Process name: tap water production, conventional treatment tap water. Location: CA-QC RAG output:CA-QC,RoW,Europe Matched Process Data:Process name: market for tap water tap water Cutoff, U. Process ID: 82903222-7cc1-3998-9e09-cc6ee15ea29a. Location: CA-QC. Fossil fuel depletion impact: 0.00017397352551936442 MJ surplus. Respiratory effects impact: 2.3099973194447373e-07 kg PM2.5 eq. Global warming impact: 0.00020567404367292623 kg CO2 eq. Carcinogenics impact: 1.4736551071803704e-09 CTUh. Smog impact: 1.4055914991839349e-05 kg O3 eq. Acidification impact: 9.47560658803483e-07</p>
--	---

	<p>kg SO2 eq. Non carcinogenics impact: 1.919052402263651e-10 CTUh. Ozone depletion impact: 2.638909385718263e-12 kg CFC-11 eq. Eutrophication impact: 6.859822600775503e-07 kg N eq. Ecotoxicity impact: 0.020446748873728073 CTUe. Process output: 1.0 kg.</p> <p>Matched Process Data: Process name: market for tap water tap water Cutoff, U. Process ID: 6fe4478a-8c99-3ed9-a912-8e0140e9c2aa. Location: RoW. Fossil fuel depletion impact: 0.0009626851384774173 MJ surplus. Respiratory effects impact: 1.5987734833585335e-06 kg PM2.5 eq. Global warming impact: 0.0012552788469849527 kg CO2 eq. Carcinogenics impact: 2.7473759959431858e-09 CTUh. Smog impact: 7.843340003133836e-05 kg O3 eq. Acidification impact: 5.678399281406367e-06 kg SO2 eq. Non carcinogenics impact: 5.66541061452198e-10 CTUh. Ozone depletion impact: 2.8870978007673556e-10 kg CFC-11 eq. Eutrophication impact: 4.863282909069511e-06 kg N eq. Ecotoxicity impact: 0.04452041192174379 CTUe. Process output: 1.0 kg.</p> <p>Matched Process Data:</p> <p>LLM input: You are answering a question about environmental impact using real data. Use the retrieved information below to craft your response, ensuring accuracy. Retrieved Data: Processes: Process name: market for tap water tap water. Location: CA-QC Process name: market for tap water tap water. Location: RoW Process name: market for tap water tap water. Location: CO Process name: market for tap water tap water. Location: Europe Process name: tap water production,</p>
--	---

	<p>direct filtration treatment tap water. Location: CO Process name: tap water production, direct filtration treatment tap water. Location: CA-QC Process name: market for tap water tap water. Location: BR Process name: tap water production, artificial recharged wells tap water. Location: CA-QC Process name: market for tap water tap water. Location: ZA Process name: tap water production, conventional treatment tap water. Location: CA-QC Process name: market for tap water tap water Cutoff, U. Process ID: 82903222-7cc1-3998-9e09-cc6ee15ea29a. Location: CA-QC. Fossil fuel depletion impact: 0.00017397352551936442 MJ surplus. Respiratory effects impact: 2.3099973194447373e-07 kg PM2.5 eq. Global warming impact: 0.00020567404367292623 kg CO2 eq. Carcinogenics impact: 1.4736551071803704e-09 CTUh. Smog impact: 1.4055914991839349e-05 kg O3 eq. Acidification impact: 9.47560658803483e-07 kg SO2 eq. Non carcinogenics impact: 1.919052402263651e-10 CTUh. Ozone depletion impact: 2.638909385718263e-12 kg CFC-11 eq. Eutrophication impact: 6.859822600775503e-07 kg N eq. Ecotoxicity impact: 0.020446748873728073 CTUe. Process output: 1.0 kg. Process name: market for tap water tap water Cutoff, U. Process ID: 6fe4478a-8c99-3ed9-a912-8e0140e9c2aa. Location: RoW. Fossil fuel depletion impact: 0.0009626851384774173 MJ surplus. Respiratory effects impact: 1.5987734833585335e-06 kg PM2.5 eq. Global warming impact: 0.0012552788469849527 kg CO2 eq. Carcinogenics impact: 2.7473759959431858e-09 CTUh. Smog</p>
--	--

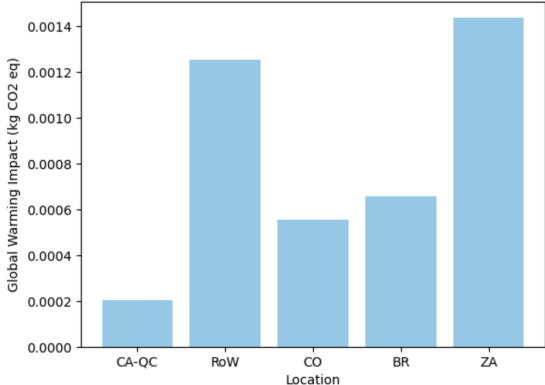
	<p>impact: 7.843340003133836e-05 kg O3 eq. Acidification impact: 5.678399281406367e-06 kg SO2 eq. Non carcinogenics impact: 5.66541061452198e-10 CTUh. Ozone depletion impact: 2.8870978007673556e-10 kg CFC-11 eq. Eutrophication impact: 4.863282909069511e-06 kg N eq. Ecotoxicity impact: 0.04452041192174379 CTUe. Process output: 1.0 kg.</p> <p>User query:What is the carbon footprint of 1kg of tap water? Answer the user query factually without making assumptions. Craft your answer to a policy maker audience and explain all technical terms and units. Do not show equations. Show the process name and location in your answer.</p> <p>LLM output: The carbon footprint of tap water refers to the total amount of carbon dioxide (CO2) emissions associated with the production and distribution of 1 kilogram of tap water. This is an important environmental metric as it helps policy makers and environmental advocates understand the impact of water supply systems on global warming. Here, we provide two examples from different locations to illustrate this concept:</p> <p>Process Name: Market for tap water tap water Cutoff, U. Location: CA-QC (Canada - Quebec) Carbon Footprint: 0.000205674 kg CO2 equivalent. This value represents the contribution to global warming by releasing this amount of CO2 during the production and distribution process for 1 kg of tap water in Quebec.</p>
--	--

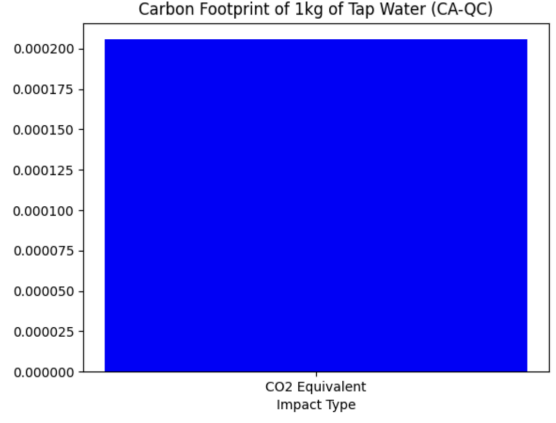
	<p>Process Name: Market for tap water tap water Cutoff, U.</p> <p>Location: RoW (Rest of World)</p> <p>Carbon Footprint: 0.001255279 kg CO2 equivalent.</p> <p>This is the carbon footprint associated with global water supply infrastructure other than specific regions, highlighting a different impact compared to Quebec, reflecting variations in local processes and efficiencies.</p> <p>For policy makers, it is crucial to understand that a lower carbon footprint indicates a more environmentally friendly process, reducing the potential contribution to global climate change. By comparing these figures, we can better evaluate the efficiency and environmental impact of regional water supply systems and strive for improvements that could further lower emissions.”</p>
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	2
Did you notice a clear difference between the two responses?	Yes, the researcher role had more information than the policymaker role, even though there were no numbers.
Did you feel that you understood a certain response better than the other?	Understood the researcher role response more because of more information with the response.
Final Questions	
What did you like most about using EcoValid?	Noted that the interface was the participant’s favorite part of the application. Felt that it was easy to use and very pretty.

What would you improve or change?	<p>The participant discussed that the chatbot gave them “what they needed”.</p> <p>Suggestion:</p> <ul style="list-style-type: none"> - More followup after questions - Referenced ChatGPT’s function after a query is asked, responds with “Let me know if there is anything else I can help with!”
Would you consider using a tool like this in your daily life or work?	Yes, if my job or daily life needed environmental information.

Question/Action	Observed - Participant # 2
Pre-Test Questions	
Have you used any tools that track sustainability or environmental impact before?	No
On a scale of 1–5, how comfortable are you using data dashboards or chatbots?	3/4
Have you ever used the software openLCA to get environmental data?	No
Task 1 - Navigate Through Introduction Page	
Tell me what you notice, and what you think this tool is about.	<ul style="list-style-type: none"> - Looks like an environmentally friendly app because it’s green - It’s AI powered - Tool description: “The tool is about educating people about the environment.”
Did you find this page helpful?	Yes
Was it hard to navigate through this page?	No
Was that task easy or difficult?	Easy

What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 2 - Select a Researcher Role	
Selected proper role?	Yes, Researcher
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 3 - Start a Query with Researcher Role	
Query submitted:	“What is the carbon footprint of 1kg of tap water?”
Output based on query:	<p>“The carbon footprint of 1 kg of tap water varies depending on the location and the specific process used. Here's a summary of the global warming impacts for different tap water markets based on the retrieved data:</p> <ol style="list-style-type: none"> 1. Market for tap water tap water Cutoff, U Location: CA-QC: <ul style="list-style-type: none"> ○ Global warming impact: 0.000205674 kg CO2 eq. 2. Market for tap water tap water Cutoff, U Location: RoW: <ul style="list-style-type: none"> ○ Global warming impact: 0.001255279 kg CO2 eq. 3. Market for tap water tap water Cutoff, U Location: CO: <ul style="list-style-type: none"> ○ Global warming impact: 0.000556628 kg CO2 eq. 4. Market for tap water tap water Cutoff, U Location: BR: <ul style="list-style-type: none"> ○ Global warming impact:

	<div>0.000656367 kg CO2 eq.</div> <div>5. Market for tap water tap water Cutoff, U Location: ZA:<ul style="list-style-type: none">Global warming impact: 0.001437038 kg CO2 eq.</div> <div>This data provides insight into the carbon footprint of tap water production across different regions, useful for researchers analyzing environmental impacts of water usage.</div> <div>Visualization:</div> <div><div>Carbon Footprint of 1kg Tap Water in Different Regions</div><table border="1"><thead><tr><th>Location</th><th>Global Warming Impact (kg CO2 eq)</th></tr></thead><tbody><tr><td>CA-QC</td><td>0.0002</td></tr><tr><td>RoW</td><td>0.00125</td></tr><tr><td>CO</td><td>0.00055</td></tr><tr><td>BR</td><td>0.00065</td></tr><tr><td>ZA</td><td>0.001437</td></tr></tbody></table></div> <div>”</div>	Location	Global Warming Impact (kg CO2 eq)	CA-QC	0.0002	RoW	0.00125	CO	0.00055	BR	0.00065	ZA	0.001437
Location	Global Warming Impact (kg CO2 eq)												
CA-QC	0.0002												
RoW	0.00125												
CO	0.00055												
BR	0.00065												
ZA	0.001437												
Was that task easy or difficult?	Easy												
What part (if any) confused you?	None												
On a scale from 1–5, how satisfied were you with this part of the experience?	5												
Task 4 - Select a Policymaker Role													
Selected proper role?	Yes, Policymaker												
Was that task easy or difficult?	Easy												

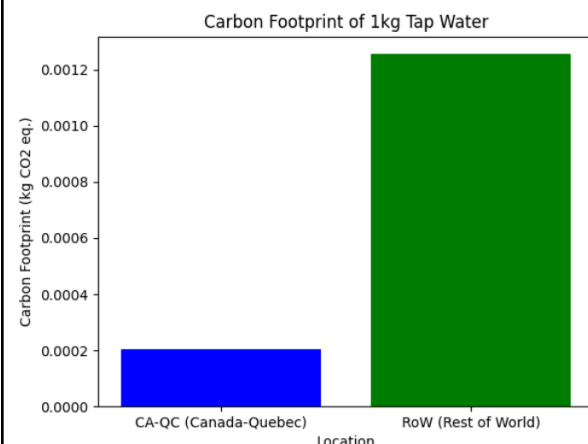
What part (if any) confused you?	No
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 5 - Start a Query with Policymaker Role	
Query submitted:	“What is the global warming potential of generating 1 kWh electricity (high voltage) via geothermal in the southeastern US?”
Output based on query:	<p>“ The carbon footprint of 1 kg of tap water varies depending on the location and treatment process. Specifically, for the process named "market for tap water tap water" located in CA-QC, the global warming impact, which is a measure of the carbon footprint, is 0.00020567404367292623 kg of CO2 equivalent per kilogram of tap water. This data point provides a useful benchmark for policymakers considering the environmental impact of tap water production in different regions.</p> <p>Visualization:</p>  <p>”</p>
Was that task easy or difficult?	Easy
What part (if any) confused you?	None

On a scale from 1–5, how satisfied were you with this part of the experience?	5
Did you notice a clear difference between the two responses?	Yes, the researcher role had more information and was more structured than the policymaker role.
Did you feel that you understood a certain response better than the other?	Understood the researcher role response more because of more information with the response.
Final Questions	
What did you like most about using EcoValid?	Liked that is specializes in environmental questions and that you can change roles to get different responses.
What would you improve or change?	Suggestion: - Having the policymaker answer contain more information
Would you consider using a tool like this in your daily life or work?	Yes.

Question/Action	Observed - Participant # 3
Pre-Test Questions	
Have you used any tools that track sustainability or environmental impact before?	No
On a scale of 1–5, how comfortable are you using data dashboards or chatbots?	4
Have you ever used the software openLCA to get environmental data?	No

Task 1 - Navigate Through Introduction Page	
Tell me what you notice, and what you think this tool is about.	<ul style="list-style-type: none"> - It describes itself as an AI assistant for something environmental - It mentions data - Seems to be a tool for environmental data
Did you find this page helpful?	Yes
Was it hard to navigate through this page?	No
Was that task easy or difficult?	“Not hard at all”
What part (if any) confused you?	“It wasn’t confusing”
On a scale from 1–5, how satisfied were you with this part of the experience?	4
Task 2 - Select a Researcher Role	
Selected proper role?	Researcher
Was that task easy or difficult?	Easy
What part (if any) confused you?	“Nothing”
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 3 - Start a Query with Researcher Role	
Query submitted:	“What is the carbon footprint of 1kg of tap water?”
Output based on query:	<p>“Eco(RAG): The carbon footprint of 1 kg of tap water can vary depending on the geographical location of the process. Based on the retrieved data, the global warming impact, expressed as CO2 equivalent for 1 kg of tap water, is as follows:</p> <p>Process Name: market for tap water tap</p>

	<p>water Cutoff, U.</p> <p>Location: CA-QC (Canada-Quebec)</p> <p>Carbon Footprint: 0.00020567404367292623 kg CO2 eq.</p> <p>Process Name: market for tap water tap water Cutoff, U.</p> <p>Location: RoW (Rest of World)</p> <p>Carbon Footprint: 0.0012552788469849527 kg CO2 eq.</p> <p>These values represent the global warming impact due to the production and market distribution of 1 kg of tap water in those respective locations. Researchers should consider these variations while studying environmental impacts or planning sustainability strategies.</p> <p>Visualization:</p>
--	--

	 <p>Carbon Footprint of 1kg Tap Water</p> <table border="1"> <thead> <tr> <th>Location</th> <th>Carbon Footprint (kg CO2 eq.)</th> </tr> </thead> <tbody> <tr> <td>CA-QC (Canada-Quebec)</td> <td>0.0002</td> </tr> <tr> <td>RoW (Rest of World)</td> <td>0.00125</td> </tr> </tbody> </table>	Location	Carbon Footprint (kg CO2 eq.)	CA-QC (Canada-Quebec)	0.0002	RoW (Rest of World)	0.00125
Location	Carbon Footprint (kg CO2 eq.)						
CA-QC (Canada-Quebec)	0.0002						
RoW (Rest of World)	0.00125						
Was that task easy or difficult?	Easy						
What part (if any) confused you?	“It took me a second to understand the answer”						
On a scale from 1–5, how satisfied were you with this part of the experience?	5						
Task 4 - Select a Policymaker Role							
Selected proper role?	Yes, Policymaker						
Was that task easy or difficult?	Easy						
What part (if any) confused you?	No						
On a scale from 1–5, how satisfied were you with this part of the experience?	5						
Task 5 - Start a Query with Policymaker Role							
Query submitted:	“What is the carbon footprint of 1kg of tap water?”						
Output based on query:	“Eco(RAG): The carbon footprint of 1 kilogram of tap water can vary depending on the region where the water is sourced and the treatment processes involved. In environmental terms, the carbon footprint is						

measured by the amount of carbon dioxide (CO₂) equivalent emissions generated during the production and supply of a product, in this case, tap water.

For the specific processes listed, here are some relevant details:

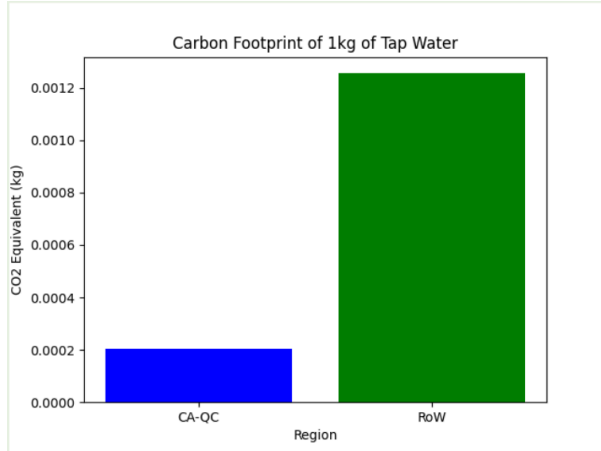
Process Name: Market for tap water | tap water. Location: CA-QC

Global Warming Impact: 0.000205674 kilograms of CO₂ equivalent per kilogram of tap water. This figure indicates the amount of CO₂ that would have the same global warming potential as the gases emitted during this process.

Process Name: Market for tap water | tap water. Location: RoW (Rest of the World)

Global Warming Impact: 0.001255279 kilograms of CO₂ equivalent per kilogram of tap water. This represents a higher carbon footprint compared to the location in CA-QC, possibly due to different energy sources or treatment methodologies used in broader regions outside Canada-Quebec.

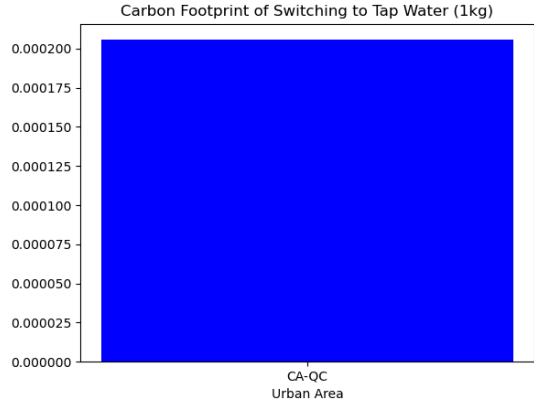
It is important for policymakers to consider these regional differences in carbon emissions, as they can inform strategies for reducing greenhouse gas emissions associated with water supply systems. Such strategies might include investing in more efficient water treatment technologies or transitioning

	<p>to renewable energy sources to power these systems, ultimately helping to mitigate climate change impacts.</p>  <p>“</p>
Was that task easy or difficult?	“Easy”
What part (if any) confused you?	“Nothing”
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Did you notice a clear difference between the two responses?	“Yes, the policymaker role explained and broke things down more.”
Did you feel that you understood a certain response better than the other?	“I understood both. It felt like the researcher perspective was more factual while the policymaker perspective was focused on how the data contributes to reducing the carbon footprint.”
Final Questions	
What did you like most about using EcoValid?	“It was easy to use. Instead of looking this stuff up online it was easy to access in a chatbot format.”
What would you improve or change?	“Perhaps add a role for people who are just interested in learning more about climate

	impacts.”
Would you consider using a tool like this in your daily life or work?	Yes

Question/Action	Observed - Participant # 4
Pre-Test Questions	
Have you used any tools that track sustainability or environmental impact before?	No
On a scale of 1–5, how comfortable are you using data dashboards or chatbots?	5
Have you ever used the software openLCA to get environmental data?	No
Task 1 - Navigate Through Introduction Page	
Tell me what you notice, and what you think this tool is about.	<ul style="list-style-type: none"> - Noticed that all the info needed to use the website is available and easy to read - It looks an AI assistant that helps you get scientifically backed data for your environmental questions
Did you find this page helpful?	Yes
Was it hard to navigate through this page?	No
Was that task easy or difficult?	Easy
What part (if any) confused you?	Nothing
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 2 - Select a Researcher Role	

Selected proper role?	Yes
Was that task easy or difficult?	Easy
What part (if any) confused you?	Nothing
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 3 - Start a Query with Researcher Role	
Query submitted:	I am studying the effects of different water usage in urban areas and their respective carbon footprints. Tell me how 1kg of tap water would affect the footprint in different urban areas if they switched over from well water.
Output based on query:	<p>Eco(RAG): When considering the environmental impact of using tap water instead of well water, specifically in terms of carbon footprint, the process data available for tap water in different regions indicates varying impacts. Unfortunately, the dataset only contains explicit carbon footprint information for one specific process, making comparative analysis limited.</p> <p>For the process "market for tap water tap water Cutoff, U" located in CA-QC, the global warming impact when providing 1 kg of tap water is recorded at 0.000205674 kg CO2 equivalent. This data provides a reference point for how urban areas transitioning from well water to this specific tap water source would contribute to carbon emissions.</p> <p>To obtain a complete analysis across different locations and types of water sources, comparable data for all processes would be necessary. Nonetheless, this CA-QC process demonstrates a concrete carbon footprint</p>

	<p>measure, essential for understanding the environmental trade-offs of switching water sources in urban contexts.</p> <p>Visualization:</p> 
Was that task easy or difficult?	Easy
What part (if any) confused you?	Nothing
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 4 - Select a Policymaker Role	
Selected proper role?	Yes
Was that task easy or difficult?	Easy
What part (if any) confused you?	Nothing
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 5 - Start a Query with Policymaker Role	
Query submitted:	I want to switch over the city's water to tap water from well water, and am curious of its effects on the environment. How would changing it by 1kg change its carbon footprint on the environment?

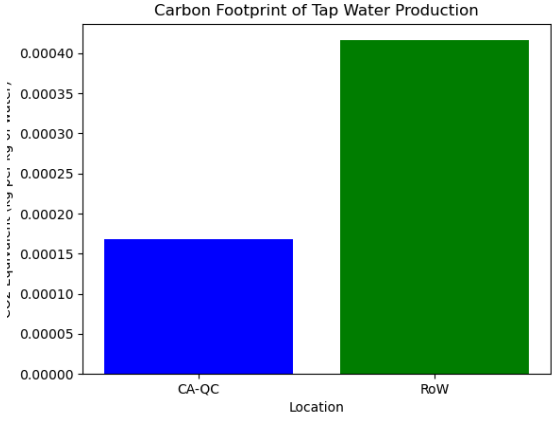
Output based on query:

Eco(RAG): As a policy maker considering the environmental impact of switching your city's water supply to tap water from well water, it's important to understand how this might affect the carbon footprint. Focusing on the production of tap water from artificially recharged wells, the impact varies depending on the specific location of the process. For example, the process titled "tap water production, artificial recharged wells | tap water" located in Quebec, Canada (CA-QC) has a global warming impact of 0.0001677 kg of CO₂ equivalent per kilogram of tap water produced. This means for each kilogram of tap water produced by this method in CA-QC, approximately 0.0001677 kg of carbon dioxide (a greenhouse gas contributing to global warming) is emitted into the atmosphere.

In contrast, the same process located in Rest of the World (RoW) has a higher global warming impact, measuring at 0.0004162 kg of CO₂ equivalent per kilogram. This indicates a greater contribution to the carbon footprint per kilogram of water produced as compared to the CA-QC location.

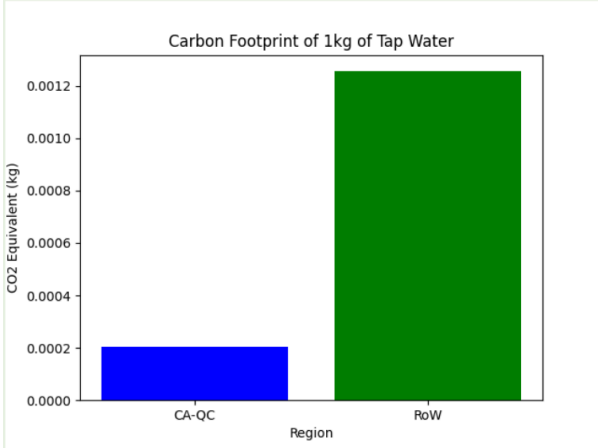
The term "CO₂ equivalent" is used to express the impact of different greenhouse gases in a common unit, making it easier to compare their potential contributions to climate change. Thus, the carbon footprint represents the total greenhouse gas emissions caused, directly or indirectly, by a process.

Making an informed decision on your city's water sourcing involves considering these environmental impacts and potentially choosing regions or methods with lower carbon footprints to reduce overall greenhouse gas emissions.

	<p>Visualization:</p>  <table border="1"> <caption>Carbon Footprint of Tap Water Production</caption> <thead> <tr> <th>Location</th> <th>CO₂ equivalent kg per kg of water</th> </tr> </thead> <tbody> <tr> <td>CA-QC</td> <td>~0.00017</td> </tr> <tr> <td>RoW</td> <td>~0.00041</td> </tr> </tbody> </table>	Location	CO ₂ equivalent kg per kg of water	CA-QC	~0.00017	RoW	~0.00041
Location	CO ₂ equivalent kg per kg of water						
CA-QC	~0.00017						
RoW	~0.00041						
Was that task easy or difficult?	Easy						
What part (if any) confused you?	Nothing						
On a scale from 1–5, how satisfied were you with this part of the experience?	5						
Did you notice a clear difference between the two responses?	The researcher's answer had way more technical terms and you would have to be in the industry to understand it fully. The policymaker's response explained it better for the general audience.						
Did you feel that you understood a certain response better than the other?	Yes the tester understood the policymaker answer better						
Final Questions							
What did you like most about using EcoValid?	<ul style="list-style-type: none"> - It gave a visualization of the data to make it easier to understand 						
What would you improve or change?	<ul style="list-style-type: none"> - Make the chatbot faster at providing answers - Highlight the key points from the answer 						
Would you consider using a tool like this in your daily life or work?	Yes if the tester's work was related to this field						

Question/Action	Observed - Participant # 5
Pre-Test Questions	
Have you used any tools that track sustainability or environmental impact before?	No
On a scale of 1–5, how comfortable are you using data dashboards or chatbots?	3
Have you ever used the software openLCA to get environmental data?	No
Task 1 - Navigate Through Introduction Page	
Tell me what you notice, and what you think this tool is about.	<ul style="list-style-type: none"> - I noticed the roles button at the top - Tool description: “The tool is about asking questions about the environment to a chatbot ”
Did you find this page helpful?	Yes
Was it hard to navigate through this page?	No
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 2 - Select a Researcher Role	
Selected proper role?	Yes, Researcher
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you	5

with this part of the experience?	
Task 3 - Start a Query with Researcher Role	
Query submitted:	“Summary of carbon footprint of 1 kg of tap water.”
Output based on query:	<p>“Sorry! We do not have that data.</p> <p>Visualization: (Not needed for this answer)”</p>
Was that task easy or difficult?	Easy
What part (if any) confused you?	None
On a scale from 1–5, how satisfied were you with this part of the experience?	1
Task 4 - Select a Policymaker Role	
Selected proper role?	Yes, Policymaker
Was that task easy or difficult?	Easy
What part (if any) confused you?	No
On a scale from 1–5, how satisfied were you with this part of the experience?	5
Task 5 - Start a Query with Policymaker Role	
Query submitted:	“What is the carbon footprint of 1kg of tap water?”
Output based on query:	<p>“Eco(RAG): The carbon footprint of 1 kg of tap water can vary depending on the geographical location of the process. Based on the retrieved data, the global warming impact, expressed as CO2 equivalent for 1 kg of tap water, is as follows:</p> <p>Process Name: market for tap water tap water Cutoff, U.</p>

	<p>Location: CA-QC (Canada-Quebec)</p> <p>Carbon Footprint: 0.00020567404367292623 kg CO₂ eq.</p> <p>Process Name: market for tap water tap water Cutoff, U.</p> <p>Location: RoW (Rest of World)</p> <p>Carbon Footprint: 0.0012552788469849527 kg CO₂ eq.</p> <p>These values represent the global warming impact due to the production and market distribution of 1 kg of tap water in those respective locations. Researchers should consider these variations while studying environmental impacts or planning sustainability strategies.</p>  <table border="1"><caption>Carbon Footprint of 1kg of Tap Water</caption><thead><tr><th>Region</th><th>CO₂ Equivalent (kg)</th></tr></thead><tbody><tr><td>CA-QC</td><td>0.00020567404367292623</td></tr><tr><td>RoW</td><td>0.0012552788469849527</td></tr></tbody></table>	Region	CO ₂ Equivalent (kg)	CA-QC	0.00020567404367292623	RoW	0.0012552788469849527
Region	CO ₂ Equivalent (kg)						
CA-QC	0.00020567404367292623						
RoW	0.0012552788469849527						
Was that task easy or difficult?	Easy						
What part (if any) confused you?	None						
On a scale from 1–5, how satisfied were you	5						

with this part of the experience?	
Did you notice a clear difference between the two responses?	Yes, the researcher role definitely had more information than the policymaker role.
Did you feel that you understood a certain response better than the other?	The researcher role response was better because of more information with the response.
Final Questions	
What did you like most about using EcoValid?	Easy and Simple UI, navigation was the best part
What would you improve or change?	The UI is too simple, there could be better enhancements Suggestions - Better colors and more buttons
Would you consider using a tool like this in your daily life or work?	Yes, if my job or daily life needed environmental information.

Time Taken	Participant #
14 minutes 32 seconds	Participant 1
15 minutes 5 seconds	Participant 2
18 minutes 43 seconds	Participant 3
20 minutes 41 seconds	Participant 4
15 minutes 42 seconds	Participant 5

Findings

- There are issues with accuracy returning different answers based on how the query is structured.
 - When Participant #1 had written a query for the same information 2 different ways, the chatbot responded with 2 different responses.

- There is desire for better follow-up after queries and personalization with responses.
 - Participant #1 expressed that they wanted the chatbot to offer more help after response.
- The difference between answers in different roles was noticeable.
 - Researcher more structured and more factual, policymaker focused more on the data contribution based on input from Participant #3.
 - There was a request from Participant #2 for having the policymaker's answer contain more information.
- There was a lot of positive feedback about the homepage from all participants and the visualization of the data from Participant #4.
- Participant #4 had requests to work on highlighting the key points from the answer for easier and faster understanding.
- Desire for speedier responses from the Participant #4.
- Participant #5 says that the chatbot requires very specific wording which makes it less user friendly. If a question is not asked as per the bot's dictionary of words, the bot sometimes has no response. The bot needs to be able to process natural language.

Revisions for the Future

1. Adding more personalization to the end of responses to user queries, such as “Let me know how I can help more!” or “Does this answer your question?”.
2. Adding a role for generic public users.
3. Structuring the policymaker role responses by having the answer contain more information, so responses are more helpful like the researcher role responses are.
4. Highlighting key information in responses for easier understanding.
5. Speed up time of responses.

Suggestions for Future Teams

Author: Madison Torres

Reviewer/Editor: Esha Mascarenhas

During the development of the EcoValid project, our team encountered several important lessons and identified key opportunities for enhancing both the functionality and security of the platform. The following recommendations reflect our insights and can guide future contributors in maintaining and extending the system:

Development Insights & Suggestions

1. Adding Clarifying Descriptions to Processes in Dataset

Providing clear, consistent descriptions for each environmental process listed in the NameData.txt dataset would improve readability and understanding for both developers and end-users. This is especially useful when building prompt templates, interpreting chatbot results, or onboarding new contributors. Enhanced descriptions reduce ambiguity, minimize logic errors in data processing, and improve the overall maintainability of the system.

Implementation Suggestion: Include a descriptive JSON or Markdown reference file that explains the process type, units, region, and impact categories associated with each entry or add to the current files with this information.

Redesign Suggestions

1. Using a Newer Embedding Model

Upgrading to a newer embedding model would improve the system's ability to interpret user queries more accurately and consistently. In the current setup, some queries return high-quality responses while others yield vague or irrelevant results, especially those involving less common or poorly described processes. This inconsistency stems from the embedding model's limited ability to semantically match queries to dataset entries with non-standardized or ambiguous names. Newer models are trained on broader and more diverse datasets, offering better contextual understanding and domain transfer. This change would improve accuracy in identifying semantically similar processes, reducing mismatches and enhancing the relevance of both factual answers and RAG-based completions.

Implementation Suggestion: Evaluate recent open-source models such as E5, Instructor, or bge series on HuggingFace, and fine-tune if needed for domain-specific embedding.

2. Implementing a Custom Vector Search Algorithm

Replacing the default vector search backend with a custom algorithm tailored to EcoValid's use cases could improve search accuracy, control over relevance ranking, and response latency. This is particularly beneficial when queries require nuanced context matching or domain-specific prioritization, such as weighting location or process types. A custom implementation would also provide greater transparency and extensibility for future optimizations. This helps ensure that similar queries yield consistent results, even if the process names in the data vary in phrasing. Such an approach can correct for limitations in both embeddings and raw data by integrating structured logic into the ranking process.

Implementation Suggestion: Build a hybrid search that combines semantic similarity, like cosine distance, with rule-based filters or weighted scoring based on metadata fields like process region or impact type.

Feature Suggestions

1. Login Functionality

Introducing a secure user authentication functionality would be a critical step towards personalizing the platform. This would work to enhance user experience, privacy, and system integrity. The platform would be allowed to provide different levels of access based on user roles. Researcher and Policymaker. There could also be new roles introduced such as Admin, Analyst, or Guest that would have other functions up to the design team. Role-based access control could be introduced to manage administrative versus general user privileges. This would enable personalized features like saving user preferences, retaining query history, and offering private dashboards. It would also enhance the security by protecting the system against bots, unauthorized data scraping, and brute force attacks.

Implementation Suggestion: Use Flask-Login or Django's built in authentication system depending on preference of framework.

2. Rate Limiting Enforcement

Enforcing rate limits on API queries would ensure stable performance, especially in high traffic conditions. By limiting the number of requests a single user can make within a given timeframe, the platform can maintain responsiveness and avoid system overload. This approach is especially useful in deterring spam, mitigating Distributed Denial of Service (DDoS) style attacks and flood attacks, and promoting fair usage of shared resources. Applying differentiated rate limits based on endpoint type can optimize resource management with efficient tracking of request counts and reset windows.

Implementation Suggestion: Use Flask-Limiter for simple integration with API routes.

3. Session-Based Logging

Integrating a session-based logging system would offer valuable insights into system usage and user behavior, while also strengthening platform security. By logging user actions such as query requests, login attempts, and session durations, the system can have real-time monitoring and facilitate the identification of suspicious patterns or potential breaches. The protection against these breaches can help with brute force login attempts or gaining unauthorized access to another user's account. To comply with data privacy standards, all logs should be anonymized and avoid capturing sensitive information.

Implementation Suggestion: Use a secure logging framework like Monolog or Python's logging module with encryption extensions.

Team Information

Front-End Development

Annabella Jackmore, arjackmo@ncsu.edu, 704-214-8550, UI/UX Developer

Madison Torres, mrtorre3@ncsu.edu, 919-671-6328, Team Leader / UI/UX Developer

Back-End Development

Sid Kaju, skaju@ncsu.edu, 984-298-4718, Server Developer / Debugger

Cole Sanders, cgsande2@ncsu.edu, 984-269-2606, RAG Specialist / API Developer

Full-Stack Development

Esha Mascarenhas, emmascar@ncsu.edu, 864-478-5077, Full Stack Developer

GUI Appendix:

Feedback and evolution:

We received feedback from our sponsors that they no longer wanted login functionality as was specified in the initial project description. Thus the create account, account creation, and login page requirements were removed. Instead, when users launch our application they will be greeted by the chatbot home page. In order to maintain the functionality of users selecting their role, the dropdown menu to select a role was moved to a left sidebar of the chatbot home page. Users are not able to interact with the chatbot until they select their role from the dropdown menu. In addition to removing login functionality, we also added UI enhancements including the green background to match the environmental theme and a box in the top right corner that displays the current role of the user. When a user changes their role all chat history is deleted.

Initial prototype:

For our initial design we created a welcome page, create account page, login page, and landing page. The welcome page allows users to log in to their existing account or create a new account. If the user chooses to create an account, they have to input a username, email, and password (like any regular chatbot login). However, our website includes a “Role” section, which determines permissions for how they can interact with the database.

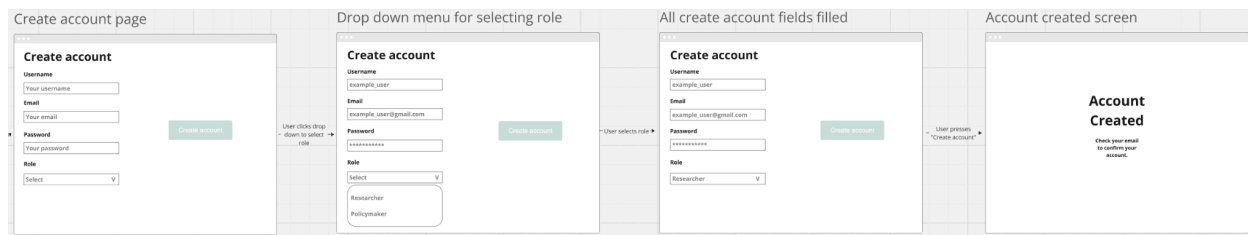


Figure 5: Prototype Create Account Wireframe

For the login page, the user must input their username and password. There's a “Forgot password” option at the bottom.

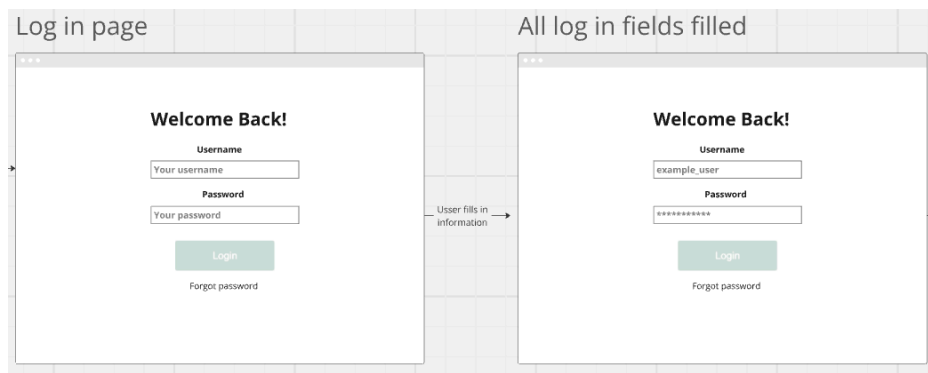


Figure 6: Prototype Login Wireframe

The landing page contains a welcoming salutation (Hello, <First Name>) and a text bar at the bottom to “Chat with Eco”. Then, the user can input a question into the text bar, and receive a response from Eco.



Figure 7: Prototype Chatbot Wireframe

Introduction Page Prototype:

An introduction page was requested by sponsors to include what the problem is currently that we are trying to solve, what EcoValid can do for them, tell users how to use EcoValid, tell users why to use EcoValid, and provide some example questions to get users started. There were two pages brought up in development to show sponsors functionality of each introduction and let them decide.

One introduction page was a pop-up box with minimal information at the top of the screen to reduce clutter. This box allowed users to still be able to chat with the chatbot after selecting a role. The box would disappear after the button “Start Chatting with Eco” was pressed by the user and the UI would display as normal. This prototype was not chosen and will be displayed in the Appendix section as Figure 8.

The second introduction page was a full scale introduction page that served as a landing page for users instead of a pop-up box. This page contained all of the necessary information the sponsors requested and gave more depth to users about EcoValid. The page would continue to the chatbot

after the button “Start Chatting with Eco” was pressed by the user and the UI would display as normal. This prototype was chosen by sponsors due to the information presented and the ease of use for users to understand the chatbot and to continue to using it.

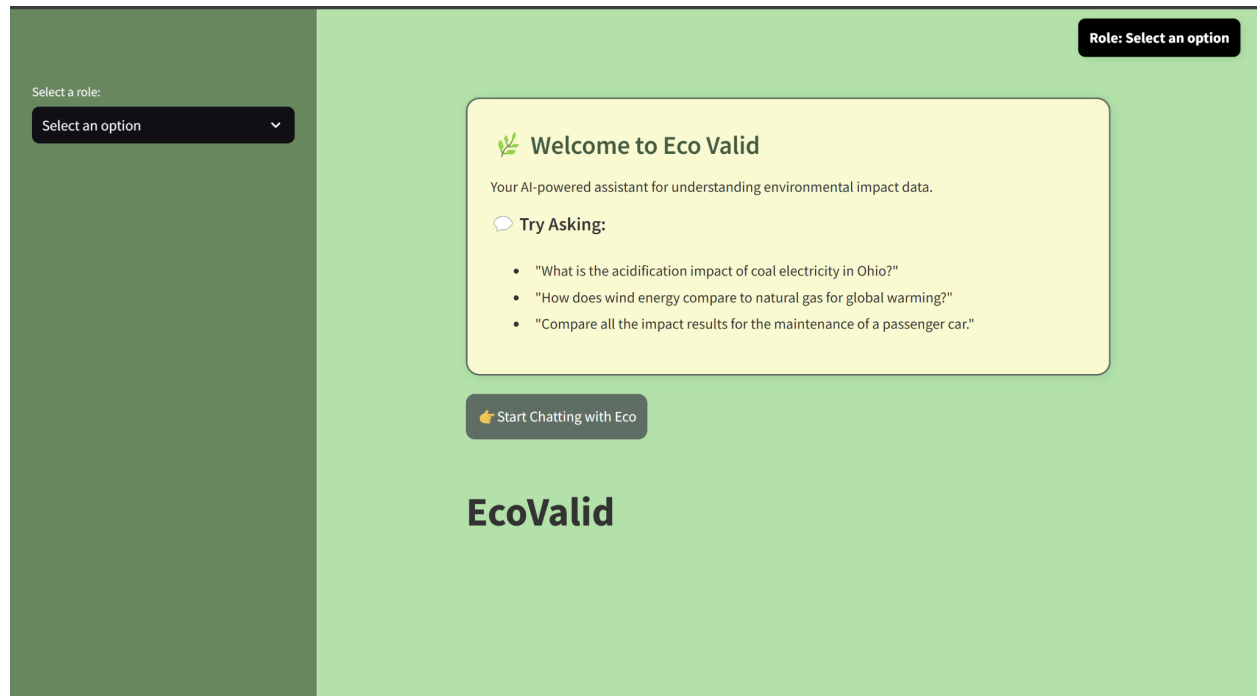


Figure 8: Introduction Page Pop-Up Box Prototype

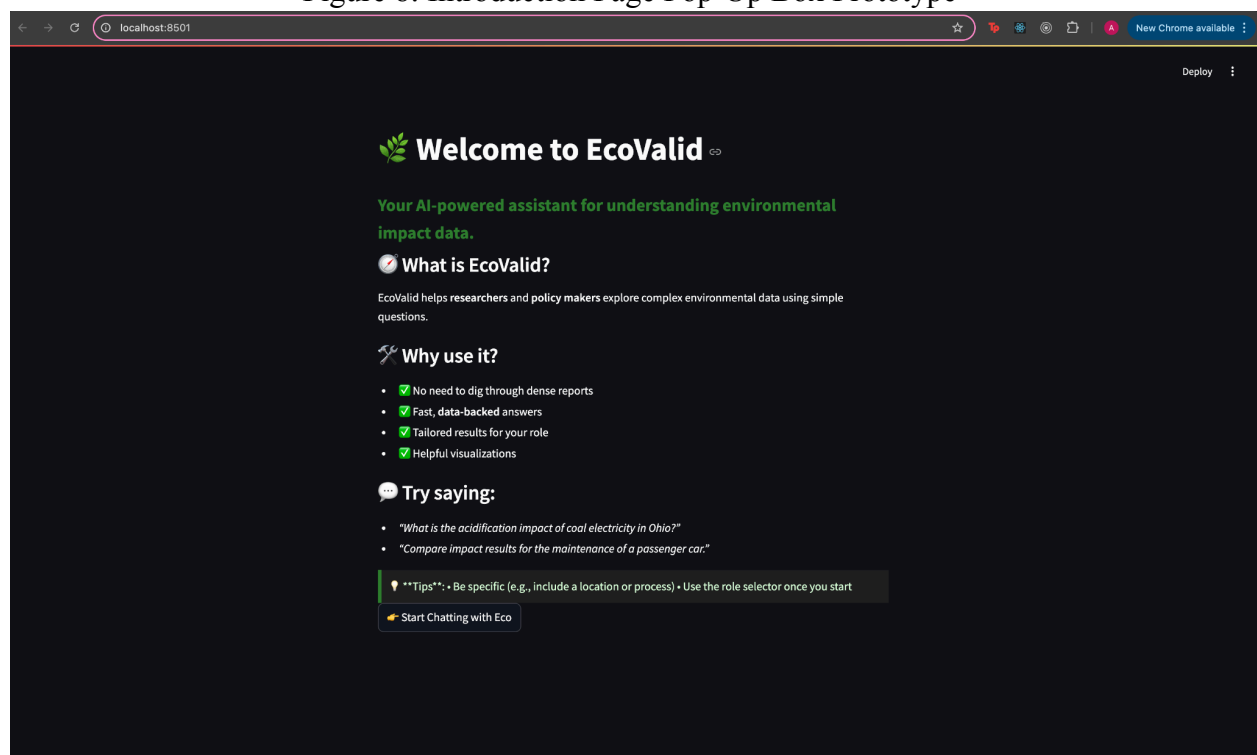


Figure 9: Introduction Page Full-Scale Page Prototype

