BCG-X Challenge 2024

GAIA
AI LANGUAGE MODEL
SUSTAINABILITY

Using Generative AI
to provide a
sustainable future

Authors

Marcelo
Data Scientist

Vitor
Data Engineer

Matheus
Software Engineer

Melissa
Data Scientist

Motivation

The Problem

Brazil ranks fourth among the top 10 countries with the highest cumulative emissions of CO_2 from fossil fuel combustion, deforestation, and land-use change.

https://adit.com.br/mudancas-climaticas-e-desenvolvimento-urbano-impactos-e-desafios/

Go to the Market

“In society, there is a perception that climate change is something that exists. People have heard about it and consider it a fact. In politics, it is different; it enters a complex universe involving various levels and issues, with climate change being one of them.”

Ph.D. Jean P. Ometto - Senior Researcher at the Brazilian National Institute for Space Research and Dean of the Earth System Science Centre.

"Promoting effective communication between public managers, technical reports, and the community can facilitate the incorporation of different perspectives into the decision-making process."

Ph.D. Evaldo Luiz Gaeta Espindola - Senior researcher in ecotoxicology and aquatic ecosystem ecology, conducting research, mentoring, and water resource management, as well as contributing to environmental policies.

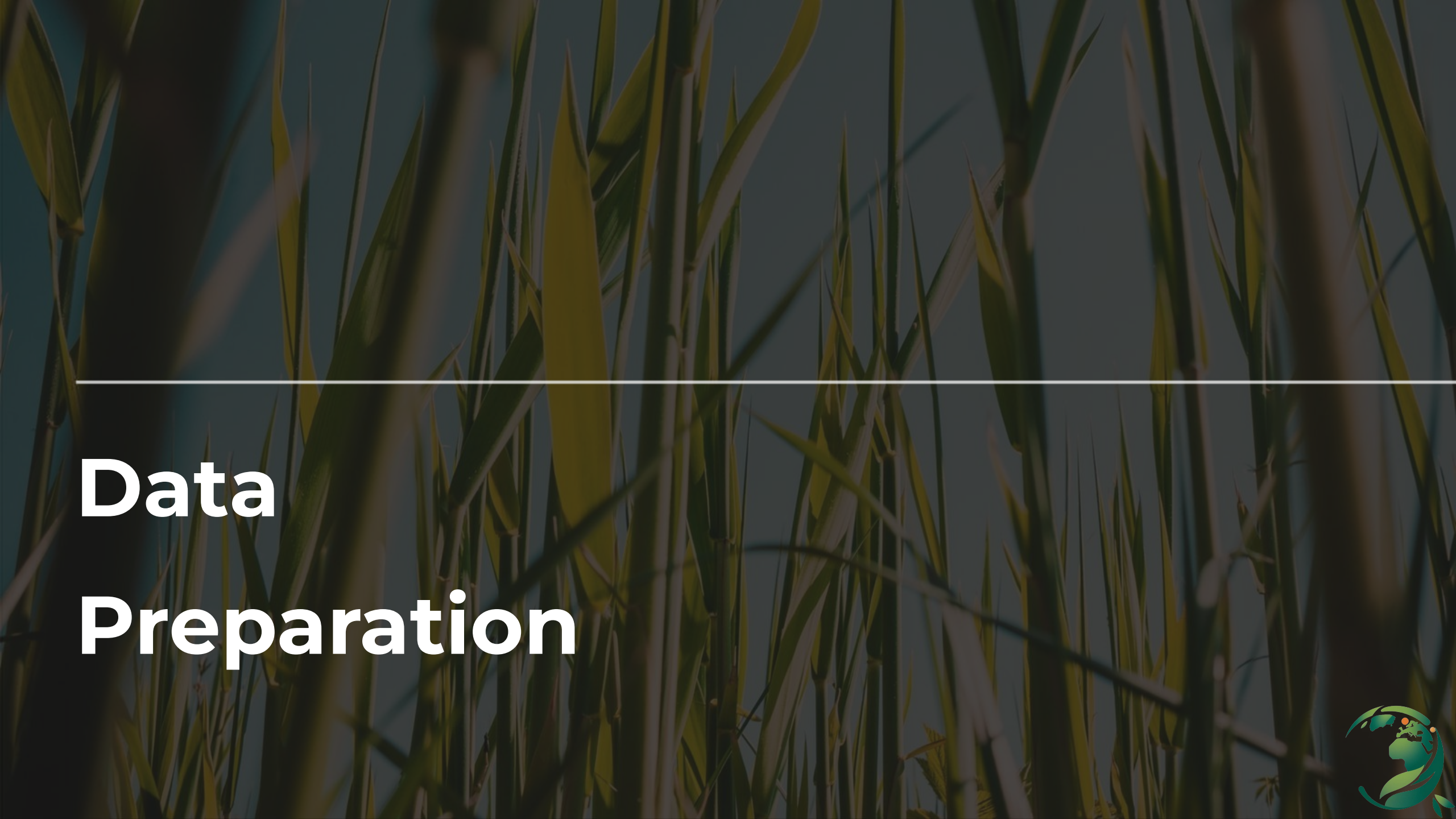
Solution

Our Proposition

GAIA
AI LANGUAGE MODEL
SUSTAINABILITY

Gestão Ambiental com Inteligência Artificial

“Intelligence for a Sustainable Future”

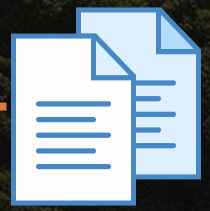


Data Preparation

Rag Flow

Extract images from the PDF files

The information inside the PDF's images is extracted using **PyTesseract** to add all useful information that is possible from the files.



Extract text from the PDF files

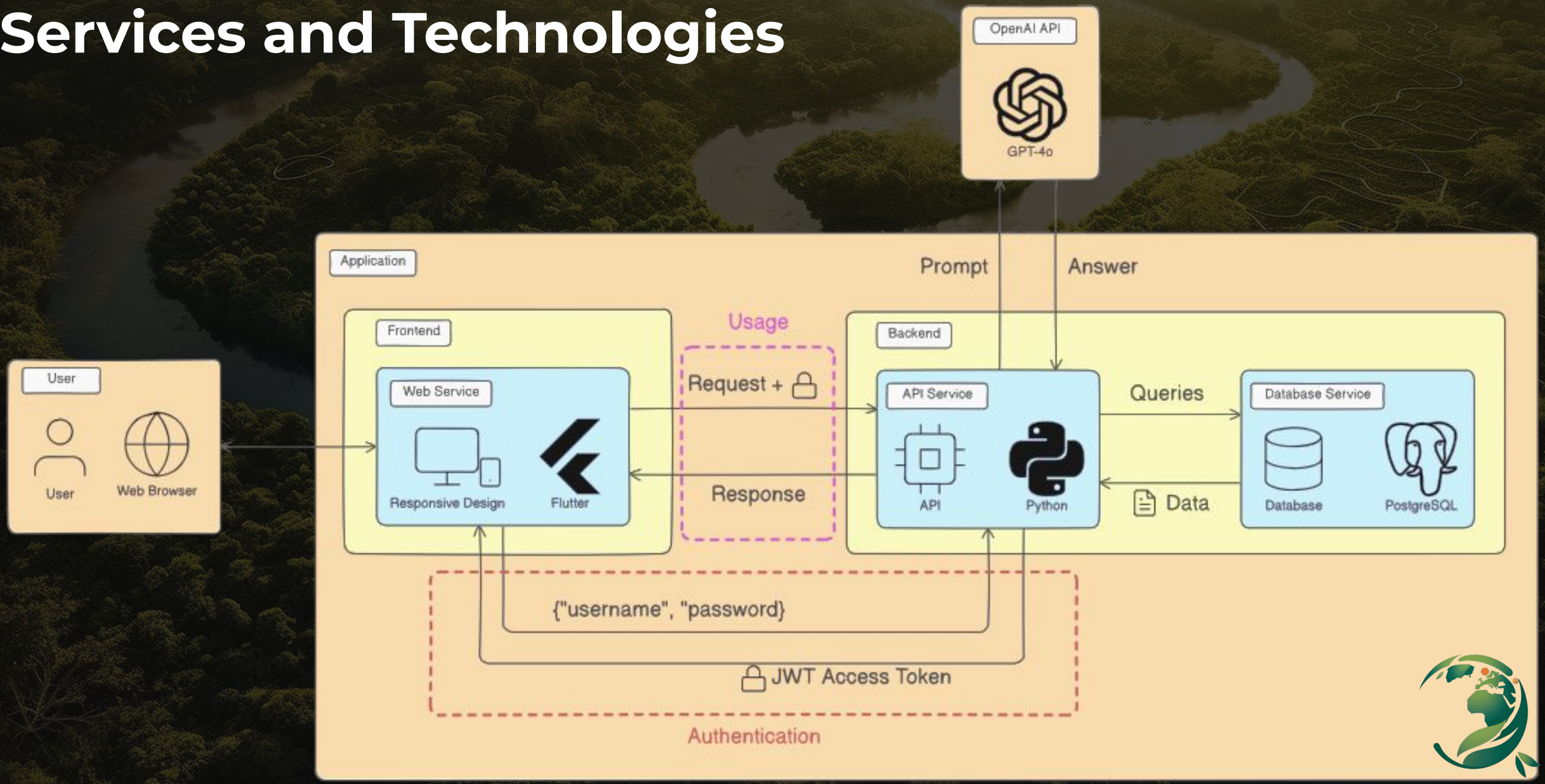
The text inside the PDF files is extracted using **PyMuPDF**.

Clean and embed texts

All the texts extracted are cleaned with regex, splitted in chunks and then, stored and indexed in a vector database using **FAISS** to guarantee the best performance.

Architecture

Services and Technologies



Key Concepts

JWT Authentication

Stateless, reduces the overhead added by sessions and removes the need to provide credentials on every request

Flutter

Responsive design, manages the state of its components.

FastAPI

Handles dependency injection of the database session and user authentication.

PostgreSQL

Stores user and chat data and implements ownership mechanisms through relationships.

Chatbot

Functionalities

Code Structure	LangChain Functionalities	• Concise Code Structure • Useful Methods • Easy Implementation and Orchestration
Question Classification	Filter as a Question Classification using LLM	• More Human like Interaction • Simple Questions are faster processed • Exclude the necessity of a RAG mechanisms for simple interactions
Memory Mechanism	Memory for handling Chat History	• Chat History used to improve answer quality • Combination of Strategies for better results • Maximum of tolkien's not reached
Guard Rails	Steps that prevent Hallucinations	• Default Message for minimum similarity score • Temperature, Frequency and Presency Penalty parameters

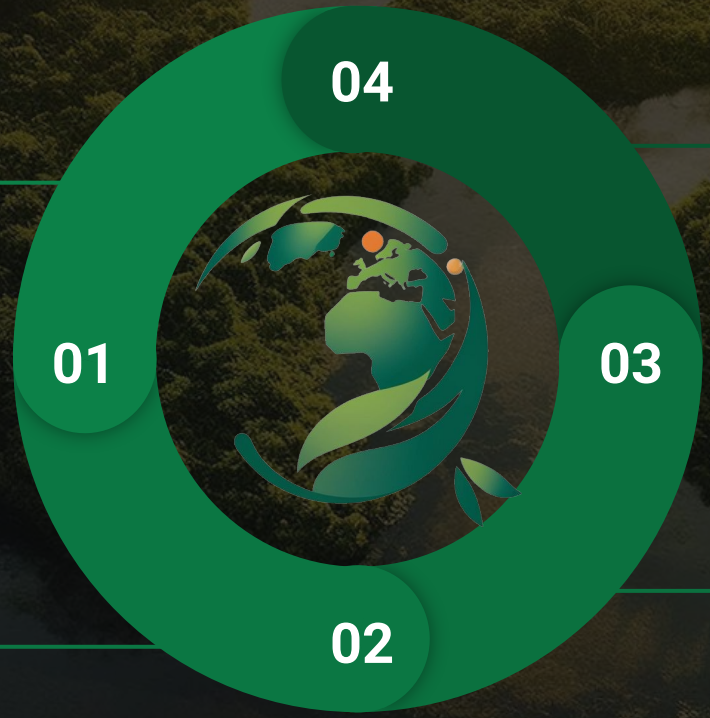
ChatBot Chain Workflow

User Input

Question Classification: General or Specific

Documents Search

RAG Vector Database Search Chain and Context retrieve



User Query Answer

User get his input answered

Prompt with Context

Query to LLM enriched with best context from RAG and Chat History

General Question → Simple LLM Request → Directly Answered

Results

Ragas and Test Dataset

Dataset for Testing

Default Dataset with 100 question generated by IA and answers validated with ambiental specialists from USP - São Carlos

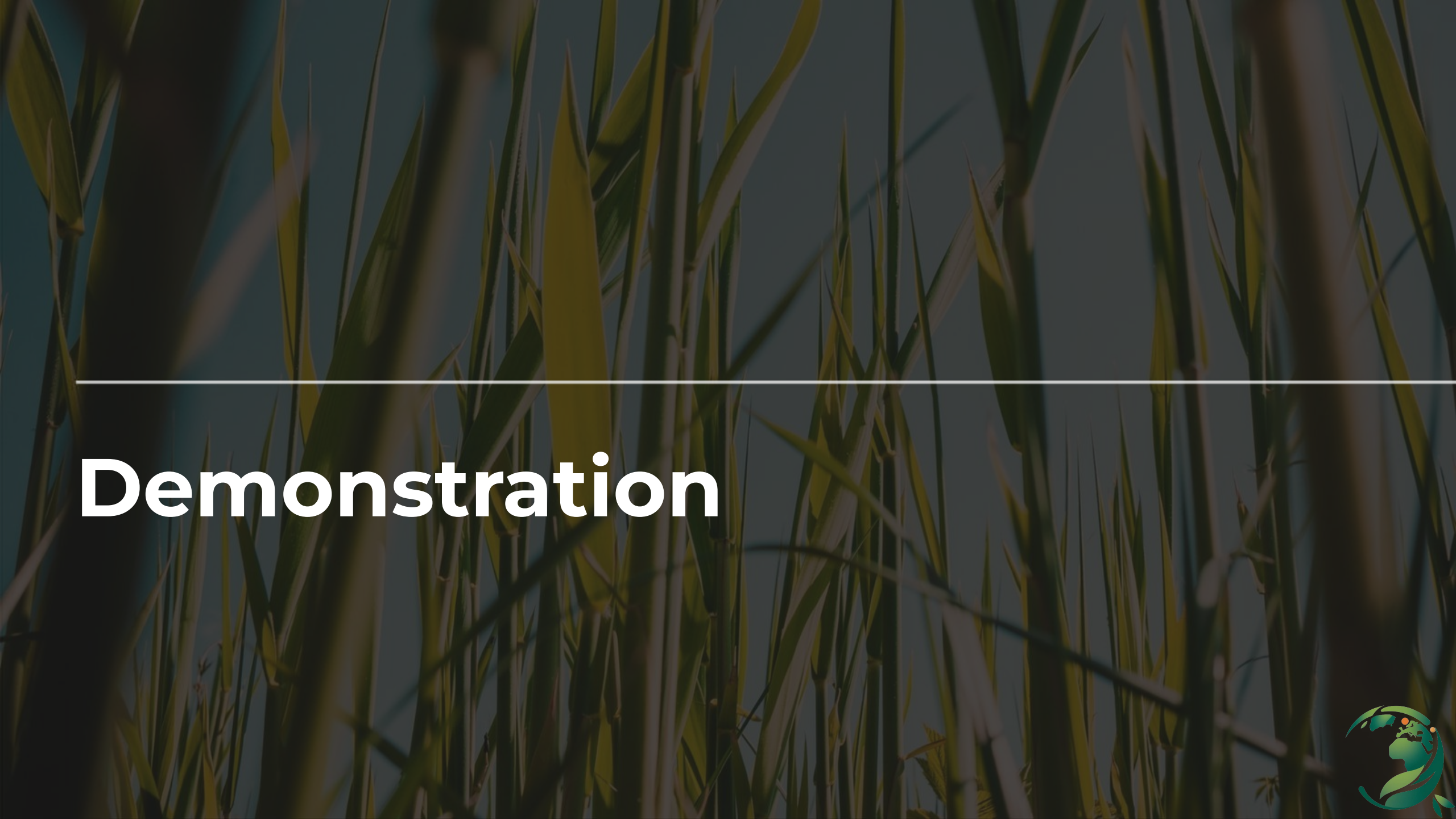
Ragas Pipeline

Use of RAGAS - A library created to help validate query results based on a dataset with the questions and the answer considered the Ground Truth.

RAGAS also has a gamma of metrics for us to use as comparison for parameters variation tests

RAGAS Metrics Range: [0,1]

Faithfulness	Measures how **faithful** the **response** is to the retrieved documents' content. Higher scores indicate that the response is more grounded in the correct information.	0,74
Answer Relevancy	Assesses how **well** the response **aligns** with the user's **question**. High scores mean the answer is directly relevant to the query.	0,87
Answer Similarity	Compares the chatbot's **response** to an ideal reference **answer**, assessing the **similarity** in both content and structure.	0,91



Demonstration

Next Steps

Future Improvements

Host the website on a **cloud provider** and develop a **mobile app**. Add **HTTPS** and token encryption.

Add **tags** to the embedding metadata so that the search may be more **accurate** and **fast**.

Introduce possibility for the client to load **other types** of **content** on the query, such as PDF files, audio, images, etc.

Create a **list of topics on the chat** so that the client may select and personalize their search on specific subjects.

Thank You

GAIA
AI LANGUAGE MODEL
SUSTAINABILITY