**IDEA EVALUATOR**

IDEA EVALUATOR

Team Name: Future Flare

Team Members:

1. Ajinkya Chafekar
2. Kavya Nagesh
3. Pratyush Singh
4. Dolsy Arora

**Table of Contents.**

1. [Problem Statement and Dataset.](#_Problem_Statement_and)
2. [Generative AI models taken into consideration.](#_Generative_AI_models)
3. [NLP Point of view.](#_NLP_Point_of)
4. [Implementation and Scoring and Metrics evaluation strategies –](#_Implementation_Scoring_and)
5. [Visualization.](#_Visualization.)
6. [Conclusion.](#_Conclusion)

## **Problem Statement and Dataset**

In the swiftly changing landscape of today's world, climate change poses a formidable challenge that profoundly affects both our daily lives and the well-being of our planet. The circular economy, which prioritizes the reuse and recycling of resources to minimize waste, emerges as a crucial strategy in addressing this challenge.

Amidst the urgency of climate change, the need to identify and implement impactful circular economy solutions has never been more pressing. The current challenge goes beyond merely devising solutions; it involves the daunting task of effectively evaluating a vast and diverse array of possibilities, discerning the most impactful ones. This process can be overwhelming, given the complexity and sheer volume of potential solutions, leading to cognitive overload for human evaluators.

After thoroughly going through the dataset with ideas that were submitted during the Innovation contest in 2023, we started with the process of brainstorming our task approaches.

**Approaches We are using:**

Idea validator**:** This guide human evaluators by establishing transparent justifications and assessments for crucial metrics, including maturity stage, market potential, feasibility, scalability, technological innovation, or adherence to circular economy principles. Ideas that align with predefined criteria will be emphasized to human evaluators.

Moonshot finder: The moonshot finder pinpoints extremely ambitious ideas with significant potential returns and heightened risk of failure. It emphasizes unconventional concepts that might be overlooked by domain experts. Despite attempting its use, we found it challenging due to the elevated risk of failure.

Idea Filter: This is designed to exclude poorly formulated, off-topic, or vague ideas, streamlining human evaluators' efforts. It directs their time and resources towards meticulously crafted, well-articulated concepts with tangible relevance.

**Deliverable:** We must provide a comprehensive description of our tool that highlights the technical and creative aspects and a demonstration of the tool in action.

**Dataset:** The Dataset contains three columns. The first one is “id,” second is the problem – this column contains the issue that the idea is focusing on and what the idea will be solving, and the solution contains a detailed approach of how the problem can be solved and why it is vital to solve it.

In this data-driven exploration, we delve into a fascinating dataset that encapsulates the how different humans are when compared to their perceptions and thought processes. Each column in this dataset represents a narrative, a story, or a detailed account of each environmental issue that we are now facing. As we embark on this journey through textual data, we aim to extract meaningful insights, uncover patterns, and gain a deeper understanding of the human brain.

# **Generative AI models taken into consideration.**

**Definition**

We have implemented the Question and Answering modules of all the models we have considered in the following to approach our Idea Evaluation Use cases.

BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly optimized BERT approach), and DistilBERT (Distilled BERT) are all transformer-based models designed for natural language processing (NLP) tasks. Hence, we considered implementing them. The pre-trained models made it very efficient for us to train the textual data we had. We implemented the Question and Answering models of the Transformers package as they have the similar question and context structure of functioning.

**Similarity** – Due to their large corpus that were pre-trained, we had to leverage this to our benefit of deriving the key points from the proposed solution.

**Differences** – Bert and DistilBert did not require any hyperparameters tuning, however, Roberta offers to optimize some hyperparameters that could potentially improve the performance.

GPT-2 and GPT-3:

GPT-2 (Generative Pre-trained Transformer 2) and GPT-3 (Generative Pre-trained Transformer 3) are both powerful language models developed by OpenAI. While they share the same underlying architecture, GPT-3 is a more advanced and larger model compared to GPT-2.

**Observations** – We found that the performance of GPT-2 was more relevant and efficient than GPT-3 in terms of question and answering, despite GPT-3 having better parameters and having trained on an extensive dataset.

**XLNet:**

XLNet (eXtreme Learning Machine Network) is another powerful transformer-based model for natural language processing (NLP), and it has been successfully applied to various tasks, including question answering.

Observations: The Performance was good when compared to all the mentioned generative AI models here. We could understand since it takes the context of a text in both the directions, the masking is effective, and it takes into count of permutations of the given textual tokens which explains its performance. But the gist from the textual data we received was incomplete and could not be potentially used in our research.

**T5**:

T5 (Text-to-Text Transfer Transformer) is a transformer-based model that approaches natural language processing tasks in a unified manner by converting all tasks into a text-to-text format.

Observations: We noticed that T5 performed poorly when it comes to contextually learn the textual data whereas outperformed when we had to extract a summary of an incredibly detailed paragraph that was not cleaned or pre-processed.

**LDA**:

Latent Dirichlet Allocation (LDA) is a topic modelling technique that can be applied in the context of question answering to discover underlying topics within a collection of documents. Taking into advantage of its probabilistic model that assumes each document is a mix of topics, and each topic is a mix of words.

LDA provides a way to uncover hidden topics within a collection of documents, making it a useful tool for organizing and understanding copious amounts of text data. When applied to question answering, it can aid in identifying relevant information and improving the efficiency of the answer extraction process.

Observations: We attempted to implement LDA with the limited knowledge we had to calculate the probabilities and weightage of the topics of a given paragraph at any point of time. However, we could not venture further into how this method of topic modelling can be helpful to us in weighing each solution against variety of parameters.

**Seq2Seq**:

Sequence-to-Sequence (Seq2Seq) models are commonly used in question answering systems, particularly in scenarios where the length of the input and output sequences can vary. It considers of LSTM and adds the Dense layers according to the requirements.

This is a good option to be considered for working with long sequences of Textual data.

# **NLP Point of view.**

Data Collection: Gather relevant text data for analysis.

Text Cleaning: Preprocess the data by removing noise, irrelevant information, or formatting issues.

Tokenization: Break down the text into smaller units (tokens) to facilitate analysis.

Stopword Removal: Exclude familiar words (stopwords) that don't carry significant meaning.

Stemming and Lemmatization: Reduce words to their base or root form for consistency.

Feature Extraction: Transform text into numerical representations, such as word embeddings or vectors.

Model Building: Choose and train an appropriate NLP model, like a neural network or machine learning algorithm.

Evaluation: Assess the model's performance using metrics like accuracy, precision, and recall.

Optimization: Fine-tune the model and parameters for better results.

Deployment: Implement the model in a real-world environment for application.

## 

### **Implementation and Scoring and Metrics evaluation strategies – Why such metrics?**

**What is prompt engineering?**

**Prompt engineering refers to the deliberate and strategic construction of prompts or input queries when working with language models or natural language processing (NLP) systems. It involves carefully crafting the instructions or queries given to the model to influence its output and guide it towards desired responses.**

**Language models, such as GPT-3, are trained to generate contextually relevant and coherent text based on the input they receive. Prompt engineering recognizes that the choice of words, structure, and context in the prompt can significantly impact the model's output. By refining and optimizing the prompts, users can shape the behaviours of the language model to better suit their specific needs.**

**We have used open AI API as OpenAI provides an API (Application Programming Interface) that allows developers to integrate and interact with their language models, such as GPT-3. Developers can use the OpenAI API to access the capabilities of these models in their applications.**

**OpenAI models have demonstrated strong natural language understanding, making them suitable for tasks like text completion, language translation, summarization, question-answering, and OpenAI's models, such as GPT-3, are among the most powerful language models available. They can generate contextually relevant and coherent text across a wide range of tasks.**

**spaCy is an open-source natural language processing (NLP) library for Python designed to handle various NLP tasks efficiently. Developed by Explosion AI, spaCy is known for its speed, accuracy, and ease of use. It provides pre-trained models for various languages and allows users to perform tasks such as tokenization, part-of-speech tagging, named entity recognition**

**en\_core\_web\_sm refers to a pre-trained English language model provided by spaCy, a popular natural language processing (NLP) library for Python. The model is part of the spaCy's English language models and is designed for general-purpose NLP tasks.**

**The similarity method is called on one problem or solution token, and the other keyword token is passed as an argument. The result is a similarity score ranging from 0 to 1, where 1 indicates the highest similarity.**

**It's important to note that to use the similarity function, spaCy requires that the models have word vectors available. Therefore, it is recommended to use medium (en\_core\_web\_md) or large (en\_core\_web\_lg) models that come with pre-trained word vectors.**

**Keep in mind that the similarity score is based on the word vectors learned during the model's training. While it can capture semantic similarity to some extent, it may not be suitable for all types of similarity tasks.**

**Metrics:**

The formula for calculating each metric is considering a variety of factors as explained below. We have considered five major metrics against which we will evaluate the ideas. They tell us how the usefulness and the relevance of an idea can be measured.

The five metrics are:

1. Environmental Impact Score (EIS):

2. Financial Viability Score (FVS):

3. Feasibility and Scalability Score (FSS):

4. Innovation and Novelty Score (INS):

5. Social Impact Score (SIS):

1) EIS:

Some of the key factors considered in calculating environmental impact score are energy consumption, greenhouse gas emissions, water usage, resource depletion, waste generation, end-of-life considerations.

EIS= (0.3×Waste Reduction) +(0.3×Resource Conservation) +(0.4×Pollution Reduction)

This gives more weightage to the Pollution as Global warming is alarmingly affected by Pollution.

2) FIS:

The common factors that are considered while calculating the Financial Viability score are Cost savings, Profitability ratio, Cash flow, Debt management and financial leverages.

3) FSS:

The calculation of feasibility and scalability scores can vary depending on the specific context, industry, and criteria established by an organization or project. These scores are often used to assess the viability and potential for growth of a business, project, or technology. Implementation is important but at the same time we also must take care of how we can reuse or apply the same solution for a different problem.

4) INS:

Measuring innovation and novelty can be challenging, as these concepts are often subjective and context dependent. However, various metrics and approaches combining the quantitative and qualitative measures is often necessary to capture the diverse aspects of innovation and novelty. Uniqueness of a solution and out of the box thinking are major factors when it comes to this metric.

5) SIS:

Finally, innovations always cause ripples of change to the society and economy we live in. Calculating a social impact score involves assessing and quantifying the positive or negative effects that an organization, project can be complex and context-dependent, as social impact is multidimensional and varies across different sectors and activities, However, we will be considering Employment opportunities and how interested is the public in this measure.

# **Visualization.**

# **A graph of a bar**

# **A screenshot of a graph**

# **Conclusion**

We have used open ai Prompt engineering model and Spacy similarity function to measure and evaluate all the ideas from the dataset. The results are similar when it comes to detecting spam ideas and the useful ones which can be worked upon further to train and test.