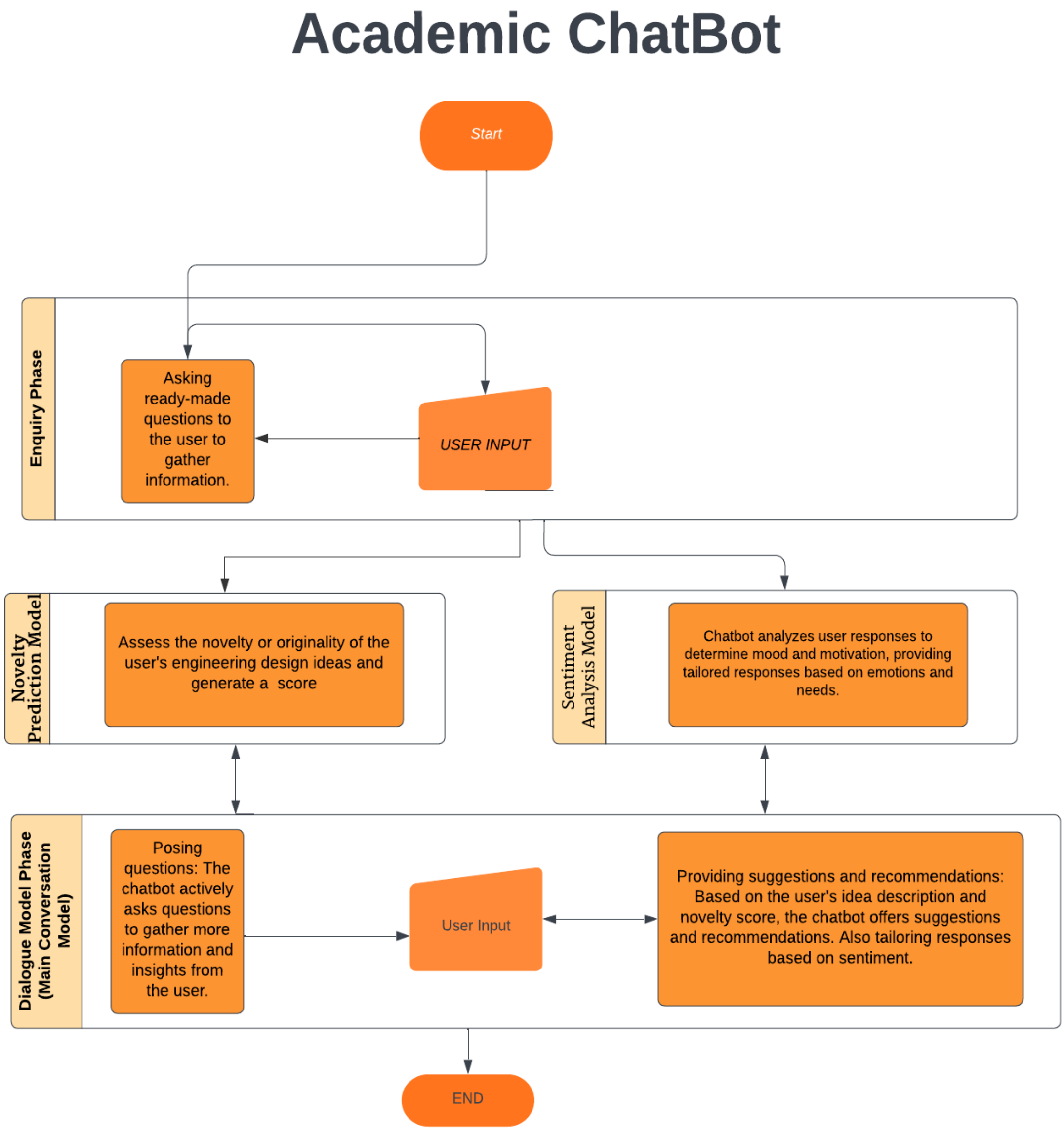
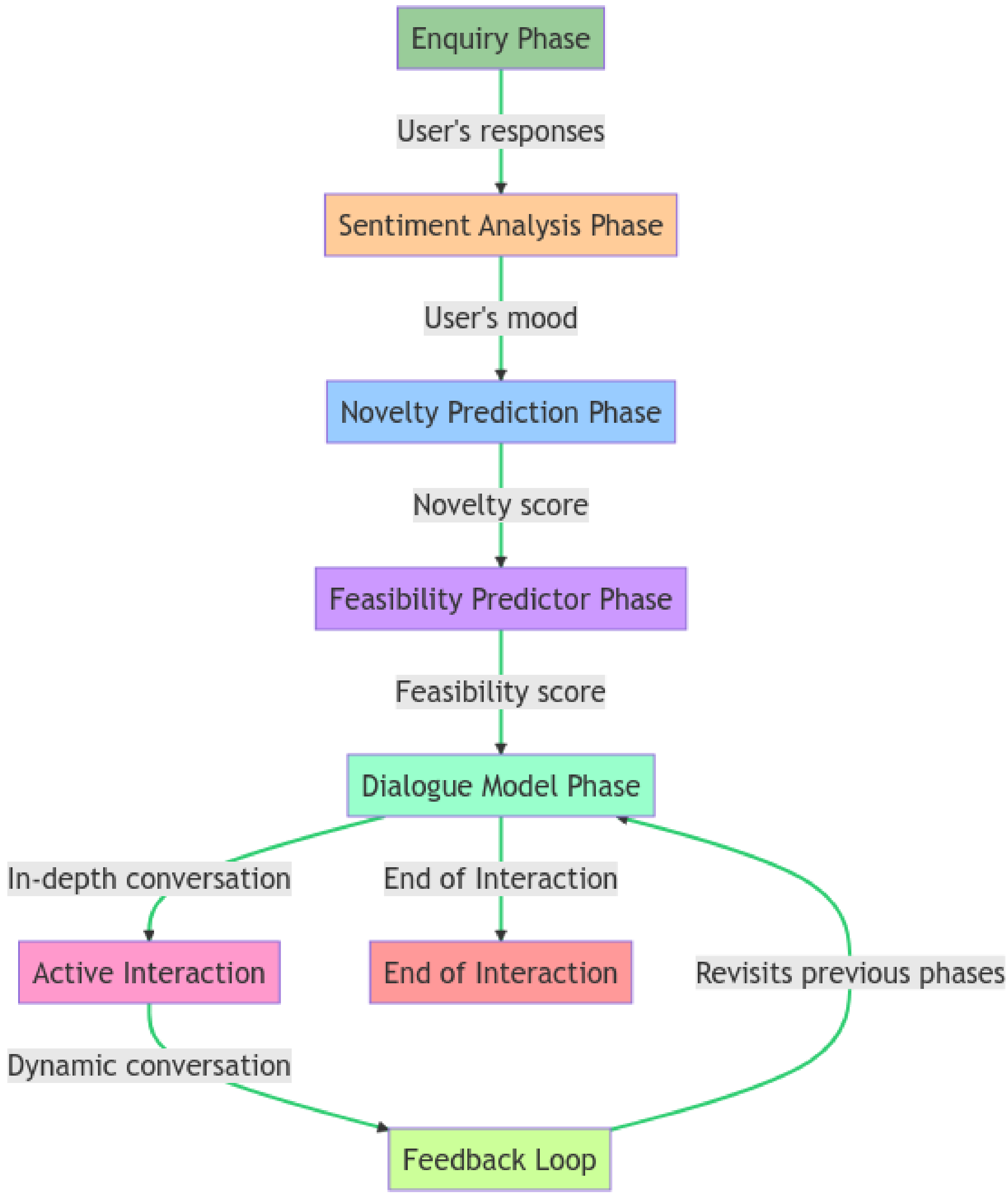
Detailed Research Project Plan

Chatbot Flow Overview




Enquiry Phase
User's responses
Sentiment Analysis Phase
User's mood
Novelty Prediction Phase
Novelty score
Feasibility Predictor Phase
Feasibility score
Dialogue Model Phase
In-depth conversation
End of Interaction
Active Interaction
End of Interaction
Revisits previous phases
Dynamic conversation
Feedback Loop

Enquiry Phase: Our chatbot initiates the conversation with pre-set questions directed at the user to understand their design project, objectives, constraints, and preliminary ideas. The user's responses are securely stored for future reference and application.

Validator Phase: We employ a validation model that is capable of recognizing context of text and will accurately be able to identify if the user's input is in fact describing ideas about a washing machine design and not any random text that is not aligned with the chatbot function to work on washing machine engineering design.

Novelty Prediction Phase: The chatbot integrates a Novelty Prediction Model in this phase to evaluate the originality and innovation of the user's engineering design ideas. This novelty score serves as a crucial parameter in the subsequent conversation with the user.

Feasibility Predictor Phase: In addition to assessing novelty, our chatbot employs a Feasibility Predictor Model to gauge the practicality and viability of the user's design ideas. This additional layer of analysis helps the chatbot to provide more well-rounded advice and feedback.

Dialogue Model Phase (Main Conversation Model): With insights from the previous phases, the chatbot then embarks on an in-depth conversation with the user. It poses relevant questions, offers valuable suggestions and recommendations, and provides motivation based on the user's sentiment, novelty score, and feasibility score. Its objective is to simulate the enriching discussion a student might have with a mentor or instructor. If the novelty score stagnates, a supporting video could be provided to help stimulate creativity.

Active Interaction: Our chatbot transcends traditional chatbot functionality by not only responding to user prompts but also taking the lead to generate prompts and questions. This interactive approach fosters an engaging and dynamic conversation.

Feedback Loop: The chatbot revisits the Sentiment Analysis, Novelty Prediction, and Feasibility Predictor Phases as required, in response to the user's replies. This feedback loop mechanism facilitates the chatbot's capacity to dynamically adapt to the user's needs and evolving progress.

End of Interaction: The interaction culminates when the user's goals are achieved, they no longer require assistance, or they choose to conclude the conversation.

Timeline and High-Level Task Breakdown

Phase 1: Enquiry [2 weeks] (Start Date: 5/24/23) Completed

1. Naveen and : Refer to literature and deploy questioning strategies to generate the enquiry questionnaire that will define the user requirements for the design (1 week)
2. Naveen and : Creating the code to integrate/hard-code this questionnaire into the chatbot and validate the responses (1 week)

Phase 2: Novelty & Feasibility Model [2 weeks] (Start Date: 6/7/23) Completed

1. Naveen: Review literature and decide on the approach for novelty and feasibility prediction. (0.5 week)
2. Naveen: Implement a model for novelty prediction and train it. (1 week)
3. Naveen: Implement a model for feasibility prediction and train it. (1 week)
4. : Perform initial testing of the model and fine-tune it. (0.5 week)

Phase 3: Validator Model [2 weeks] (Start Date: 7/21/23) Completed

1. Naveen: Review models and select one for your application. (0.5 week)
2. Naveen: Implement the chosen Validation model and train it. (1 week)
3. Naveen: Test and fine-tune the model. (0.5 week)

Phase 4: Main Conversation Model [3 weeks] (Start Date: 7/5/23) Completed

1. Naveen and (together): Research and choose a suitable language model for conversation. (0.5 week)
2. : Create a dataset tailored to the conversation style of an AI tutor/instructor. (0.5 week)
3. Naveen: Fine-tune the pre-trained model on your dataset. (1 week)
4. Naveen and (together): Integrate the conversation model with the sentiment analysis model and the novelty prediction model. (1 week)

Phase 5: Integration and Testing [2 weeks] (Start Date: 7/26/23) In progress

1. Naveen and (together): Test the entire system to ensure all components are working together properly. (1 week)
2. Naveen and (together): Debug any issues that come up during testing. (1 week)

Phase 6: Final Touches [2 weeks] (Start Date: 8/9/23) Not started

1. Naveen: Improve the models based on testing feedback.
2. : Make final touches to the system (UI improvements, documentation, etc.).

End of Project: 8/20/23 In progress

Please Note : This timeline can be unpredictable and these are just tentative estimates. We will have to regularly revisit and update our plan as we progress to help keep the project on track.

Project Work

Naveen

Week	Tasks	Hours
Week 1	• React and Flask Application • Research on Pre-trained Models • Chatbot with various Pre-trained models from hugging face	• 15 • 3 • 15
Week 2	• Research on fine-tuning pre-trained models • Performed small scale Fine-tuning on chatbot • Project planning and documentation	• 4 • 10 • 6
Week 3	• Research on Questioning strategies • Updated Project plan and documentation • Generated enquiry phase bot prompts	• 3 • 4 • 3
Week 4 (June 5 - 9)	• Worked on the frontend React code for implementing the Enquiry phase • Logic of enquiry phase response prompt and sequential flow based on user response • Fixed Bugs and added JSON storage of result	• 8 • 6 • 2
Week 5 (June 12-16)	• Literature review on multi-label NLP classification • Research on NLP classification using HuggingFace Transformers • Research on Dataset cleaning required for training the transformer model to perform classification • Proposed a methodology to implement the Novelty and feasibility Model • Data Pre-processing (did it manually at first and then tried AutoTokenizer from HuggingFace which made my life easier, had to truncate the design description and use padding if it fell short of 50 characters due to model constraints) • Generating new data using ChatGPT (using few-shot prompting technique to generate more data for training the model)	• 1 • 2 • 3 • 1 • 6 • 2
Week 6 (June 19-23)	• Developed a Novelty and feasibility prediction model • Worked on generating more accurate results • Exploring option to create 2 separate models for unary prediction of each target class	• 6 • 6 • 4

Week	Tasks	Hours
Week 7 (June 26 - 30)	● Developed a new model for Novelty Prediction ● Generated new data using chatgpt ● Pre-trained and saved the fine tuned model ● Tested and created a function to call the model	● 12 ● 1 ● 3 ● 4
Week 8 (July 3 - 7)	● Created more data using ChatGPT ● Trained model on new and larger data ● Developed the feasibility Model ● Tested the models	● 2 ● 7 ● 3 ● 3
Week 9 (July 10-14)	● Balancing the data for both datasets (improved the prompt based on feedback from our meeting) ● Re-trained the feasibility Model. ● Research on Validation Models	● 10 ● 4 ● 4
Week 10 (July 17-21)	● Experimenting on the Models to create the validation model ● Modularising model code for easier integration	● 16 ● 4
Week 11 (July 24-28)	● Development of the validation Model ● Preparing the server to train the NaF models ● Training NaF models for higher number of epochs	● 5 ● 5 ● 0
Week 12 (July 31 Aug - 4)	● Developing a new Validation model ● Fixed the server issues and Training the Novelty Model on the server for 100 epochs	● 10 ● 15
Week 13 (Aug 7-11)	● Completed the Validation Model ● Completed fine tuning Novelty and validation model ● Feasibility Model Fine tuning in progress ● Re-visiting the chatbot code using Flask and React	● 10 ● 6 ● 0 ● 4
Week 14 (Aug 14 - 18)	● Introduced Sagar to the project ● Modularised the 3 models	● 1 ● 6
Week 15 (Aug 21 - Aug 25)	● Completed modularising the backend models and their prediction/classification function ● Built 3 different backend APIs using Flask for Validation, Novelty and Feasibility models	● 5 ● 5

Project Progress

Phase 1: Enquiry Phase

Tasks -

1. Naveen and : Refer to literature and deploy questioning strategies to generate the enquiry questionnaire that will define the user requirements for the design (1 week) Completed
2. Naveen and - Work on comments (fallbacks, examples, intro) Completed
3. Naveen and : Creating the code to integrate/hard-code this questionnaire into the chatbot and validate the responses (1 week) Completed

User Requirements: In this phase, the chatbot aims to gather essential information from the user to understand their design project and objectives. The user requirements include:

- Design project description: The chatbot will prompt the user to provide a detailed description of their engineering design project, including the problem they are trying to solve and any specific goals they have.

- Constraints and limitations: The chatbot will ask the user about any constraints or limitations they need to consider during the design process, such as budget, materials, or time constraints.

- Initial ideas: The chatbot will encourage the user to share their initial ideas or concepts related to the design project. This will help the chatbot understand the user's thought process and provide tailored guidance.

- Design objectives: The chatbot will inquire about the user's desired outcomes or objectives for the design project. This will help the chatbot align its suggestions and recommendations with the user's goals.

Questioning Strategies: The chatbot will employ various strategies to ask questions effectively and elicit valuable information from the user like:

- Open-ended questions: The chatbot will ask questions that require descriptive responses from the user, allowing them to express their ideas in detail.

- Probing questions: The chatbot will ask follow-up questions to delve deeper into specific aspects of the user's design project, seeking clarifications or additional information.

- Prompting for examples: The chatbot will request the user to provide examples or concrete instances related to their design project to better understand their requirements and objectives.

- Seeking preferences: The chatbot will ask questions to determine the user's preferences, such as preferred design aesthetics, target audience, or specific functionalities they desire.

Example of Phase 1 conversation -

Bot: Please provide a detailed description of your engineering design project.
User: I'm working on a solar-powered water heater design for residential use. It should be affordable and energy-efficient.

Bot: Are there any constraints or limitations you need to consider for your design project?
User: Yes, I have a budget constraint of $1000 and the water heater should be suitable for a family of four.

Bot: What are your initial ideas or concepts for the design project?
User: I'm considering a design that uses technology to capture solar energy and transfer it to heat the water.

Bot: What are your desired outcomes or objectives for the design project?
User: My objective is to create a water heater that reduces energy costs for households and promotes sustainability.

Enquiry Phase result

{
    User Requirements: {
    - Design Description: I'm working on a solar-powered water heater design for residential use. It should be affordable and energy-efficient.
    - Constraints: Yes, I have a budget constraint of $1000 and the water heater should be suitable for a family of four.
    - Initial Ideas: I'm considering a design that uses solar energy to transfer it to the water.
    - Objectives: My objective is to create a water heater that reduces energy costs for households and promotes sustainability.

    }
}

This result will then be used by the sentiment analysis model, feasibility score predictor and the novelty score predictor which we can refer to as the **SAM, FSP** and **NSP** henceforth. Once those sentiments, feasibility scores and novelty scores are generated, the chatbot will enter into the main dialogue phase.

Literature Review

1. Questioning Techniques: A Study of Instructional Practice

User Requirement Gathering (Bot texts) -

Name CHATBOT: ANN(Artificial neural networks)

Hello, I'm ANN. I'm here to assist you in designing an exceptional engineering project. Whether you're looking to explore a new idea or seeking guidance to improve an existing project, I'm here to help. Your insights and details are crucial for me to offer the most helpful guidance, so thank you in advance for your time and effort in sharing your project's specifics with me. To get started, I'd like to provide some context on generating a new project idea. By understanding the context of your project, I can better assist you in formulating an innovative and impactful engineering design.

So, let's embark on this journey together and create something extraordinary!

Before we begin, may I ask how you would like to be addressed?

START OF ENQUIRY PHASE

Title of Your Design Project

Prompt: Great! Now, let's move on to the exciting part of naming your engineering design project. Please take a moment to think about a title that accurately represents your project. You can draw inspiration from any past experiences you've had or even explore hypothetical ideas. Feel free to be creative and descriptive with your title. Once you have it, kindly share it with me.

Fallback Example: If you're finding it challenging, think of something that describes your project's main idea or objective. For example, if your project is about creating a more efficient washing machine, you could call it "Smart Water-Efficient Washing Machine".

Motivation and Inspiration: Don't worry, naming your project can be a creative process. Remember, the title should capture the essence of your vision, and there's no rush. Take your time to think about it.

Design Project Description

Prompt: Next, could you describe your project to me in simple terms? I would appreciate it if you could provide a brief description of your project. The more details you can provide about the problem you're trying to solve and your specific goals, the better I can understand your needs. To help you break it down, you can start by telling me what problem you're aiming to address with your project and what you hope to achieve.

Fallback Example: If you're unsure where to start, you could begin by talking about the purpose of your project, its target audience, or any unique attributes it may have. For instance, if your project is about a washing machine, you could discuss its capacity, energy sources, cost-efficiency, and so on.

Motivation and Inspiration: I understand that summarizing your project can be challenging, but don't worry. Even small details can be very helpful in making your project more clear. Take your time, and remember that each bit of information brings us closer to understanding your vision better.

Initial Ideas and Concepts

Prompt:How would you implement your project? Could you tell me about the technologies or methods you're considering using? Sharing your initial ideas or concepts related to the project will help me understand your thought process better.

Fallback Example: If you're finding it difficult to express your ideas at this point, don't worry! You might start by explaining the core technology or method you're considering for your project. For a smart washing machine, you could discuss how you plan to integrate smart technologies, or any innovative features you want to include, such as AI-based fabric care.

Motivation and Inspiration: Don't worry if you're finding it challenging to articulate your ideas. All big projects start with small ideas. Feel free to share any thoughts, however preliminary they may be. Remember, creativity knows no bounds!

Design Objectives

Prompt: To better align my guidance with your needs, could you please provide more details about how you envision implementing your project and the specific technologies you plan to use? Understanding your goals and implementation approach will allow me to provide tailored recommendations and suggestions.

Fallback Example: If you're unsure, try thinking about the impact you want your project to have. If your project is a smart washing machine, do you aim to make it affordable, highly efficient, user-friendly, or something else?

Motivation and Inspiration: I understand that setting objectives can sometimes be challenging, but it's an essential step in shaping your project. These goals will serve as a roadmap for your design process. Remember, no goal is too big or small, and I'm here to help you achieve it!

Thank you again for sharing all this valuable information! Let's continue shaping your idea into a successful project!

END OF ENQUIRY PHASE

Phase 2: Novelty & Feasibility Model Completed

Tasks

1. Review literature and decide on the approach for novelty and feasibility prediction. (0.5 week) Completed
2. Implement a model for novelty prediction and train it. (1 week) Completed
3. Implement a model for feasibility prediction and train it. (1 week) Completed
4. Perform initial testing of the model and fine-tune it. (0.5 week) Completed

Literature Review

- Large-Scale Multi-Label Text Classification on EU Legislation

The research paper titled "Large-Scale Multi-Label Text Classification on EU Legislation" focuses on Large-Scale Multi-Label Text Classification (LMTC) in the legal domain. The authors introduce a new dataset of 57k legislative documents from EUR-LEX, annotated with approximately 4.3k EUROVOC labels suitable for LMTC, few- and zero-shot learning.

The researchers experiment with several neural classifiers and find that Bi-directional Gated Recurrent Units (BIGRUs) with label-wise attention perform better than other current state-of-the-art methods. They also investigate which zones of the documents are more informative and show that considering only the title and recitals of each document leads to almost the same performance as considering the full document. Furthermore, the authors fine-tune BERT,obtaining the best results for all but zero-shot learning labels.

In the context of our research project on classifying engineering design project ideas as novel and feasible, this paper provides valuable insights into the use of large-scale multi-label text classification. The methods and findings of this paper could potentially be adapted to classify engineering design ideas based on various criteria, such as novelty and feasibility. The use of BIGRUs with label-wise attention and the fine-tuning of BERT could be particularly relevant for developing a robust classification model.

- Medical Code Prediction From Discharge Summary

The research paper titled "BERT and Sequence Attention for Automatic ICD Coding of Clinical Notes" discusses the importance of accurately classifying International Classification of Diseases (ICD) codes in clinical notes. Manual classification is time-consuming and costly, prompting the need for automated solutions.

The authors propose a model based on BERT and sequence attention for automatic ICD code assignment. This model outperforms the state-of-the-art model on the MIMIC-III dataset, a widely used dataset for medical informatics research.

The study contributes a method for using BERT on documents and sequence attention to capture important sequence information in documents. The authors highlight that many studies have been conducted using machine learning and deep learning algorithms to predict various classification tasks in NLP

In the context of your research project on classifying engineering design project ideas as novel and feasible, this paper provides insights into the use of BERT and sequence attention for automatic code assignment. While the paper focuses on ICD codes in clinical notes, the methods and findings could potentially be adapted to classify engineering design ideas. The use of BERT and sequence attention could help capture important information in the descriptions of the design ideas, aiding in their classification based on novelty and feasibility.

Possible Pre-trained Models that can be used

Hugging Face pre-trained models have been utilized in the literature for novelty and feasibility prediction tasks using a variety of popular approaches and methodology. Some significant examples are provided below:

1. Fine-tuning BERT for classification: BERT (Bidirectional Encoder Representations from Transformers) is a well-liked pre-trained model from Hugging Face. In order to forecast the uniqueness and viability of engineering projects, researchers have enhanced BERT by adding a classification layer on top of the pre-trained model. The model is trained using methods such as binary classification or multi-class classification. The input data often consists of project descriptions or pertinent textual material.
2. GPT's text creation and classification needs some fine-tuning: Another popular pre-trained model is GPT (Generative Pre-trained Transformer). GPT has been improved for originality and feasibility prediction by completing classification tasks or project description generation during training. GPT has occasionally been used to come up with project ideas or assess the viability of given project concepts.
3. Using transformer-based designs: Other transformer-based architectures included in Hugging Face's Transformers library have been investigated in addition to BERT and GPT. For jobs requiring novelty and feasibility prediction, models like RoBERTa, DistilBERT, and XLNet have been improved. These versions offer different iterations of the transformer architecture and can be customized to meet the needs of certain projects.
4. Transfer learning with domain-specific data: In order to adapt pre-trained models to domain-specific data in the engineering field, researchers have used transfer learning approaches. The models can learn domain-specific features and enhance prediction performance by being fine-tuned with annotated datasets unique to novelty and feasibility prediction.
5. Ensemble models: To improve the precision and robustness of novelty and feasibility prediction, some studies have integrated numerous pre-trained models from Hugging Face in ensemble designs. The ensemble approach can produce more accurate forecasts and reduce individual model biases by combining predictions from various models.

It is also important to note that the precise strategy and fine-tuning methods may change depending on the particular specifications of the novelty and feasibility prediction objective and the data at hand. The particular attributes of the engineering projects and the targeted performance metrics should serve as a guidance for selecting models and methodologies.

In order to comprehend the fine-tuning procedure and learn how to modify the pre-trained models to your particular purpose, consult the documentation and examples offered by Hugging Face's Transformers library before putting the chosen strategy into practice.

Chat during feasibility and novelty phase -

Novelty Analysis

1. (If the novelty score is high) Your idea is indeed innovative! This project has the potential to bring fresh perspectives to the field. Let's continue exploring this exciting journey.

2. (If the novelty score is low) While the idea behind your project is robust, similar concepts have been explored in the past. But don't worry, we can work together to enhance its novelty. Try considering using different methodologies, materials, or technologies.

Feasibility Analysis

1. *(If the feasibility score is high)* Your project appears to be quite feasible given the constraints and objectives you've shared. We're on the right path to creating a successful design.

2. *(If the feasibility score is low)* Although there might be some challenges in executing this project, I'm confident we can address them. Try another way to improve the feasibility of your design.

Fantastic work so far! Together, we'll refine your design until it's both novel and feasible.

Proposed Methodology for Novelty and Feasibility Classification Model

The process involves several steps, including data preparation, model selection, training, and evaluation.

Data Preparation
The first step in our methodology is data preparation. We start with a dataset that contains text descriptions of engineering design project ideas. Each idea in the dataset has **two labels: one for novelty and one for feasibility.** The data preparation process involves tokenizing the text descriptions into a format that can be understood by the machine learning model. We use a pre-trained tokenizer for this purpose.

Model Selection
Next, we select a pre-trained model for our task. In this case, we use a model called DistilBERT, which is a smaller version of the BERT model. (This model does have a few constraints though which we can currently ignore as it won't be a big problem for our use case as of now - input text size limit)

Model Training
Once we have prepared our data and selected our model, the next step is to train the model on our dataset. During training, the model learns to associate the text descriptions of the project ideas with their corresponding novelty and feasibility labels. The training process is carried out using the Hugging Face's Trainer class.

Model Evaluation
After the model has been trained, we evaluate its performance on a separate test set that the model has not seen during training. This gives us an unbiased estimate of how well the model is likely to perform on unseen data.

Conclusion
In conclusion, our methodology involves preparing the data, selecting a pre-trained model, training the model on our dataset, and evaluating its performance.

----

1. Use scores and re-run the QnA till we get satisfactory results. Then end
2. A model to detect and check if the responses are proper,(some kind of sentiment analysis) give option to end the conversation if the user isn't in the right spirits.

Model Development

Novelty Model

1. **Result using distilbert-base-uncased, on a novelty classification model**

```
              precision    recall  f1-score   support

           0       0.62      0.84      0.71        82
           1       0.32      0.12      0.18        49

    accuracy                           0.57       131
   macro avg       0.47      0.48      0.44       131
weighted avg       0.50      0.57      0.51       131

Enter a description: A washing machine that uses solar power to run
Novelty Score: 0.4237995445728302
```

Observations -

The model is not capable of correctly predicting the results

2. **Using bert-base-uncased**

Observations -

1. When I used the 623 row dataset, I was able to get an accuracy of 67% in 2 epochs, but when I tested the model out, it was not accurate, it couldn't correctly determine the novelty score.

3. Using ChatGPT data

I am now adding 200 more rows to the dataset by generating descriptions and novelty score from chatGPT and running for 7 epochs

Total rows = 800

Result = Accuracy of 70%

Observations -

> The accuracy has remained constant at around 70% after 5 epochs. But upon testing, I am still not convinced at the models ability to correctly classify. But this model can be used as a prototype

4. Trying to use https://huggingface.co/SamLowe/roberta-base-go_emotions model

```
Some weights of RobertaModel were not initialized from the model checkpoint a
y initialized: ['roberta.pooler.dense.bias', 'roberta.pooler.dense.weight']
You should probably TRAIN this model on a down-stream task to be able to use
huggingface/tokenizers: The current process just got forked, after parallelis
to avoid deadlocks...
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(tru
Validation Accuracy: 0.0
Validation Accuracy: 0.0
Validation Accuracy: 0.0
Validation Accuracy: 0.0
Validation Accuracy: 0.0
(venv) naveenrenji@Naveens-MacBook-Air BackEnd % b
```

> Observations - This model failed to classify at all

5. Added Total 800 more rows from ChatGPT

Number of novel descriptions: 722
Number of non-novel descriptions: 673

6. Training for 100 epochs

After training on the server for 100 epochs, I noticed that though there wasnt a significant jump in the model accuracy, it peaked at 84% at multiple random epochs indicating a slight randomness in accuracy between the ranges of 79-84%.

```
Train loss 0.6118211783468723
Epoch 000: 100%|██████████| 40/40 [05:34<00:00,  6.55s/it,
loss=0.612]          precision    recall  f1-score   support

           0       0.70      0.89      0.79        66
           1       0.88      0.66      0.75        74

    accuracy                           0.77       140
   macro avg       0.79      0.78      0.77       140
weighted avg       0.79      0.77      0.77       140

Train loss 0.4345655746757984
                                                                    precision    recall  f1-score
support
Epoch 001: 100%|██████████| 40/40 [05:32<00:00,  6.53s/it,
loss=0.435]
           0       0.74      0.91      0.82        66
           1       0.90      0.72      0.80        74

    accuracy                           0.81       140
   macro avg       0.82      0.81      0.81       140
weighted avg       0.82      0.81      0.81       140

Train loss 0.3301972871646285
Epoch 002: 100%|██████████| 40/40 [05:32<00:00,  6.52s/it,
loss=0.33]          precision    recall  f1-score   support

           0       0.72      0.91      0.81        66
           1       0.89      0.69      0.78        74

    accuracy                           0.79       140
   macro avg       0.81      0.80      0.79       140
weighted avg       0.81      0.79      0.79       140

Train loss 0.25225600469857457
                                                                    precision    recall  f1-score
support
Epoch 003: 100%|██████████| 40/40 [05:30<00:00,  6.44s/it,
loss=0.252]
           0       0.74      0.89      0.81        66
           1       0.88      0.72      0.79        74

    accuracy                           0.80       140
   macro avg       0.81      0.81      0.80       140
weighted avg       0.81      0.80      0.80       140
```

```
Train loss 0.21181377209722996
Epoch 004: 100%|██████████| 40/40 [05:29<00:00,  6.48s/it,
loss=0.212]              precision    recall  f1-score   support

           0       0.70      0.89      0.79        66
           1       0.88      0.66      0.75        74

    accuracy                           0.77       140
   macro avg       0.79      0.78      0.77       140
weighted avg       0.79      0.77      0.77       140

Train loss 0.145381412608549
                                                                              precision    recall  f1-score
support
Epoch 005: 100%|██████████| 40/40 [05:29<00:00,  6.46s/it,
loss=0.145]
           0       0.73      0.88      0.80        66
           1       0.87      0.72      0.79        74

    accuracy                           0.79       140
   macro avg       0.80      0.80      0.79       140
weighted avg       0.81      0.79      0.79       140

Train loss 0.12519404166378081
Epoch 006: 100%|██████████| 40/40 [05:27<00:00,  6.44s/it,
loss=0.125]              precision    recall  f1-score   support

           0       0.77      0.83      0.80        66
           1       0.84      0.78      0.81        74

    accuracy                           0.81       140
   macro avg       0.81      0.81      0.81       140
weighted avg       0.81      0.81      0.81       140

Train loss 0.10868801819160581
                                                                              precision    recall  f1-score
support
Epoch 007: 100%|██████████| 40/40 [05:30<00:00,  6.50s/it,
loss=0.109]
           0       0.79      0.82      0.81        66
           1       0.83      0.81      0.82        74

    accuracy                           0.81       140
   macro avg       0.81      0.81      0.81       140
weighted avg       0.81      0.81      0.81       140

Train loss 0.13661045250482856
```

```
Epoch 008: 100%|██████████████████████| 40/40 [05:29<00:00,  6.49s/it,
loss=0.137]           precision    recall  f1-score   support

         0       0.83      0.74      0.78        66
         1       0.79      0.86      0.83        74

   accuracy                           0.81       140
  macro avg       0.81      0.80      0.80       140
weighted avg       0.81      0.81      0.81       140

Train loss 0.14223848176188766
                                                                                   precision    recall  f1-score
support
Epoch 009: 100%|██████████████████████| 40/40 [05:30<00:00,  6.51s/it,
loss=0.142]
         0       0.79      0.79      0.79        66
         1       0.81      0.81      0.81        74

   accuracy                           0.80       140
  macro avg       0.80      0.80      0.80       140
weighted avg       0.80      0.80      0.80       140

Train loss 0.0868520706659183
Epoch 010: 100%|█████████████████████| 40/40 [05:30<00:00,  6.42s/it,
loss=0.0869]           precision    recall  f1-score   support

         0       0.72      0.86      0.79        66
         1       0.85      0.70      0.77        74

   accuracy                           0.78       140
  macro avg       0.79      0.78      0.78       140
weighted avg       0.79      0.78      0.78       140

Train loss 0.052843635971657935
                                                                                   precision    recall  f1-score
support
Epoch 011: 100%|█████████████████████| 40/40 [05:30<00:00,  6.57s/it, loss=0.0528]
         0       0.72      0.88      0.79        66
         1       0.87      0.70      0.78        74

   accuracy                           0.79       140
  macro avg       0.80      0.79      0.79       140
weighted avg       0.80      0.79      0.78       140

Train loss 0.033097050385549664
Epoch 012: 100%|█████████████████████| 40/40 [05:34<00:00,  6.54s/it,
loss=0.0331]           precision    recall  f1-score   support
```

```
         0       0.80      0.83      0.81        66
         1       0.85      0.81      0.83        74

    accuracy                           0.82       140
   macro avg       0.82      0.82      0.82       140
weighted avg       0.82      0.82      0.82       140

Train loss 0.025647191412281244
                                                                                               precision    recall  f1-score
support
Epoch 013: 100%|████████████████████| 40/40 [05:31<00:00,  6.44s/it,
loss=0.0256]
         0       0.82      0.80      0.81        66
         1       0.83      0.84      0.83        74

    accuracy                           0.82       140
   macro avg       0.82      0.82      0.82       140
weighted avg       0.82      0.82      0.82       140

Train loss 0.01943690625485033
Epoch 014: 100%|████████████████████| 40/40 [05:29<00:00,  6.47s/it,
loss=0.0194]              precision    recall  f1-score   support

         0       0.77      0.85      0.81        66
         1       0.85      0.77      0.81        74

    accuracy                           0.81       140
   macro avg       0.81      0.81      0.81       140
weighted avg       0.81      0.81      0.81       140

Train loss 0.01509126354358159
                                                                                               precision    recall  f1-score
support
Epoch 015: 100%|████████████████████| 40/40 [05:31<00:00,  6.57s/it,
loss=0.0151]
         0       0.81      0.83      0.82        66
         1       0.85      0.82      0.84        74

    accuracy                           0.83       140
   macro avg       0.83      0.83      0.83       140
weighted avg       0.83      0.83      0.83       140

Train loss 0.016022863640682772
Epoch 016: 100%|████████████████████| 40/40 [05:34<00:00,  6.52s/it,
loss=0.016]              precision    recall  f1-score   support

         0       0.77      0.85      0.81        66
         1       0.85      0.77      0.81        74
```

accuracy 0.81 140
macro avg 0.81 0.81 0.81 140
weighted avg 0.81 0.81 0.81 140

Train loss 0.01854711733176373
precision recall f1-score
support
Epoch 017: 100%|██████████| 40/40 [05:33<00:00, 6.53s/it,
loss=0.0185]
0 0.82 0.82 0.82 66
1 0.84 0.84 0.84 74

accuracy 0.83 140
macro avg 0.83 0.83 0.83 140
weighted avg 0.83 0.83 0.83 140

Train loss 0.025062907882966102
Epoch 018: 100%|██████████| 40/40 [05:32<00:00, 6.42s/it,
loss=0.0251] precision recall f1-score support

0 0.77 0.83 0.80 66
1 0.84 0.78 0.81 74

accuracy 0.81 140
macro avg 0.81 0.81 0.81 140
weighted avg 0.81 0.81 0.81 140

Train loss 0.01669023621943779
precision recall f1-score
support
Epoch 019: 100%|██████████| 40/40 [05:31<00:00, 6.61s/it,
loss=0.0167]
0 0.77 0.85 0.81 66
1 0.85 0.77 0.81 74

accuracy 0.81 140
macro avg 0.81 0.81 0.81 140
weighted avg 0.81 0.81 0.81 140

Train loss 0.018511902453610674
Epoch 020: 100%|██████████| 40/40 [05:34<00:00, 6.55s/it,
loss=0.0185] precision recall f1-score support

0 0.80 0.80 0.80 66
1 0.82 0.82 0.82 74

accuracy 0.81 140

```
  macro avg       0.81      0.81      0.81       140
weighted avg       0.81      0.81      0.81       140

Train loss 0.007889027759665624
                                                                              precision    recall  f1-score
support
Epoch 021: 100%|██████████| 40/40 [05:33<00:00,  6.51s/it, loss=0.00789]
           0       0.78      0.85      0.81        66
           1       0.85      0.78      0.82        74

    accuracy                           0.81       140
   macro avg       0.82      0.82      0.81       140
weighted avg       0.82      0.81      0.81       140

Train loss 0.009270745079265907
Epoch 022: 100%|██████████| 40/40 [05:33<00:00,  6.56s/it, loss=0.00927]
precision    recall  f1-score   support

           0       0.81      0.83      0.82        66
           1       0.85      0.82      0.84        74

    accuracy                           0.83       140
   macro avg       0.83      0.83      0.83       140
weighted avg       0.83      0.83      0.83       140

Train loss 0.006975636151037179
                                                                              precision    recall  f1-score
support
Train loss 0.0002815419800754171
                                                                              precision    recall  f1-score
support
Epoch 093: 100%|██████████| 40/40 [05:33<00:00,  6.55s/it, loss=0.000282]
           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

Train loss 0.00024140790810633916
Epoch 094: 100%|██████████| 40/40 [05:33<00:00,  6.45s/it, loss=0.000241]
precision    recall  f1-score   support

           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
```

```
weighted avg       0.79      0.79      0.79       140

Train loss 0.0002071537788651767
                                                                                          precision    recall  f1-score
support
Epoch 095: 100%|██████████| 40/40 [05:31<00:00,  6.45s/it, loss=0.000207]
           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

Train loss 0.00019605679644882912
Epoch 096: 100%|██████████| 40/40 [05:32<00:00,  6.53s/it, loss=0.000196]
precision    recall  f1-score   support

           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

Train loss 0.00018419180323689944
                                                                                          precision    recall  f1-score
support
Epoch 097: 100%|██████████| 40/40 [05:35<00:00,  6.62s/it, loss=0.000184]
           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

Train loss 0.0001677649066550657
Epoch 098: 100%|██████████| 40/40 [05:30<00:00,  6.50s/it, loss=0.000168]
precision    recall  f1-score   support

           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

Train loss 0.00016084624567156425
```

```
                                                                          precision    recall  f1-score
support
Epoch 099: 100%|██████████████████████████| 40/40 [05:30<00:00,  6.49s/it, loss=0.000161]
           0       0.76      0.80      0.78        66
           1       0.81      0.77      0.79        74

    accuracy                           0.79       140
   macro avg       0.79      0.79      0.79       140
weighted avg       0.79      0.79      0.79       140

saving it
saved it
```

```
  support: 100%|████████████████████| 23/23 [07:04<00:00, 16.40s/it, loss=0.337]

           0       0.87      0.58      0.69        45
           1       0.62      0.89      0.73        35

    accuracy                           0.71        80
   macro avg       0.74      0.73      0.71        80
weighted avg       0.76      0.71      0.71        80

Train loss 0.2996451051338859
Epoch 006: 100%|██████████████████████| 23/23 [06:48<00:00, 14.84s/it, loss=0.3]
  support

           0       0.82      0.60      0.69        45
           1       0.62      0.83      0.71        35

    accuracy                           0.70        80
   macro avg       0.72      0.71      0.70        80
weighted avg       0.73      0.70      0.70        80

saving it
saved it
○ (venv) naveenrenji@Naveens-MacBook-Air BackEnd %
```

Feasibility Model

I have started working on the feasibility model and realized that there is a severe unbalance in the data. There are very few rows that classify text as non feasible. This lack of balance will result in a very biased classifier that will not be able to accurately classify between feasible and infeasible.

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         5
           1       0.92      1.00      0.96        61

    accuracy                           0.92        66
   macro avg       0.46      0.50      0.48        66
weighted avg       0.85      0.92      0.89        66
```

As you can see, we have achieved an accuracy of 92% while testing 66 rows, but out of the 66 rows, only 5 were 0 (non-feasible) whereas 61 were 1 (feasible)

So I performed an Exploratory Data Analysis

Number of feasible descriptions: 575
Number of non-feasible descriptions: 77

I will therefore now try to rebalance the Data by generating more non-feasible rows using ChatGPT.

After ChatGPT Data -

Number of feasible descriptions: 663
Number of non-feasible descriptions: 552

```
Train loss 0.142470434175006
Epoch 004: 100%|██████████| 35/35 [10:05<00:00, 13.26s/it, loss=0.142]              precision    recall  f1-score   support

           0       0.82      0.80      0.81        46
           1       0.88      0.89      0.89        76

    accuracy                           0.86       122
   macro avg       0.85      0.85      0.85       122
weighted avg       0.86      0.86      0.86       122
```

Validation Model

First -

A model that will determine whether the user's input is in fact describing an engineering design for a washing machine.

The idea is used to create a context about washing machine designs and ideas and use that to determine if a users input aligns with the context using the Bert model that was trained on the stanford QnA dataset.

This approach did not work out after multiple various approaches and tries.

Second -

So now I have started developing a new Validation model that uses Bert and I am generating a dataset of valid and invalid data using chatgpt for the preliminary training.

Dataset size is only - 238 rows of equally distributed valid and invalid descriptions

Observations - on a limited dataset, the model was able to predict with 93.75% accuracy. But this result may be unreliable and requires further investigations with a bigger dataset trained for more epochs.

Given a query -

```python
# Example usage
description = "have a washing machine that can detect if clothes have stains."
print(f'Prediction for description: {predict_washing_machine(description)}')
```

Output -

```
Accuracy: 93.75%
Prediction for description: yes
(venv) naveenrenji@Naveens-MacBook-Air Validation %
```

Third -

I made modifications to the dataset and created a more diverse set of descriptions.

```
Accuracy: 95.83%
Prediction for description:  have a washing machine that can detect if clothes have stains. = yes
Prediction for description: have a juice blender that runs on solar energy. = no
```

We can now notice that it is able to accurately differentiate between descriptions of washing machines and descriptions of other things.

Phase 3 - Creating The Pre-trained models and functions to use them for the backend API

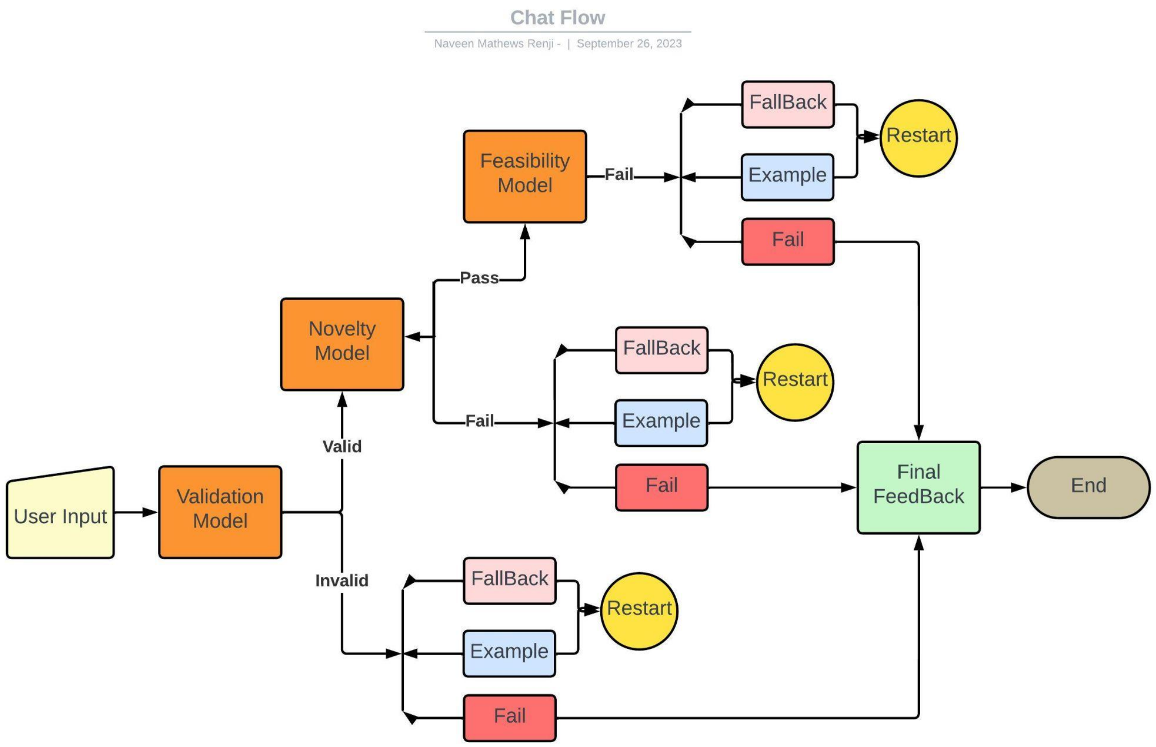
Created the condition to save a fine tuned model as it reached the global maximum accuracy.
Completed the finetuning for Novelty model and for Validation Model, feasibility model is in progress

Completed the fine tuning and classification function of the feasibility model

Created API -

1. Validation model - /validation (post request - send the message and returns yes for valid and no for invalid)
2. Novelty model - /novelty (post request - send the message and returns the probability of message being novel)
3. Feasibility model - /feasibility (post request - send the message and returns the probability of message being feasible)

Chat Flow



- Chatbot Flow

1. Input -> Validation Test -> Valid or Invalid
    a. If Valid ->Novelty Evaluation
    b. If Invalid -> 1st time is a, 2nd time is b, third and final time is c
        i. Fallback -> restart
        ii. Example -> restart
        iii. Fail -> Final Feedback -> end
2. Novelty Evaluation -> Pass or Fail
    a. Pass -> Feasibility Evaluation
    b. Fail -> 1st time is a, 2nd time is b, third and final time is c
        i. Fallback -> restart
        ii. Example -> restart
        iii. Fail -> Final Feedback -> end
3. Feasibility Evaluation -> Pass or Fail
    a. Pass -> Final Feedback -> end
    b. Fail -> 1st time is a, 2nd time is b, third and final time is c
        i. Fallback -> restart
        ii. Example -> restart
        iii. Fail -> Final Feedback -> end

Start

- Initial Greeting: Hello! My name is Aida. I am here to help you design the next generation of sustainable washing machine.
- Now, tell me your idea/your thoughts. Please use complete sentences/write at least 10 words so I have a better understanding blah blah.

Input Validation

- Valid Input: Your idea is Great! Let's move on to the novelty evaluation.
- Invalid Input (1st time): Oops! It looks like there's something missing or incorrect in your input. Could you please try again?
- Invalid Input (2nd time): Still not quite right. Here's an example of how to format your input: "The machine is manufactured from parts that generate less electronic waste and is easy to take apart and recycle." . Please try again.
- Invalid Input (3rd time): Unfortunately, the input is still not valid. Please review the guidelines and try again later. [Final Feedback]

Novelty Evaluation

- Pass: Your idea is indeed innovative! This project has the potential to bring fresh perspectives to the field. Let's continue exploring this exciting journey.
- Fail (1st time): While the idea behind your project is robust, similar concepts have been explored in the past. But don't worry, we can work together to enhance its novelty. Try considering using different methodologies, materials, or technologies.
- Fail (2nd time): It seems the idea still lacks uniqueness. Have you considered pivoting the project's focus? Let's try again.
- Fail (3rd time): Despite our efforts, the novelty aspect isn't meeting the criteria. You may want to revisit the drawing board. [Final Feedback]

Feasibility Evaluation

- Pass: Your project appears to be quite feasible given the constraints and objectives you've shared. We're on the right path to creating a successful design.
- Fail (1st time): Although there might be some challenges in executing this project, I'm confident we can address them. Try another way to improve the feasibility of your design.
- Fail (2nd time): The feasibility still seems to be a concern. Have you considered alternative approaches or materials? Let's try again.
- Fail (3rd time): Unfortunately, the design still doesn't meet the feasibility criteria. It might be time to reconsider some fundamental aspects. [Final Feedback]

End

- Success: Fantastic work so far! Together, we've refined your design until it's both novel and feasible. Thank you for using the Engineering Design Evaluation Chatbot. Failure: Thank you for your efforts. While the design didn't meet all the criteria this time, don't be discouraged. Keep iterating and come back when you're ready.

Feedback

- Thank you for your time, I hope you enjoyed this experiment. Please share your valuable feedback on your experience by filling out the survey.

Alternate Flow

- Chatbot: Hello! I'm Aida, an AI chatbot to help you generate engineering design concepts. Today, we will work together on washing machines. Let's begin by having you describe your concept.
- User provides inputs
- Chatbot: Thank you for sharing those details on your washing machine design concept.
- Validity Analysis:
- 1st failure: I appreciate you taking the time to describe your idea. However, I'm having some trouble understanding the key details pertaining specifically to a washing machine design.
- Could you please rephrase the description and objective to focus on the core washing machine features and purpose? Providing more specifics would really help me understand your idea.
- 2nd failure: Unfortunately I'm still unable to fully comprehend the washing machine design concept based on the current description. Please don't be discouraged! With a bit more clarification on the machine's intended functionality and features, I'll be better equipped to assess validity. Let's collaborate to articulate an accurate portrayal of this inventive washing machine idea.

- 3rd failure: I seem to be missing some critical information needed to validate this as a washing machine design concept. For my evaluation, please rephrase the description and objective to only focus on detailing a novel washing machine idea, including the key features and functionality. I'm committed to working together until I fully understand your vision.
- (If valid): Excellent, thank you for providing a clear and valid washing machine design description. Let's now assess novelty:
- Novelty Analysis:
- 1st failure: I see potential for innovation in this washing machine concept, though some similarities exist with previous designs. Let's ideate tweaks to differentiate your approach, like using alternative washing methods or creative new technologies. I'm excited to help make this more novel!
- 2nd failure: While aspects seem familiar, I believe originality is within reach. Let's explore modifying certain components or processes to create a truly distinctive washing machine. Bounce ideas off me - I'm happy to collaborate on enhancing novelty.
- 3rd failure: It seems we need a real breakthrough concept here to distinguish this from prior washing machine designs. Take some time to freely imagine without limits - this will lead to innovation! Picture what's currently missing in the field, and how your vision could fill that gap. I know that working together, we can develop something unprecedented.
- (If novel): Your washing machine concept is highly innovative! You've achieved something original in this space. Let's proceed to evaluate feasibility:
- Feasibility Analysis:
- 1st failure: Let's examine potential feasibility barriers and brainstorm solutions. Perhaps different materials, tools, or methods could improve viability. I believe revising certain elements will lead to a more executable washing machine design.
- 2nd failure: It seems there are still some challenges to feasibly implementing this concept as described. Don't be discouraged! Let's pinpoint just one or two changes that might significantly improve feasibility. This could involve the process, resources, or techniques. I'm excited to optimize collaboratively.
- 3rd failure: While the current washing machine concept presents some challenges, I believe we can find ways to make this design work. Think creatively for solutions. Even small tweaks to the approach could greatly improve feasibility. I'm committed to working together until we shape something both novel and viable.
- (If feasible): Your washing machine design appears quite feasible given the constraints and objectives. We've developed an innovative concept that can be successfully executed - congratulations!
- Final Feedback: Thank you for sharing your inventive washing machine design ideas with me today. As we conclude, I would greatly appreciate any brief feedback you can provide about your experience using this chatbot to evaluate novelty, feasibility, and validity of washing machine engineering concepts. Please let me know if you have any suggestions for how I can improve. Your insights will help me enhance this chatbot experience for washing machine designers.

Names -

1. EDGE (Engineering Design Guide and Evaluator)
2. Aida (Artificial Intelligence Design Assistant)

Interaction flow (high-inspiration/low empathy version)

- Bot: Hello! My name is Aida. I am here to help you design the next generation of sustainable washing machines. Please use complete sentences to describe your idea.
- [Wait for any user input; if not, prompt the next message]
- Now, tell me your idea of the washing machines.
- [User input]

Input Validation

- Bot (Valid Input): Thanks for your input.
- Invalid Input (1st time): Your input is not recognized as a valid washing machine design. Please try to make more specific descriptions about washing machine design.
- Invalid Input (2nd time): It's still not quite relevant to washing machines. Here's an example of the design idea: "The machine is manufactured from parts that generate less electronic waste and is easy to take apart and recycle." Please try again.
- Invalid Input (3rd time): Unfortunately, the input seems irrelevant to washing machine design.

Novelty Evaluation

- Pass: Your idea is indeed innovative! This idea has the potential to bring novel perspectives to the design field.
- Fail (1st time): The idea is robust, but similar concepts have been explored in the past. You may consider using different methodologies, materials, or technologies. Please try again.

- Fail (2nd time): The idea still lacks uniqueness. Have you considered pivoting the idea's focus?
- Here is an example" Sense the load automatically and automatically determine the time to wash". Please elaborate the idea or brainstorm another one.
- Fail (3rd time): Thanks for your effort. There are some notions similar to the one you proposed.

Feasibility Evaluation

- Pass: Your idea appears to be quite feasible. You're on the right path to creating a successful design.
- Fail (1st time): It may be challenging in executing this idea. Please provide more details or propose a more feasible idea.
- Fail (2nd time): The feasibility still seems to be a concern. Here is an example of the feasible design idea: "The machine is manufactured from parts that generate less electronic waste and are easy to take apart and recycle." Please elaborate the idea or brainstorm another one.
- Fail (3rd time): Thanks for your effort. Unfortunately, the idea seems infeasible given the current technology.

Second Round (if pass both the first time)

- You did a great job! Your design is novel and feasible. Please come up with another idea of sustainable washing machines, potentially with a different focus. [Restart the loop]

End

- Thank you for participating in this design task. Please share your valuable feedback on your experience by filling out the survey [Survey linkhttps://qualtricsxmmq6rxgmlg.qualtrics.com/jfe/form/SV_5nHNE5TiFYEd85E]