PERSONALITY DETECTION
(Approach Methodology)

Dataset choices

1. **Essays**

   Essay is a large dataset based on the stream of consciousness that was collected by Pennebaker and Laura King according to the text generated by 2467 users between 1997 and 2004 were labeled based on classes of personality traits including Neuroticism (NEU), Extraversion (EXT), Openness to experience (OPN), Agreeableness (AGR), and conscientiousness (CON))

2. **My personality**

   MyPersonality was developed by Facebook which collected data by allowing users to participate in psychological research by filling in a personality questionnaire.

Adapting Large Models (LLMs) to New Tasks

Zero-shot learning - "*prompting*"

Few-shot learning - "*prompting with examples*"

Fine-Tuning - *dozens or fewer examples*

1. ***Full finetuning***
2. ***PEFT(Parameter efficient fine tuning)***
   - ***LoRA(Low-Rank Adaptation of large language)(Paper_link)***
   - ***QLoRA(Quantized Low-Rank Adaptation of large language)(paper_link)***
   - ***Prompt tuning(Paper_link)***
   - ***Prefix tuning(Paper_link)***

Model Training

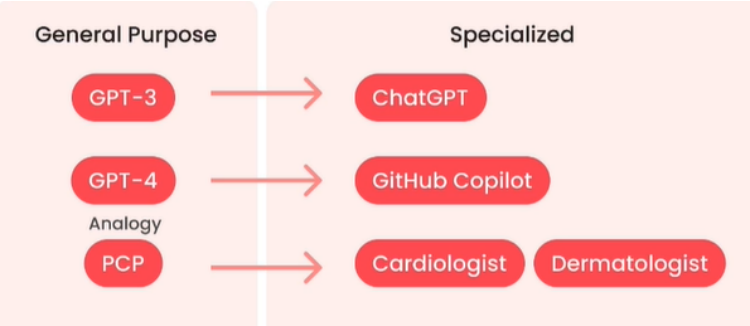
Model evaluation

Choosen LLM models

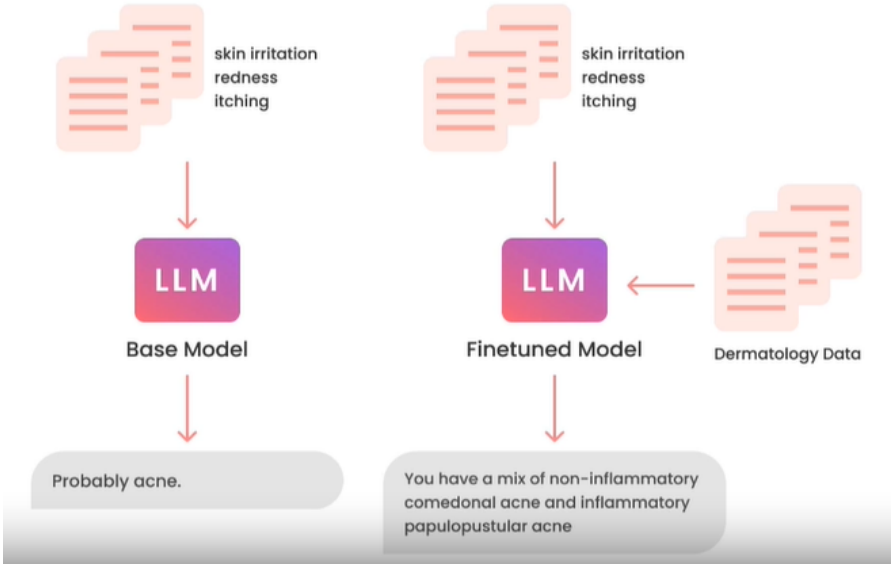
1. **Llama**
2. **Flan-t5**
3. **Falcon**

What is finetuning

- Making LLms to carry out specialized tasks



General purpose model(Base model) vs Finetuned model



- The base model can only predict the next token/output based on the input prompt and **does not have knowledge about a specific domain or task**, making it more generalized.

- On the other hand, during fine-tuning, we add knowledge to the base models about a specific domain/task, making it **highly specialized for that particular task.**

Approach 01- Prompting Technique and Human Evaluation

Some generalized LLM models were able to give good reliable predictions on our specific task(**Personality detection**) without finetuning to the specific task
By **zero shot** and **few shot prompting**

GPT3.5 turbo- - - > Chat gpt3.5

The output was only generated by prompting in **Chatgpt3.5(175 billion Parameter)** which was powered by **GPT3.5 turbo** LLM model which was trained for **chatting/(Q&A)** task

> The primary **Big Five personality trait** displayed in this statement is **Neuroticism (Emotional Stability)**. The writer's thoughts and emotions are reflective, expressing a range of feelings, concerns about growing old, self-doubt, discomfort with their writing being read, and a sense of introspection about their personal experiences and fears.

GPT-4- - - > Chat gpt4

GPT-4(1.76 trillion Parameter) excels at providing detailed responses **based on prompts**
The same input prompt which was used in the GPT-3.5 was used here but GPT-4 gives better and more detailed token prediction

> In summary, the writer's stream of consciousness touches upon multiple facets of the Big Five personality traits, with particular elements suggesting tendencies toward Openness, Conscientiousness, and Neuroticism. However, it's essential to understand that a single piece of writing might not capture the entirety of a person's personality, and the above analysis is based solely on the given text.

Flan-t5 (*Colab_notebook_link*)
(*Model_link_Huggingface*)

Tried for Facebook **My personality dataset**

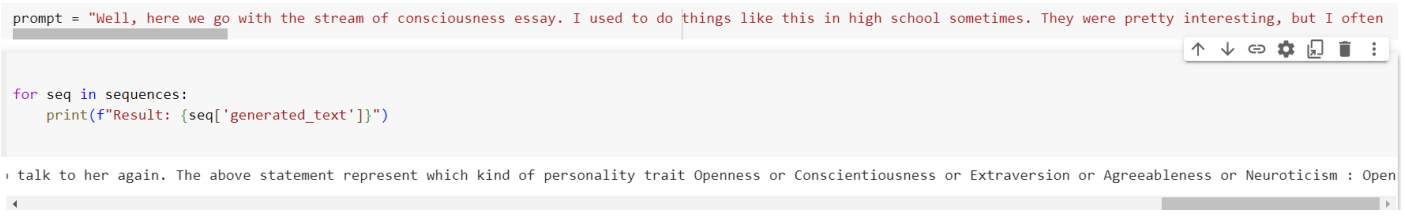
Tried with **Zero shot** prompting

```
---------------------------------------------------------------------------------------------------
INPUT PROMPT:

Predict whether the above sentence is Openness or Conscientiousness or Extraversion or Agreeableness or Neuroticism :

GO STATE BEAT THE BOBCATS!!!!!

Summary:

---------------------------------------------------------------------------------------------------
BASELINE HUMAN SUMMARY:
agreeableness

---------------------------------------------------------------------------------------------------
MODEL GENERATION - ZERO SHOT:
Conscientiousness
```

But it didn't match the original output

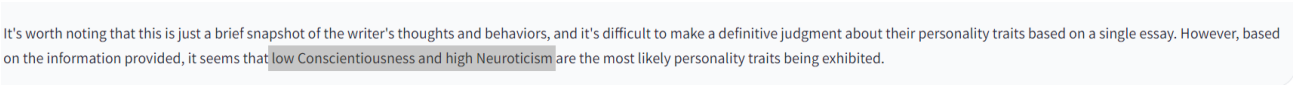
Falcon-7b-instruct(*Colab-notebook-link*)
(*Modellink-Huggingface*)

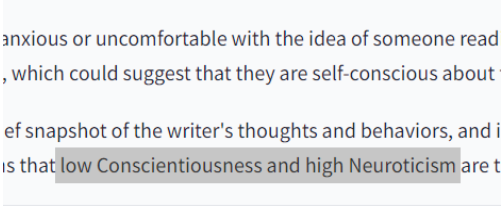
Tried with **zero shot** prompting



LLama 2-7b(*Model_link_Huggingface*)

Zero shot prompt





Vicuna 7b(*Model_link_Huggingface*)

It was unable to predict personality traits while prompting

> It is difficult to predict the personality traits of the writer based on this stream of consciousness essay. Personality traits are complex a
> piece of writing. Additionally, the writer may not be fully aware of their own personality traits and how they come across in their writir
> unique perspective and voice. A more in-depth analysis of the writer's behavior, thoughts, and feelings would be necessary to accurat

We can try out promoting with top LLM models in the space here....(Link)

In this research, I experimented with both **single-shot and few-shot prompting** with the LLM models mentioned above. Each model produced better predictions compared to zero-shot prompting. However, only **GPT-4** gives a reasonably good output.

For demonstration here I added screenshots of a **single and few shot prompting** for the **flan-t5** model on Facebook **my personality dataset**

Zero shot prompt

```
---------------------------------------------------------------------------------------------------
Example  1
---------------------------------------------------------------------------------------------------
INPUT PROMPT:

Predict whether the above sentence is Openness or Conscientiousness or Extraversion or Agreeableness or Neuroticism :

About to use my new alarm clock that projects time on the ceiling ;p Laziest thing ever and I'm gonna love it!

Summary:

---------------------------------------------------------------------------------------------------
BASELINE HUMAN SUMMARY:
conscientiousness
---------------------------------------------------------------------------------------------------
MODEL GENERATION - ZERO SHOT:
Neuroticism
```

One shot prompt

```
-----------------------------------------------------------
BASELINE HUMAN SUMMARY:
conscientiousness

-----------------------------------------------------------
MODEL GENERATION - ONE SHOT:
Conscientiousness
```

Few shot prompt

```
------------------------------------------------------------------
BASELINE HUMAN SUMMARY:
conscientiousness

------------------------------------------------------------------
MODEL GENERATION - FEW SHOT:
Conscientiousness
```

As a first step for getting a model to do a specific task according to our use case Before going to the finetuning technique we have to evaluate the LLM model by prompting and how effectively these tasks are done by the model.

The main drawback in the prompting is there is no specific way for model evaluation. We have to manually do the evaluation of the performance.

NOTE:- For prompting its better to use instruct models of LLMs on there the base model is trained additionally with chat power and have additional knowledge to understand the domain.

Below I listed top leader board **instruct LLMs**

1.'Platypus2-70B-instruct'-*Model_link*

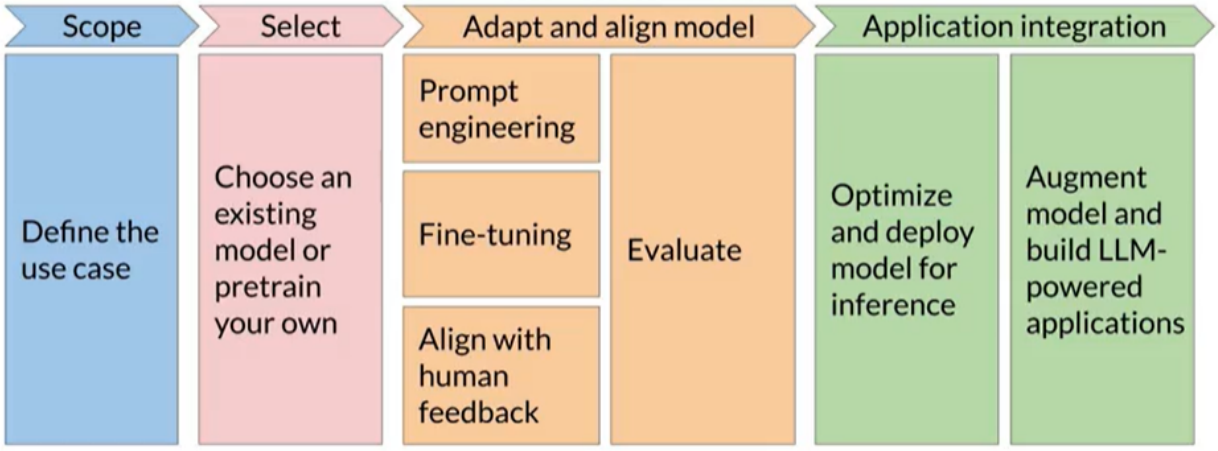
2.'Llama-2-70b-instruct'* -*Model_link

3.'falcon-40b-instruct' -*Model_link*

4.'CodeLlama-13b-Instruct-hf' - *Model_link*

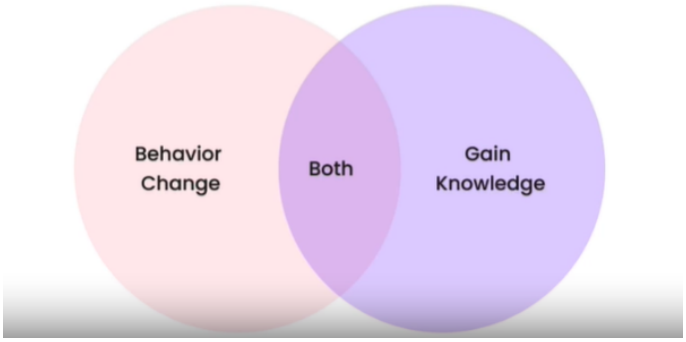
Planning...

1. Do the task by prompt Engineering on an LLM
2. Evaluate that the task is 'OK' with LLM
3. Evaluate with 1000 input-output pairs —> better than 'OK' from LLM
4. Finetune an LLM on this data
5. Use parameter-efficient finetuning(PEFT) to reduce the computational cost



Overall the research pipeline

Approach 02- Finetuning LLM



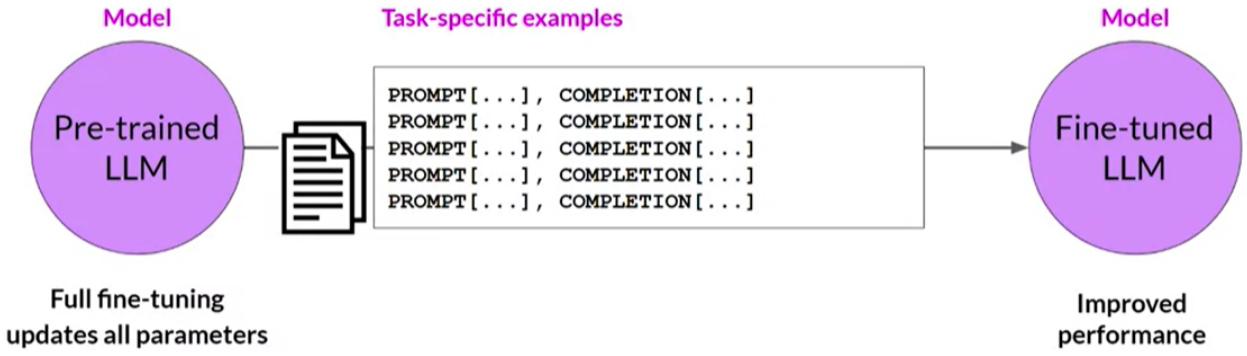
By finetuning LLM we change the base model to learn to respond more consistently and increase the knowledge of new specific concepts.

Fine-tuning tasks based on two approaches

Our personality detection research comes under **Extractio**n where we extract the big five personality traits from the given dialogue or conversation.

INSTRUCTION FINETUNING

Fine-tuning- It involves training an **LLM model** on a new, typically smaller, dataset**.**
Instruction finetuning- This is a variant of fine-tuning where the model is trained to follow instructions given in the input prompt.



Outline pipeline of how instruction finetuning happens

1.Single task instruction finetuning -Targets a specific task or domain.
2.Multi task instruction finetuning - Simultaneously targets multiple tasks or domains.

Our personality detection task comes under **Single task instruction finetuning** where train our model to do only a specific task for prediction of personality trait

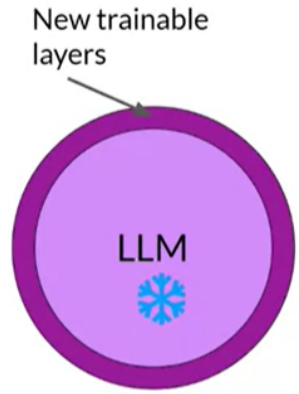
FINE-TUNING

Both **Full-finetuning** and **PEFT** are used to improve the performance of LLm model

1. **Full finetuning -** Full Fine-tuning is taking a pre-trained LLM model and training it further on a new task with new data. The entire pre-trained model is usually trained in **fine-tuning, including all its layers and parameters.**
   This process can be computationally expensive and time-consuming, especially for large models.

2. **Parameter-efficient finetuning(PEFT) -** parameter-efficient fine-tuning is a method of fine-tuning that focuses on training only a **subset of the pre-trained model's parameters**. This approach involves identifying the **most important parameters** for the new task and only updating those parameters during training.
   PEFT techniques
   1. **LoRA**
   2. **QLoRA**
   3. **Prefix-tuning**
   4. **Prompt-tuning(Soft prompt)**

Parameter-efficient finetuning(PEFT)



LLM with additional layers for PEFT

In PEFT the original weight gets frozen and an additional **trainable layer** is added according to our use case.

PEFT methods.

1. **Selective**- Select a subset of initial LLM parameters to finetune.
2. **Reparameterization**- Reparameterize model weights using low-rank representation
    - **LoRA**
    - **QLoRA**
3. **Additive-** Add trainable layers or parameters to the model
    - **Adapters**
    - **soft prompts**

LoRA

1. Freeze most of the original LLM weight
2. Inject 2 rank decomposition matrices
3. Train the weights of the smaller matrices while keeping the other weights freeze.

Colab notebooks trained with LoRA technique.

1.LoRA_on_Facebook_Mypersonality_Dataset- Flan-t5