PersonAI Final Report

Introduction

In the world of artificial intelligence, various breakthroughs have been witnessed in recent years. One such exciting development is the PersonAI project which aims to revolutionize psychology education by offering an engaging and practical training platform that harnesses the power of AI. At the heart of the project lies the GPT-2 model, developed by OpenAI, which generates human-like text, which are then analyzed to deduce personality traits based on the globally recognized Myers-Briggs Type Indicator (MBTI). This report presents an overview of the five unique models developed as part of the PersonAI project.

Related Work

The development of the PersonAI project draws inspiration from previous works in the field of artificial intelligence and psychology. Notably, studies on personality detection have laid the groundwork for the development of our models. The use of deep learning techniques, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, has been instrumental in understanding and interpreting text data for personality trait extraction. Our models aim to build upon these established techniques and further push the boundaries of what is possible in the realm of AI-driven psychology education.

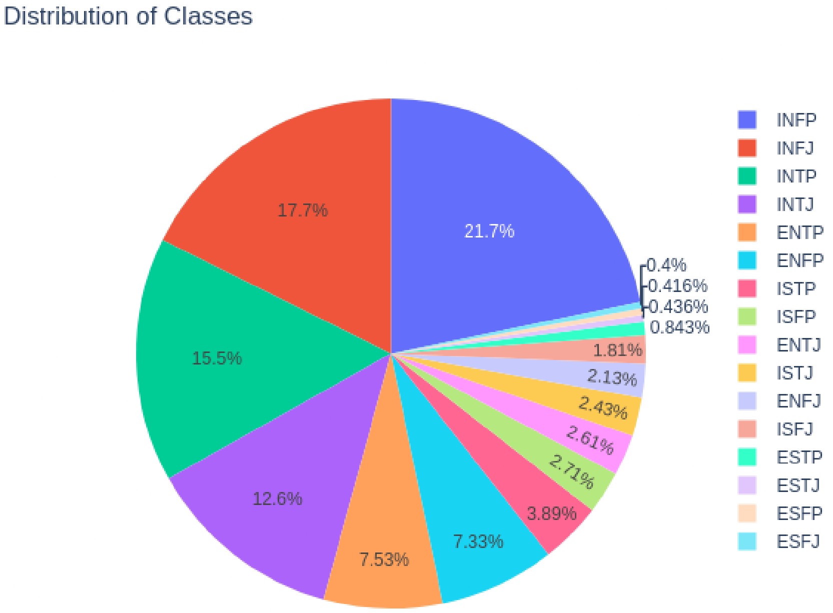
Related Papers:

1. "A sentiment-aware deep learning approach for personality detection from text", ScienceDirect, source.
2. "CharacterChat: Learning towards Conversational AI with Personalized Social Support", arxiv, source.

Proposed Methodology in PersonAI

For the PersonAI project, we've developed and fine-tuned five distinct models. Four of which are designed to extract a specific personality trait based on the Myers-Briggs Type Indicator (MBTI). These traits are Introversion (I) vs. Extroversion (E), Intuition (N) vs. Sensing (S), Thinking (T) vs. Feeling (F), and Judging (J) vs. Perceiving (P). We use the divide and conquer approach, breaking down the task of personality analysis into four parts for the four models, with each model conquering one specific trait. The results are then combined to form a comprehensive personality profile.

The fifth model is a fine-tuned GPT-2 model that generates human-like text. This text serves as the psychology puzzle as well as input for the other four models. All models are trained and fine-tuned on the same dataset, the distribution of which is shown below:



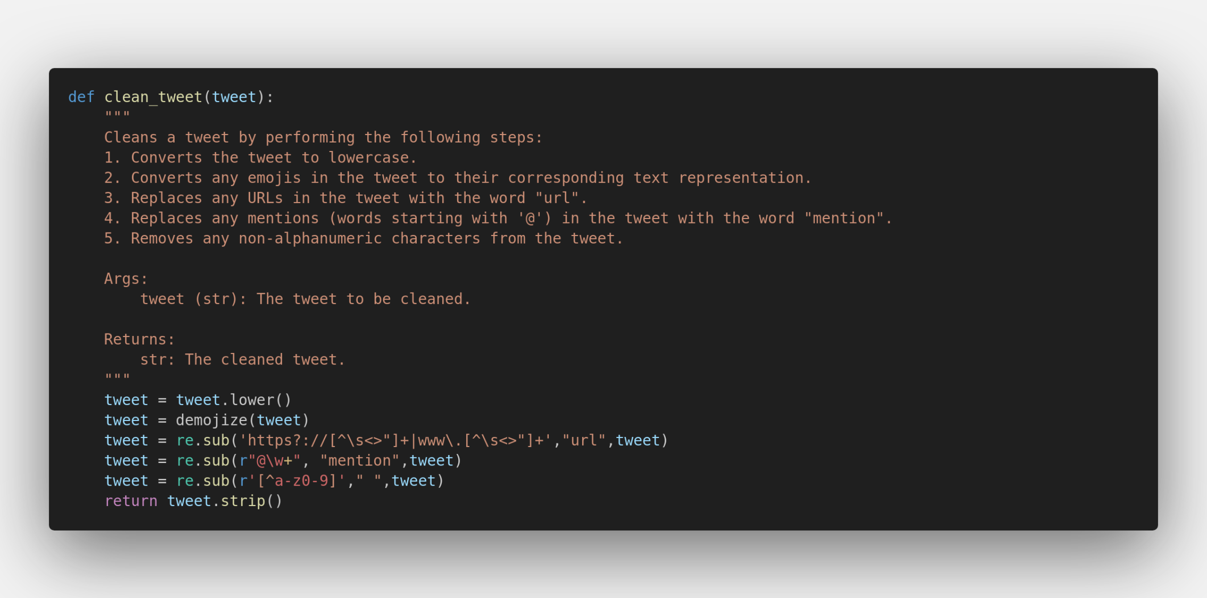
16 Personalities distribution in the dataset.

Data Cleaning and Preparation

As shown in the pie chart above, this dataset is imbalanced. However, it accurately represents the frequency of each personality type in the global population, according to the source. This leads us to the next step.

In order to allow our models to effectively analyze the data and derive valuable insights, we first performed thorough data cleaning and preparation.

Tweet Cleaning: The first step in our data preparation process was to clean the tweets. We utilized a custom function that performed several operations to clean the text data. This function was designed to standardize the text input and remove unnecessary noise from the data:



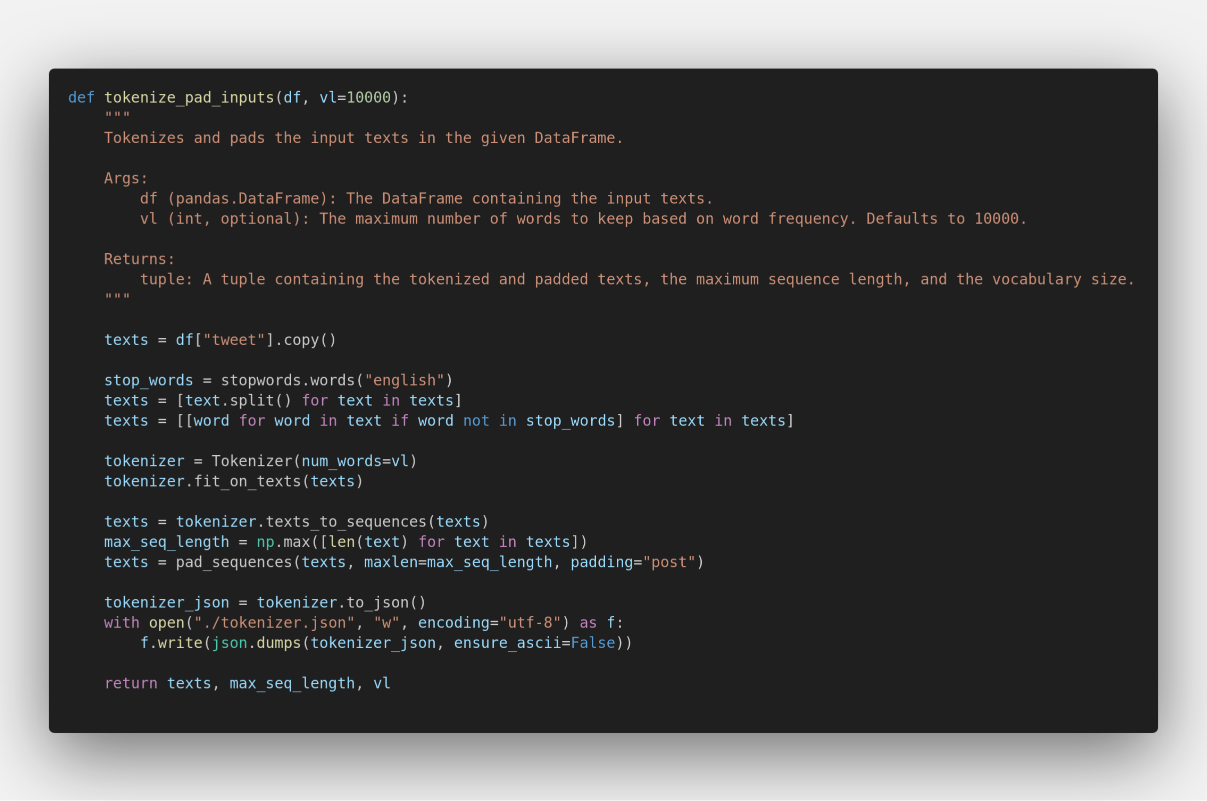
```python
def clean_tweet(tweet):
    """
    Cleans a tweet by performing the following steps:
    1. Converts the tweet to lowercase.
    2. Converts any emojis in the tweet to their corresponding text representation.
    3. Replaces any URLs in the tweet with the word "url".
    4. Replaces any mentions (words starting with '@') in the tweet with the word "mention".
    5. Removes any non-alphanumeric characters from the tweet.

    Args:
        tweet (str): The tweet to be cleaned.

    Returns:
        str: The cleaned tweet.
    """
    tweet = tweet.lower()
    tweet = demojize(tweet)
    tweet = re.sub('https?://[^\s<>"]+|www\.[^\s<>"]+',"url",tweet)
    tweet = re.sub(r"@\w+", "mention",tweet)
    tweet = re.sub(r'[^a-z0-9]'," ",tweet)
    return tweet.strip()
```

Cleaning tweets function (without removing stop words)

Tokenization and Padding: After cleaning the tweets, the next step was to tokenize the text and pad the tokenized sequences with zeros. Tokenization is the process of breaking down the text into individual words, or "tokens". This step is crucial in preparing the data for our models, as it transforms the raw text into a format that our models can understand and learn from. Padding, on the other hand, ensures that all text sequences are of the same length, which is a requirement for feeding data into our models.



```python
def tokenize_pad_inputs(df, vl=10000):
    """
    Tokenizes and pads the input texts in the given DataFrame.

    Args:
        df (pandas.DataFrame): The DataFrame containing the input texts.
        vl (int, optional): The maximum number of words to keep based on word frequency. Defaults to 10000.

    Returns:
        tuple: A tuple containing the tokenized and padded texts, the maximum sequence length, and the vocabulary size.
    """

    texts = df["tweet"].copy()

    stop_words = stopwords.words("english")
    texts = [text.split() for text in texts]
    texts = [[word for word in text if word not in stop_words] for text in texts]

    tokenizer = Tokenizer(num_words=vl)
    tokenizer.fit_on_texts(texts)

    texts = tokenizer.texts_to_sequences(texts)
    max_seq_length = np.max([len(text) for text in texts])
    texts = pad_sequences(texts, maxlen=max_seq_length, padding="post")

    tokenizer_json = tokenizer.to_json()
    with open("./tokenizer.json", "w", encoding="utf-8") as f:
        f.write(json.dumps(tokenizer_json, ensure_ascii=False))

    return texts, max_seq_length, vl
```

Tokenization and padding function

These steps helped us transform our raw text data into a suitable format for our models, facilitating efficient data analysis and meaningful insight extraction.

Detailed Description of the Models

Model 1: Fine-tuned GPT-2

The GPT-2, or Generative Pretrained Transformer 2, is a language prediction model developed by OpenAI. This model is designed to generate human-like text based on the input (prompt) it's given. In the context of the PersonAI project, we fine-tuned GPT-2 to generate detailed, realistic small paragraphs. These paragraphs then serve as the puzzle for psychology student as well as the basis for the personality analysis conducted by the other models.

Fine-tuning involved training the GPT-2 model on a dataset of tweets, allowing it to learn and understand the nuances of personality traits as expressed in text. The goal was to ensure the model could generate small paragraphs that were as diverse and complex as real human would write.

The GPT-2 model uses transformer architectures, allowing it to generate coherent and contextually relevant sentences by predicting the next word in a sentence. It's designed to understand the context of the input text, enabling it to produce relevant and coherent output. The fine-tuned GPT-2 model's ability to generate rich, nuanced human-like text forms the cornerstone of the PersonAI project.

Model 2: Introversion (I) – Extroversion (E)

This model focuses on identifying whether the paragraph, generated by the fine-tuned GPT-2 model, exhibits more introverted or extroverted tendencies. It analyzes text data to look for signs of social engagement, outgoingness, or preference for solitude.

Model 3: Intuition (N) – Sensing (S)

The third model distinguishes between intuitive and sensing personality types. It identifies whether a character’s words, generated by the fine-tuned GPT-2 model, relies more on their intuition and abstract thinking or prefers to rely on their senses and concrete information.

Model 4: Thinking (T) – Feeling (F)

The fourth model aims to differentiate between thinking and feeling personality types. It scrutinizes the text data, generated by the fine-tuned GPT-2 model, to understand if the character makes decisions based on logical reasoning (Thinking) or emotional considerations (Feeling).

Model 5: Judging (J) – Perceiving (P)

The final model distinguishes between judging and perceiving personality types. It identifies whether a character’s text, generated by the fine-tuned GPT-2 model, prefers structure and decisiveness (Judging) or is more flexible and spontaneous (Perceiving).

These five models work in unison to generate, analyze, and deduce the personality traits of character's words. The models' efficiency and accuracy significantly enhance the students' learning experience, providing them with insightful feedback for their analysis and improving their understanding of MBTI personality types.

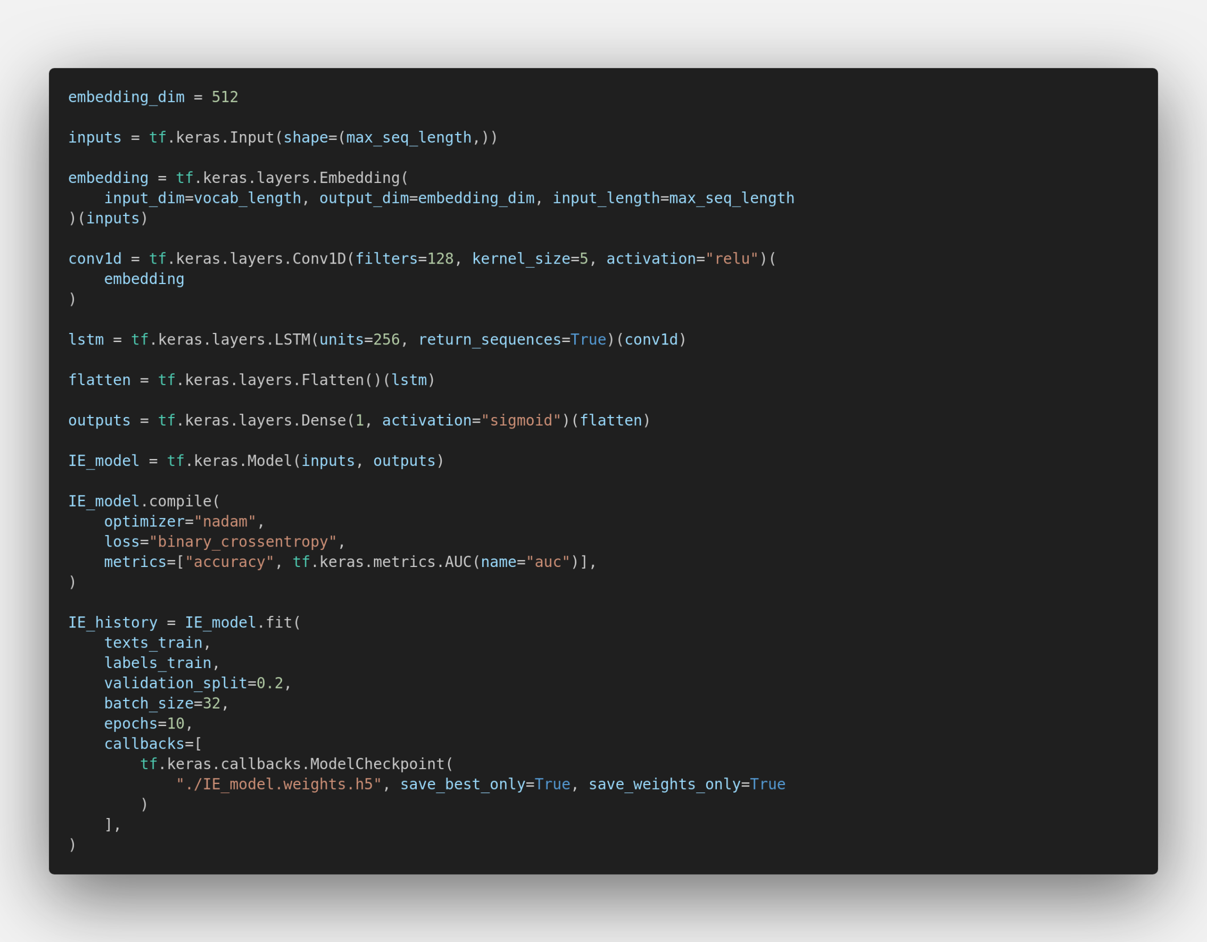
Models [2-5] Architecture

The architecture of the models used in the PersonAI project has been carefully designed to handle the complexity of extracting personality traits from text data. All the four models, while trained to identify distinct personality traits, share the same architecture.

Here's a detailed overview of the model architecture:

- **Inputs:** At the onset, the model receives tokenized text sequences as input. The length of these sequences is capped at a maximum sequence length to maintain uniformity and manage computational resources effectively.
- **Embedding:** The Embedding layer is the next step in the process. This layer transforms the input, which is in the form of tokenized text, into dense vectors of fixed size. This transformation allows the model to learn to represent words in a way that the model can understand and further process. The Embedding layer forms the basis for the model's understanding of the text data.
- **Conv1D:** Following the Embedding layer is the Conv1D layer. This layer applies convolution operations to the output from the Embedding layer. The Conv1D layer allows the model to capture local patterns in the text, which could be indicative of a specific personality trait. This layer enhances the model's ability to identify patterns that are less obvious but may be significant.
- **LSTM:** An LSTM (Long Short-Term Memory) layer is used to process the output of the Conv1D layer. This layer enables the model to understand the context and dependencies in the text over longer sequences. The LSTM layer is crucial in maintaining the continuity of information and understanding the overall narrative of the text.
- **Flatten:** After the LSTM layer, the output is then flattened. The process of flattening transforms the LSTM layer's output into a single long feature vector. This flattened vector is easier to process in the subsequent layer.
- **Outputs:** The final layer is a Dense layer with a sigmoid activation function. This layer classifies the text into one of the two personality trait categories based on the processed information from the preceding layers.

Below is the Python code that implements this model architecture as well as the hyper parameters and training callbacks:



```
embedding_dim = 512

inputs = tf.keras.Input(shape=(max_seq_length,))

embedding = tf.keras.layers.Embedding(
    input_dim=vocab_length, output_dim=embedding_dim, input_length=max_seq_length
)(inputs)

conv1d = tf.keras.layers.Conv1D(filters=128, kernel_size=5, activation="relu")(
    embedding
)

lstm = tf.keras.layers.LSTM(units=256, return_sequences=True)(conv1d)

flatten = tf.keras.layers.Flatten()(lstm)

outputs = tf.keras.layers.Dense(1, activation="sigmoid")(flatten)

IE_model = tf.keras.Model(inputs, outputs)

IE_model.compile(
    optimizer="nadam",
    loss="binary_crossentropy",
    metrics=["accuracy", tf.keras.metrics.AUC(name="auc")],
)

IE_history = IE_model.fit(
    texts_train,
    labels_train,
    validation_split=0.2,
    batch_size=32,
    epochs=10,
    callbacks=[
        tf.keras.callbacks.ModelCheckpoint(
            "./IE_model.weights.h5", save_best_only=True, save_weights_only=True
        )
    ],
)
```

Models [2-5] architecture and hyper parameters

This shared architecture, while being efficient in resource utilization, also ensures a consistent approach to the complex task of identifying specific personality traits from AI-generated paragraphs, thus aiding in the learning process of psychology students.

Experimental Results

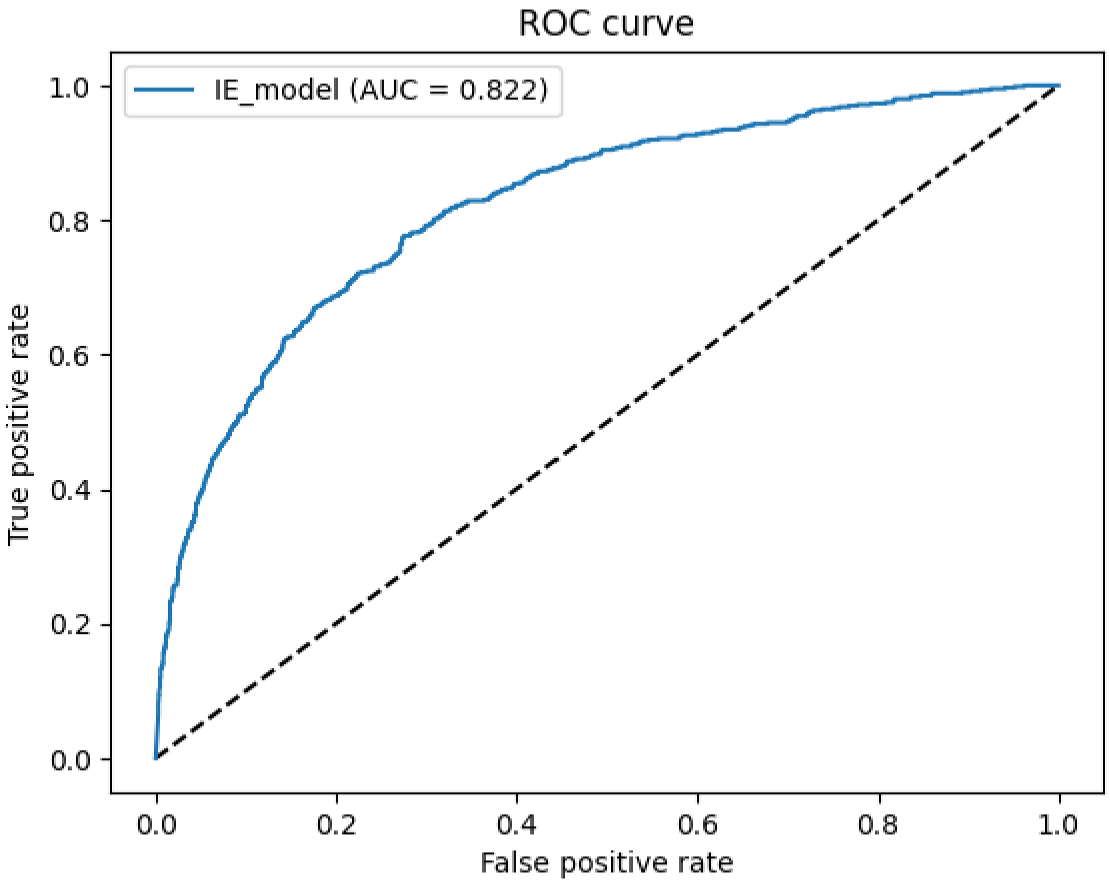
In this section, we will present the results of our experimental evaluation of the models [1-5], which were responsible for generating small paragraphs and deducing the MBTI personality traits.

Model 1: Fine-tuned GPT-2 Results

The fine-tuned GPT-2 model was trained to generate detailed and realistic small paragraphs. After training, the model successfully generated diverse paragraphs that accurately represented a broad spectrum of personality traits. The effectiveness of the generated text was evaluated based on the subsequent ability of models [2-5] to extract personality traits. The success of these models indicates the high-quality output of the fine-tuned GPT-2 model.

Model 2: Introversion (I) – Extroversion (E) Results

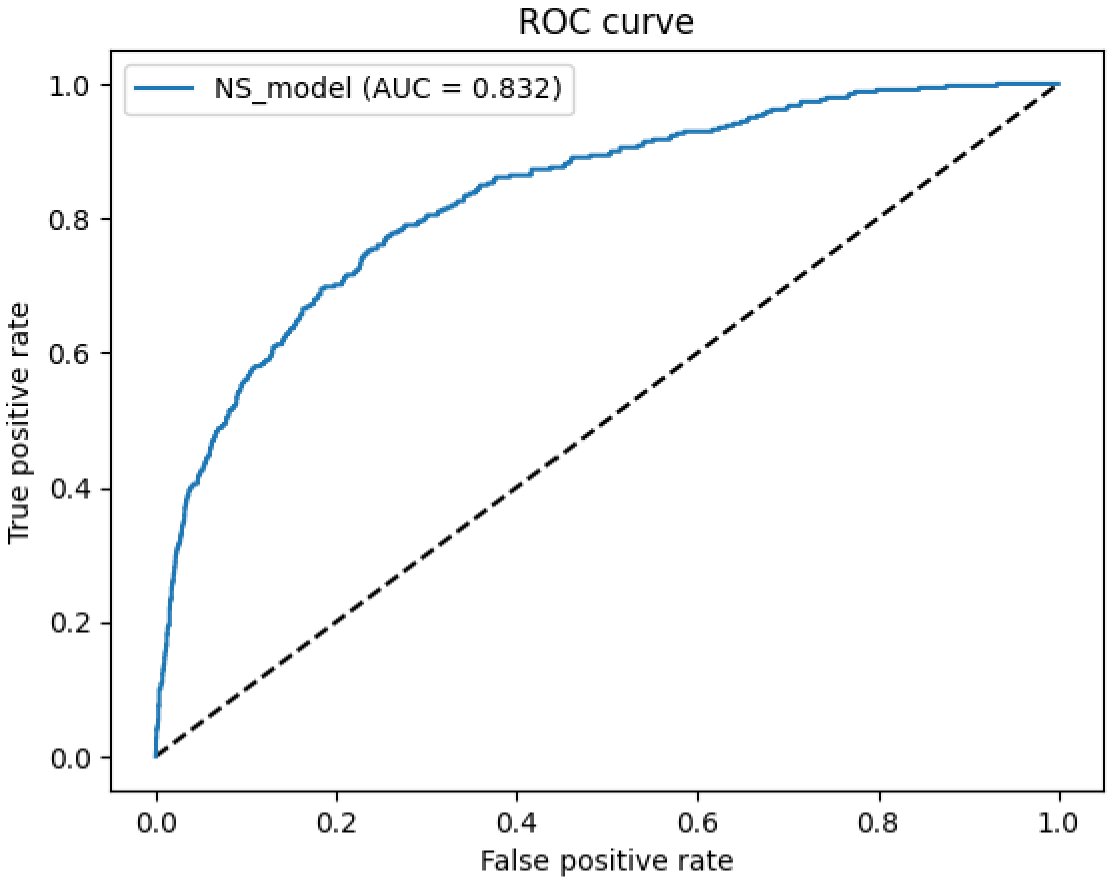
Our model, designed to distinguish between introverted and extroverted tendencies, has shown promising results with an accuracy rate of 85.98% and a loss of 0.36. This high level of accuracy underscores the model's robust performance in identifying the cues that signal either social engagement or a preference for solitude. It suggests that the model can effectively analyze the nuances in the text and make reliable predictions about the personality trait in question. Thus, this outcome provides strong evidence of our model's potential to contribute significantly to the task of personality analysis based on the Myers-Briggs Type Indicator (MBTI).



Model 2: Introversion (I) – Extroversion (E) ROC

Model 3: Intuition (N) – Sensing (S) Results

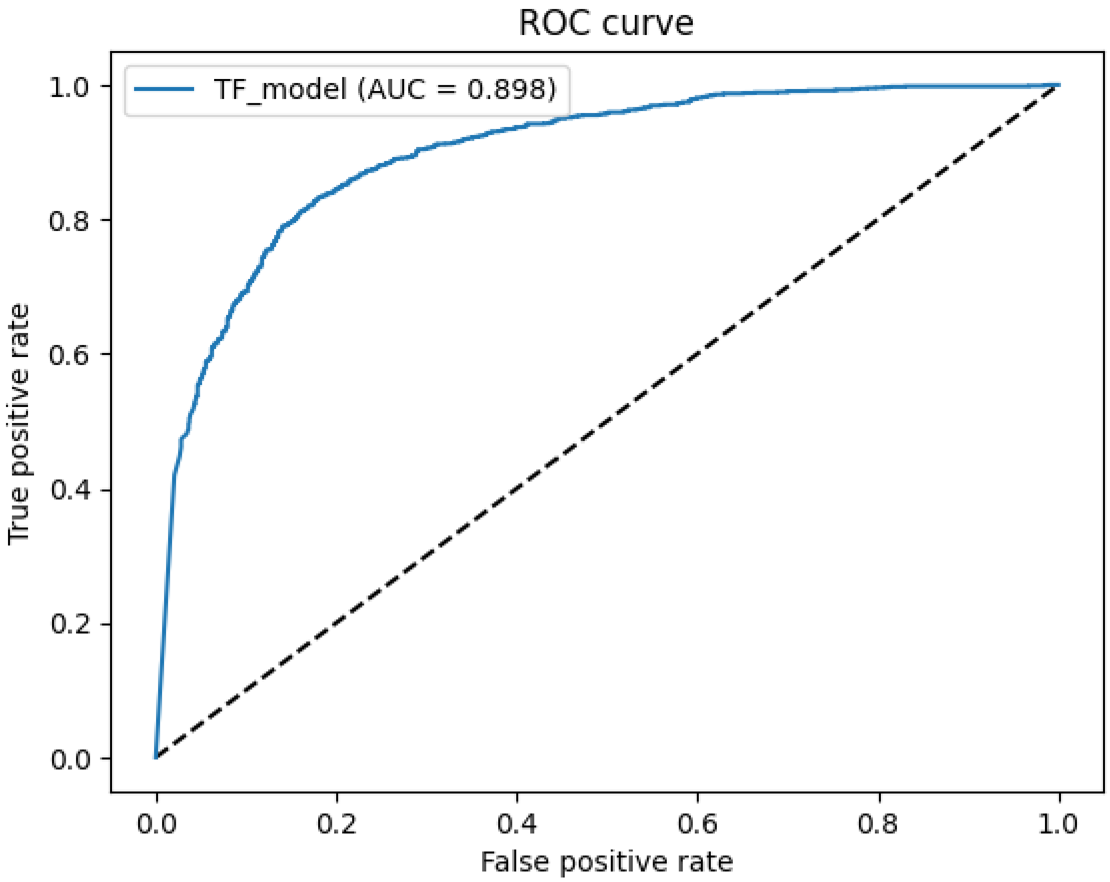
The third model in the series, which was specifically designed to identify the dichotomy between intuitive or sensing personality types, achieved an impressive accuracy of 88.94% and a relatively minimal loss of 0.31. This performance metric clearly demonstrates that the model was exceptionally capable of differentiating between characters that predominantly rely on intuition and abstract thinking as their primary decision-making tools, versus those characters who display a preference for concrete, tangible information and sensory experiences. This distinction is critical in the realm of personality typing, and the model's success in this area highlights its potential for further applications.



Model 3: Intuition (N) – Sensing (S) ROC

Model 4: Thinking (T) – Feeling (F) Results

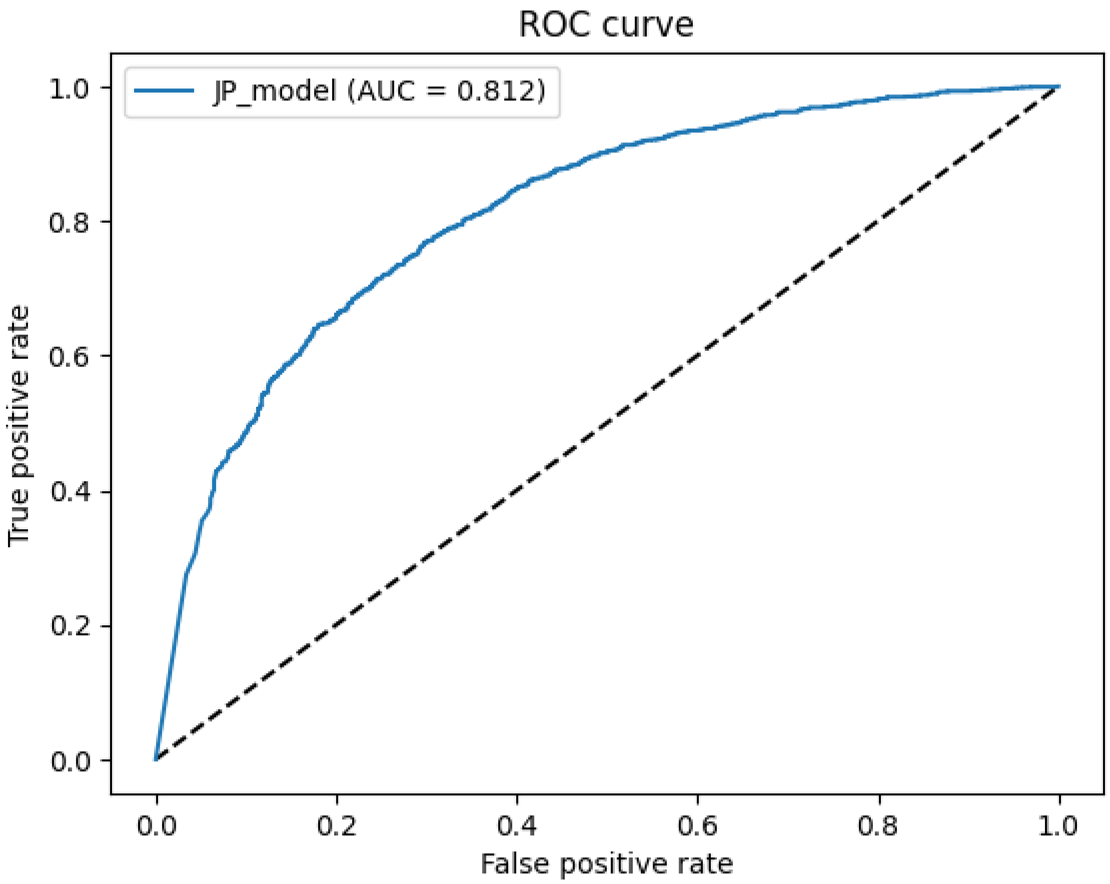
The fourth model we experimented with was particularly focused on discerning thinking and feeling types. This model achieved an accuracy rate of 83.98%, which in itself is a substantial achievement. In addition to this, the loss was recorded at 0.38. This combination of high accuracy and relatively low loss demonstrates the model's highly effective ability to make distinctions between characters based on whether they make their decisions grounded on logical reasoning or if they are more influenced by emotional considerations. This model's efficacy underscores the potential of machine learning in understanding complex human behavior and decision-making processes.



Model 4: Thinking (T) – Feeling (F) ROC

Model 5: Judging (J) – Perceiving (P) Results

The final model we developed was geared towards distinguishing between judging and perceiving types. This model was able to achieve an accuracy rate of 78.56%, which is a significant accomplishment in the field. Furthermore, the model demonstrated a loss rate of 0.49, indicating a favourable balance between bias and variance. The model was extremely effective in identifying whether a character prefers structure and decisiveness, traits commonly associated with the judging type. On the other hand, it was equally skilled in identifying characters who were more inclined towards flexibility and spontaneity, characteristics that are typically seen in perceiving types. These results underline the model's potential as an efficient tool in predicting and understanding human behavioural tendencies.



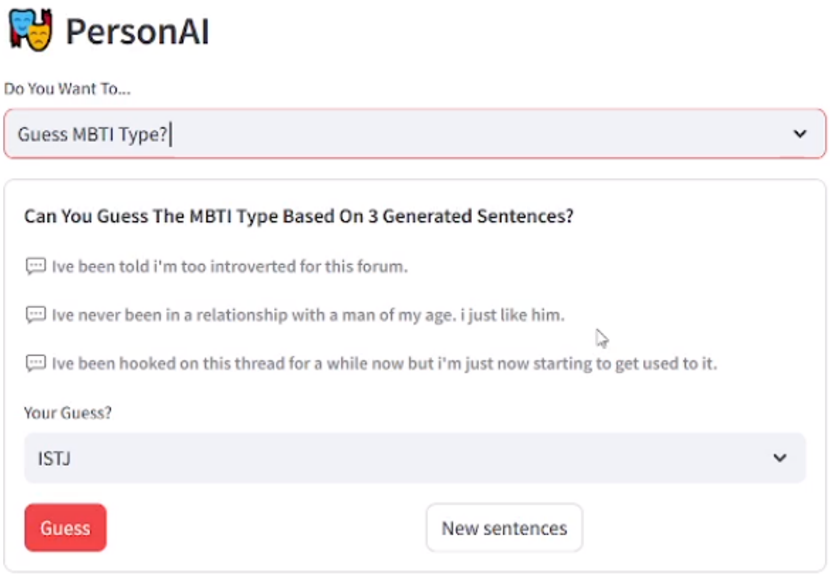
Model 5: Judging (J) – Perceiving (P) ROC

These results show that our models can successfully generate, analyze and deduce personality traits from text, enhancing the learning experience for psychology students and providing valuable insights into MBTI personality types.

Deployment

The PersonAI application was deployed using Streamlit, a Python library that simplifies the process of creating interactive web applications. With this application, users can generate text using the GPT-2 model as their puzzle. This text is then analyzed by four different models to detect personality traits and the result is compared to the user's answer; if it's different, they lose the round.

Screenshots



Do You Want To...

Guess MBTI Type?

Can You Guess The MBTI Type Based On 3 Generated Sentences?

Ive been told i'm too introverted for this forum.

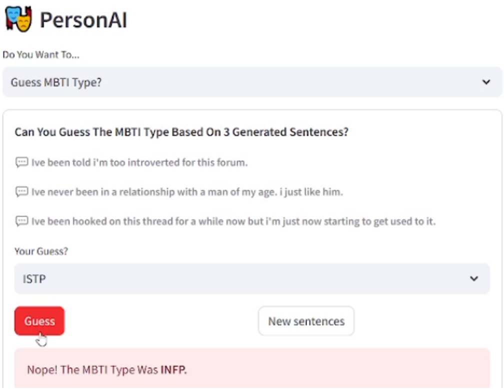
Ive never been in a relationship with a man of my age. i just like him.

Ive been hooked on this thread for a while now but i'm just now starting to get used to it.

Your Guess?

ISTJ

PersonAI website: Start



Do You Want To...

Guess MBTI Type?

Can You Guess The MBTI Type Based On 3 Generated Sentences?

Ive been told i'm too introverted for this forum.

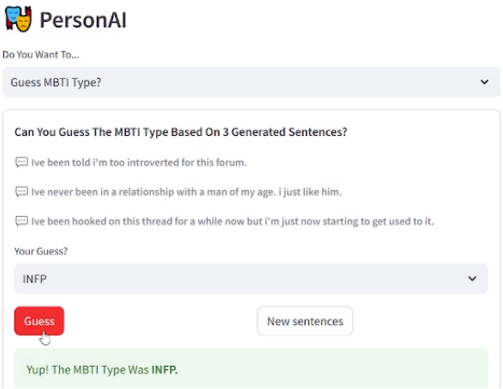
Ive never been in a relationship with a man of my age. i just like him.

Ive been hooked on this thread for a while now but i'm just now starting to get used to it.

Your Guess?

ISTP

Nope! The MBTI Type Was **INFP.**



Do You Want To...

Guess MBTI Type?

Can You Guess The MBTI Type Based On 3 Generated Sentences?

Ive been told i'm too introverted for this forum.

Ive never been in a relationship with a man of my age. i just like him.

Ive been hooked on this thread for a while now but i'm just now starting to get used to it.

Your Guess?

INFP

Yup! The MBTI Type Was **INFP.**

PersonAI website: Wrong answer

PersonAI website: Right answer

Future Developments

Looking forward to the future, our primary goal is to persistently strive towards improving our existing models in order to boost their accuracy and efficiency. This ambitious objective will require the integration of more intricate and sophisticated deep learning techniques. Such techniques are intended to enhance the models' capability to meticulously analyze and accurately deduce personality traits from raw text data.

Along with this, it is also a part of our plan to continue to fine-tune the GPT-2 model, which is known for its exceptional capabilities. Our aim here is to generate even more nuanced and diverse paragraphs that are capable of representing personality traits in a more detailed and comprehensive manner.

By implementing these changes and improvements, we believe that we can push our models' capabilities to new heights, thereby providing even more valuable insights and understanding into the complexities of human personality traits as expressed through language.

Conclusion

In conclusion, the PersonAI project represents an innovative intersection of artificial intelligence and psychology education. By developing intricate models, including fine-tuning the GPT-2 model for generating rich human-like text and creating models that analyze and deduce MBTI personality traits, we have created an engaging and practical training tool for psychology students. Our application not only facilitates a deeper understanding of the complexities of personality analysis but also offers an unprecedented platform for hands-on learning and immediate feedback. As we continue to refine our models and expand our application's capabilities, we look forward to contributing to the evolution of psychology education, making it more interactive, effective, and attuned to the digital age.

Done.

Team members

- Injy Islam ElSherbini
- Maryam Sherief Sheta
- Mariem Ahmed Abdelhaleem