**Project: LanguageLearned**

**Helping bridge the gap**

SUBMITTED BY:

Max Presley, MDS2000185, maxwell.presley@mds.ac.nz

26 03 2024

**Brief Overview /Abstract**

* **Introduction and Background**

The area my project explores is exclusivly text base. No speech recgonition is used in my project as to define a well balanced scope. The objective of the project is to help Korean speakers learn and become more familiar with the English language. It does this thruough the use of 4 main modules designed to aid in different ways. Such as conversations, quizs, translations and word analysis. This is done with 3 different NLP techniques. This project is important as there is a big language barrier between Koreans and English speaking nations. Korea culturally, technologically and socially is huge, helping to bridge this gap will enhance both Koreans and English speakers in many ways.

* **Methodology**
* **Methodology:** I will use a range of NLP techniques to help solve this problem. I will make use of large language models to help create a realistic conversation module. A translation model will be used to translate from korean to english and visaversa. The project will use a trained sentiment model and POS tagging to gain insights to words.
* **Tools/Techniques:** The code will be done exclusivly in Python. Jupyter Notebook will be the chosen IDE. Gemini/Bard will be used for the large laguage model.
* **Project architecture**
* **Results and Evaluation**

I will use a Sentiment Analysis Dataset that I have found on Kaggle. This has very good data already split into train and test. Not a lot of preprocessing will need to be done. My code once ran trains the model for later use and opens a GUI (graphical user interface) design by myself. From the GUI you can choose which module you wish to access. The performance of the model is accurate at 0.69. The advantages of using my trained model is that I can modify and change how it is working and also change/modify the dataset to get better performance. Disadvantages are that it struggles to determine the correct sentiment when the word 'not' is present like 'i am not happy' would be flagged as positive due to not understanding the corrolation of the whole sentence.

* **Discussion and Conclusions**

In conclusion, the project runs very well and can help Koreans learn English. All the modules work but the quiz module is not always accurate due to the uncontrolable nature of the Gemini model. The project can be explanded in many ways. The quiz module could be extended into a full course that uses AI to target your weak spots. Additionally, speak recognition could be a good way to expand futher learning.

* **References**

deep-translator. (2023, June 28). PyPI. <https://pypi.org/project/deep-translator/>

Gemini API: Quickstart with Python. (n.d.). Google AI for Developers. <https://ai.google.dev/tutorials/python_quickstart>

Jain, Y. (2022, March 5). POS Tagging [NLP, Python] - Yash Jain - Medium. Medium. <https://medium.com/@yashj302/pos-tagging-nlp-python-41df5243da78>

Monigatti, L. (2023, December 4). Visualizing Part-of-Speech Tags with NLTK and SpaCy. Medium. <https://towardsdatascience.com/visualizing-part-of-speech-tags-with-nltk-and-spacy-42056fcd777e>

Sentiment Analysis Dataset. (2021, January 3). Kaggle. <https://www.kaggle.com/datasets/abhi8923shriv/sentiment-analysis-dataset>

W&B. (2024, March 26). Weights & biases. W&B. <https://wandb.ai/onlineinference/Gemini/reports/The-Gemini-Pro-API-With-Python-Quickstart--Vmlldzo2MjU3OTQz>