TNSDC - GENERATIVE AI FOR ENGINEERING

FINAL PROJECT

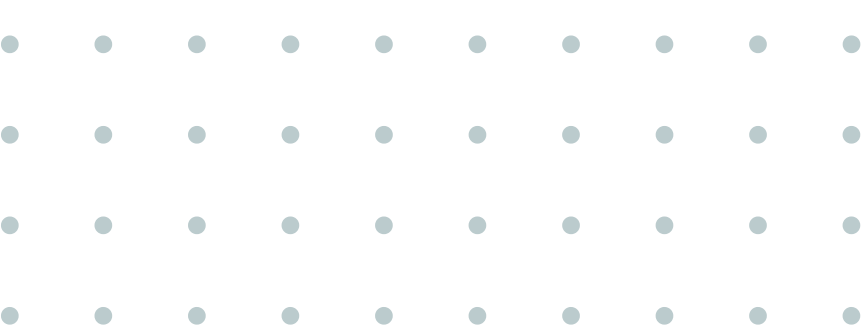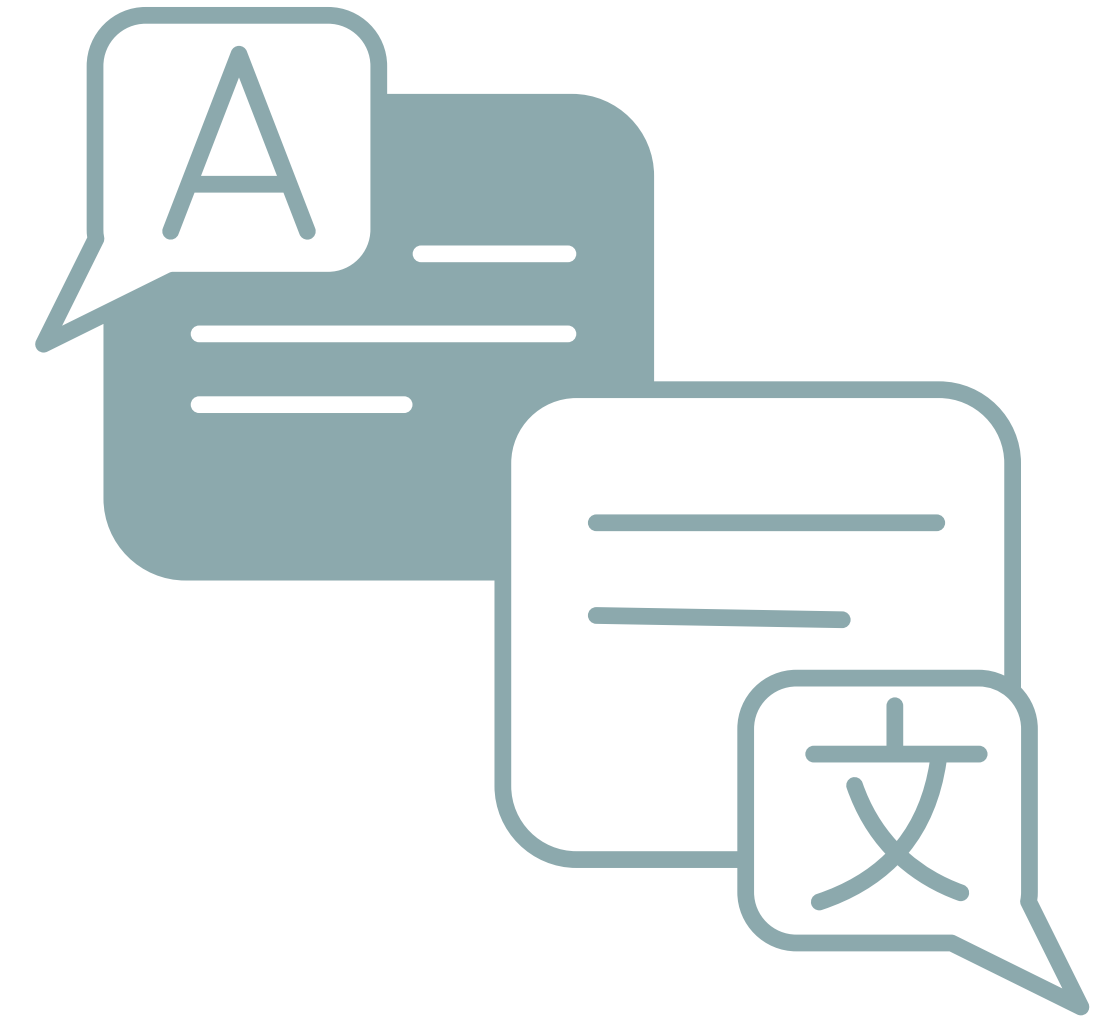Submitted by

Abishek S
(311521104003)

MULTILINGUAL TEXT TRANSLATOR

TABLE OF CONTENT

PROBLEM STATEMENT

Develop a Python program using generative AI and the Transformers library to facilitate text translation between different languages. The program should prompt the user to input the source and target languages, as well as the text to be translated. It should then utilize pre-trained models to generate accurate translations and display the results to the user. Exception handling should be implemented to ensure smooth execution even in case of errors.

PROJECT OVERVIEW

The project incorporates the MarianMTModel from the Transformers library, a state-of-the-art model designed for machine translation tasks. This model utilizes an encoder-decoder architecture with attention mechanisms to generate accurate translations between different languages.

The code follows a series of steps to facilitate text translation. It begins by prompting the user to input the source language, target language, and the text to be translated. It then tokenizes the input text and utilizes the pre-trained MarianMTModel to generate the translation. Finally, the translated text is decoded and presented to the user, ensuring a seamless translation experience.

Through this process, the program demonstrates the power of advanced generative AI in overcoming language barriers and facilitating effective cross-cultural communication. By leveraging the capabilities of the MarianMTModel, the code provides users with a reliable and efficient solution for translating text between diverse languages.
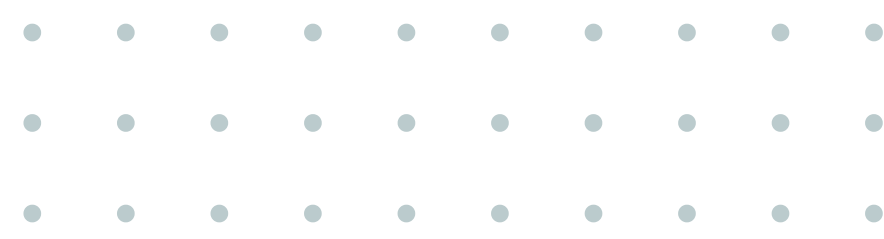
END USERS

1. **Travelers and Tourists:** Individuals traveling to foreign countries for leisure or business purposes who need to translate signs, menus, directions, and other text-based information in real-time.
2. **Language Learners:** Students and enthusiasts learning a new language who require assistance in translating texts to aid their comprehension and language acquisition process.
3. **Multinational Corporations:** Businesses operating globally that need to translate documents, emails, reports, and other communication materials to facilitate collaboration and communication among their international teams and clients.
4. **Researchers and Academics:** Professionals in the field of linguistics, anthropology, sociology, and related disciplines who utilize translation services to analyze and study texts from different languages and cultures.
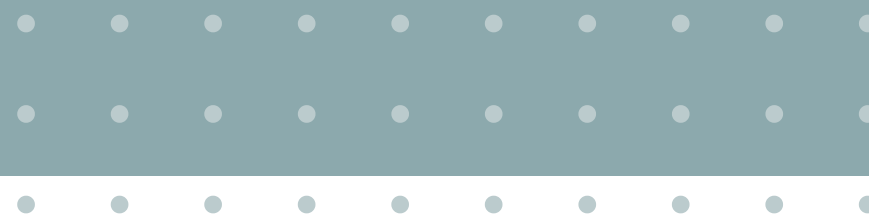
SOLUTION
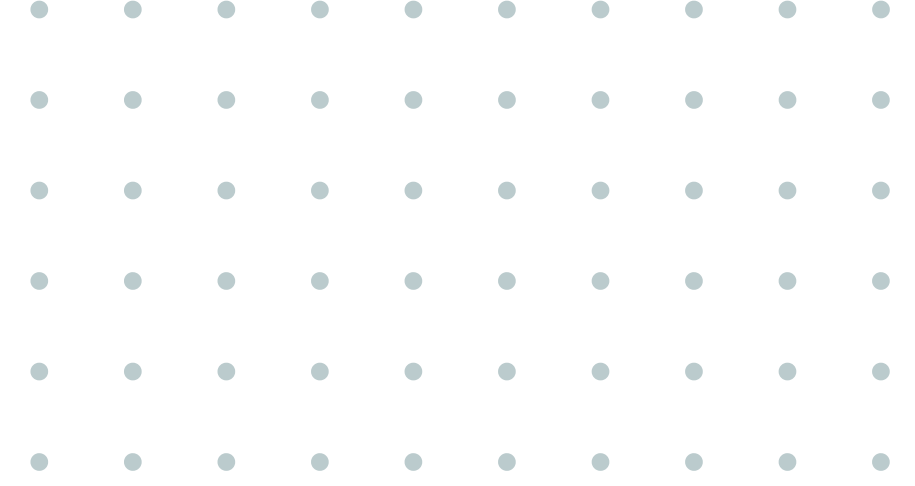
The solution provided by the text translation program addresses the problem statement by offering a user-friendly platform for translating text between different languages. Leveraging generative AI and the Transformers library, the program accurately translates input text from a source language to a target language. Users can easily interact with the program by specifying the source and target languages along with the text to be translated. The program utilizes pre-trained models to generate precise translations, ensuring effective communication across linguistic boundaries. With its intuitive interface and wide language coverage, the solution streamlines the translation process, saving time and resources while facilitating seamless cross-cultural communication.

VALUE PROPOSITION
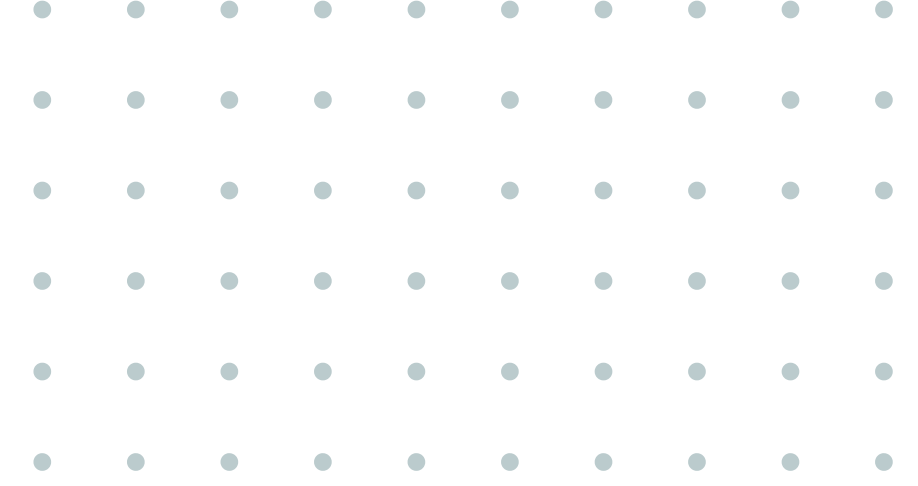
The value proposition of the text translation program lies in its ability to provide accurate and efficient language translation services, enabling seamless communication across linguistic barriers. By leveraging generative AI and the Transformers library, the program ensures precise translations, maintaining the integrity of the original message. Users benefit from a user-friendly interface that simplifies the translation process, saving time and effort. With support for a wide range of languages, the program caters to diverse linguistic needs, making it indispensable for individuals, businesses, and organizations operating in a globalized world. Overall, the program offers enhanced accessibility, efficiency, and effectiveness in cross-cultural communication, empowering users to connect and collaborate effectively regardless of language differences.

The "wow" factor in the solution is its seamless integration of cutting-edge generative AI and the Transformers library to deliver remarkably accurate translations. This advanced technology empowers users to effortlessly bridge language barriers with precision and efficiency. Additionally, the program's intuitive interface and wide language coverage ensure accessibility and usability for a diverse range of users. Whether it's translating complex business documents or navigating foreign environments during travel, the program offers a transformative solution that simplifies and enhances cross-cultural communication. Its ability to facilitate smooth, accurate, and quick translations truly sets it apart, eliciting a "wow" response from users encountering its capabilities.

MODELLING

The model used in the code is the MarianMTModel from the Transformers library. Specifically, it employs the pre-trained models available under the "Helsinki-NLP/opus-mt-{source_language}-{target_language}" format. These models are trained for machine translation tasks, allowing the program to translate text from a source language to a target language. The MarianMTModel utilizes encoder-decoder architecture and incorporates attention mechanisms for generating accurate translations. It is trained on large multilingual datasets to capture complex language patterns and nuances, enabling high-quality translation outputs.
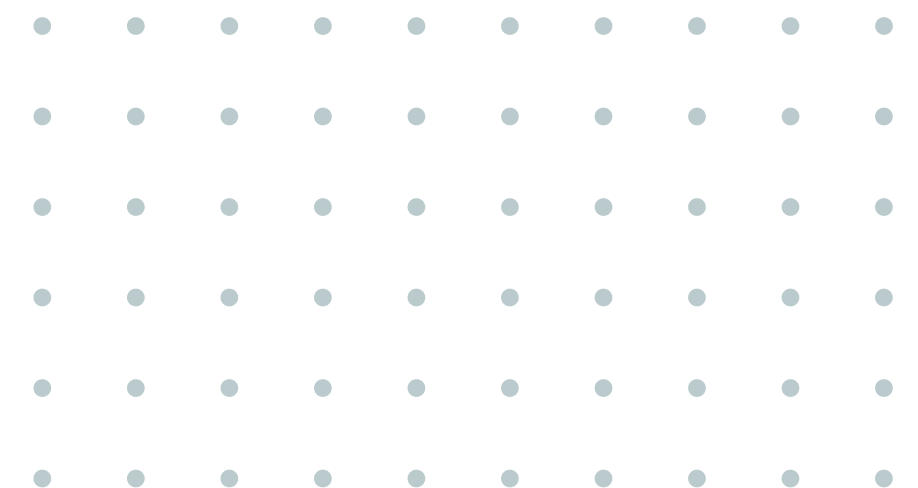
RESULTS

```
Welcome to the language translator!!!
Supported languages
Enter source language code (e.g., en for English): en
Enter target language code (e.g., fr for French): fr
Enter text to be translated: image caption generation is a challenging task

config.json: 100%           ████████████████        1.42k/1.42k [00:00<00:00, 54.7kB/s]

pytorch_model.bin: 100%     ████████████████        301M/301M [00:04<00:00, 39.4MB/s]

generation_config.json: 100% ███████████████        293/293 [00:00<00:00, 12.2kB/s]

tokenizer_config.json: 100% ████████████████        42.0/42.0 [00:00<00:00, 1.52kB/s]

source.spm: 100%            ████████████████        778k/778k [00:00<00:00, 11.1MB/s]

target.spm: 100%            ████████████████        802k/802k [00:00<00:00, 2.38MB/s]

vocab.json: 100%            ████████████████        1.34M/1.34M [00:00<00:00, 2.58MB/s]
/usr/local/lib/python3.10/dist-packages/transformers/models/marian/tokenization_marian.py:197: UserWarning: Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")

Original text (en): image caption generation is a challenging task
Translated text (fr): la génération de légendes d'image est une tâche difficile

Do you want to translate another text? (yes/no): yes
Welcome to the language translator!!!
Supported languages
Enter source language code (e.g., en for English): en
Enter target language code (e.g., fr for French): fr
Enter text to be translated: Taking my time on my ride
/usr/local/lib/python3.10/dist-packages/transformers/models/marian/tokenization_marian.py:197: UserWarning: Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
```

```
Original text (en): Taking my time on my ride
Translated text (fr): J'ai pris mon temps en voiture.

Do you want to translate another text? (yes/no): yes
Welcome to the language translator!!!
Supported languages
Enter source language code (e.g., en for English): en
Enter target language code (e.g., fr for French): fr
Enter text to be translated: Today is sunny

Original text (en): Today is sunny
Translated text (fr): Aujourd'hui est ensoleillé

Do you want to translate another text? (yes/no): yes
Welcome to the language translator!!!
Supported languages
Enter source language code (e.g., en for English): fr
Enter target language code (e.g., fr for French): en
Enter text to be translated: Aujourd'hui est ensoleillé
config.json: 100%          1.42k/1.42k [00:00<00:00, 72.0kB/s]

pytorch_model.bin: 100%          301M/301M [00:04<00:00, 41.0MB/s]

generation_config.json: 100%          293/293 [00:00<00:00, 15.2kB/s]

tokenizer_config.json: 100%          42.0/42.0 [00:00<00:00, 1.71kB/s]

source.spm: 100%          802k/802k [00:00<00:00, 13.9MB/s]

target.spm: 100%          778k/778k [00:00<00:00, 25.1MB/s]

vocab.json: 100%          1.34M/1.34M [00:00<00:00, 24.3MB/s]

Original text (fr): Aujourd'hui est ensoleillé
Translated text (en): Today is sunny
```