

APLIKASI TEXT GENERATION DARI BERBAGAI BAHASA MENGUNAKAN MODEL GENERATIVE PRE-TRAINED TRANSFORMER DAN NO LANGUAGE LEFT BEHIND

Naufal Setyawan

Program Studi Teknik Informatika, Fakultas Bisnis dan Teknologi Informasi

Universitas Teknologi Yogyakarta

Jl. Ringroad Utara Jombor Sleman Yogyakarta

E-mail : naufalsetiawan64@gmail.com

ABSTRACT

GPT models are widely used for generating text, but they currently struggle to understand languages other than English, and even if GPT models that use different languages exist, they are often too large to be easily deployed because it uses monolithic design. To address this problem, we propose a microservice design for a GPT model that uses NLLB (No Language Left Behind) for translating the prompt and generated text. This allows for better performance and understanding in languages other than English. The use of a microservice design makes it easier to deploy and scale the model. The model consists of two main components, the GPT model and the NLLB translation module. The GPT model is responsible for generating text based on the given prompt, while the NLLB module is responsible for translating the prompt and generated text to the desired language. The use of a microservice design allows for easy scaling and the ability to add more languages and resources as needed. The model is trained on a diverse dataset of various languages, which improves the performance and understanding of the generated text in multiple languages. Overall, our proposed model addresses the problem of GPT models not being able to understand languages other than English, and the difficulty of deploying larger models for other languages by incorporating a microservice design and the NLLB translation module. This allows for better performance and understanding in a variety of languages, making it more accessible to a global audience.

Keywords : NLLB, GPT, Microservice, Monolithic, Translator, Generator.