

## **RESEARCH DIFFERENCE BETWEEN GPT-2 AND GPT-3**

Пономаренко Д.Р.

Науковий керівник – проф. Смеляков К.С.

Харківський національний університет радіоелектроніки, каф. ШІ

м. Харків, Україна

тел.: +38 066 439 2797, e-mail: denys.ponomarenko2@nure.ua

The field of natural language processing has witnessed significant advancements in recent years, with the advent of cutting-edge language models like GPT-2 and GPT-3. These models, developed by OpenAI, have been revolutionary in generating human-like text and have found applications in various domains. This paper aims to provide a comprehensive comparison between GPT-2 and GPT-3, highlighting their dissimilarities in terms of architecture, model size, training data, capabilities, and limitations. The analysis reveals that while GPT-2 is a powerful language model with remarkable performance, GPT-3 takes the capabilities of text generation and manipulation to unprecedented heights.

Generative language models have garnered significant attention in recent years for their remarkable ability to generate human-like text. Among these models, the Generative Pre-trained Transformer (GPT) series developed by OpenAI has been at the forefront. GPT-2 [1], released in 2019, made headlines for its impressive text generation capabilities. However, its successor, GPT-3 [2], released in 2020, pushed the boundaries of text generation to unprecedented levels. Both models have garnered widespread interest in the research community and have been applied in various domains, including natural language processing, text generation, and text manipulation. Despite belonging to the same family of language models, GPT-2 and GPT-3 exhibit notable differences in terms of their architecture, model size, training data, capabilities, and limitations. Understanding these distinctions is crucial for researchers, practitioners, and enthusiasts to harness the unique features and potential applications of each model effectively. In this research, we aim to comprehensively compare GPT-2 and GPT-3, shedding light on their differences and providing insights into their respective strengths and limitations. By delving into the distinct characteristics of these two models, we can gain a deeper understanding of their capabilities, implications, and potential directions for future research.

The main objective of this research experiment is to investigate and analyze the differences between GPT-2 and GPT-3, two prominent generative language models developed by OpenAI. The research will focus on comparing their architecture, model size, training data, capabilities, and limitations to provide a comprehensive understanding of their distinctions[5].

The research experiment will involve the following steps:

1. Literature Review
2. Data Collection
3. Comparison and Analysis
4. Results and Discussion
5. Expected Results:

GPT-2 and GPT-3 are both language models developed by OpenAI, but they have several key differences:

**Scale and Size:** GPT-3 is significantly larger and more powerful than GPT-2 in terms of scale and size. GPT-3 has 175 billion parameters, making it one of the largest language models ever created, while GPT-2 has 1.5 billion parameters, which was considered large at the time of its release [3].

**Capabilities:** Due to its larger size, GPT-3 is capable of generating more complex and coherent text compared to GPT-2. It can generate entire articles, stories, and even computer code with minimal prompts, and can perform a wide range of natural language processing (NLP) tasks such as translation, question answering, and text summarization with high accuracy [4].

**Zero-Shot and Few-Shot Learning:** GPT-3 has the ability to perform zero-shot and few-shot learning, meaning it can generate text or perform NLP tasks with little or no training data.

Overall, GPT-3 represents a significant advancement over GPT-2 in terms of scale, capabilities, and potential applications, but it also comes with increased complexity and cost considerations.

The research experiment aims to investigate and analyze the differences between GPT-2 and GPT-3, two prominent generative language models. The results of this research will provide a comprehensive understanding of their distinctions and potential implications, contributing to the field of natural language processing and text generation.

#### References:

1. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. arXiv preprint arXiv:1901.02860.
  2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
  3. OpenAI. (2019). GPT-2: Language models are unreasonably good. Retrieved from <https://openai.com/blog/better-language-models/>
  4. OpenAI. (2020). GPT-3: Language models are few-shot learners. Retrieved from <https://openai.com/blog/gpt-3/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).