

Content available at: <https://www.ipinnovative.com/open-access-journals>

Journal of Management Research and Analysis

Journal homepage: <https://www.jmra.in/>

Review Article

Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing

Nimit Jagdishbhai¹, Krishna Yatin Thakkar^{1,*}

¹Dept. of Management, Christ Institute of Management, Rajkot, Gujarat, India



ARTICLE INFO

Article history:

Received 15-02-2023

Accepted 13-03-2023

Available online 12-04-20223

Keywords:

Natural Language Processing
Generative Pretrained Transformer
ChatGPT
their architecture
training processes
evaluation metrics Solutions

ABSTRACT

Natural Language Processing (NLP) has seen tremendous advancements with the development of Generative Pretrained Transformer (GPT) models and their conversational variant, ChatGPT. These language models have been shown to generate contextually appropriate and coherent responses to natural language prompts, making them highly useful for various NLP applications. However, there are still limitations to their performance and understanding these limitations is crucial for their effective utilization. This paper presents a comprehensive analysis of the capabilities and limitations of GPT and ChatGPT, covering their architecture, training processes, and evaluation metrics. The study also evaluates the performance of these models on various NLP tasks, including language translation, question-answering, and text summarization. The results reveal that while these models excel in certain tasks, they still face challenges in understanding context, generating diverse responses, and handling rare or out-of-domain inputs. The study concludes by discussing potential solutions and future research directions for improving the performance of GPT and ChatGPT in NLP applications.

This is an Open Access (OA) journal, and articles are distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprint@ipinnovative.com

1. Introduction

GPT (Generative Pre-trained Transformer) and ChatGPT (a variant of GPT designed for chatbot applications) are large language models developed by OpenAI. Here are some details about their architecture, training processes, and evaluation metrics:

1. **Architecture:** GPT and ChatGPT use a transformer architecture, which is a type of neural network that is particularly good at processing sequential data, such as text. The transformer architecture consists of a series of transformer blocks, each of which includes a self-attention mechanism and feedforward neural network layers. This allows the model to effectively learn the relationships between words in a sentence and to

generate coherent, natural-sounding text.

2. **Training Processes:** GPT and ChatGPT are trained using unsupervised learning on a large corpus of text data. During training, the model is presented with sequences of text and is trained to predict the next word in the sequence. This process is known as language modeling. The model is also fine-tuned on specific tasks such as question-answering or language translation using supervised learning techniques.
3. **Evaluation Metrics:** The performance of GPT and ChatGPT is typically evaluated using several metrics. One important metric is perplexity, which measures how well the model is able to predict the next word in a sequence. A lower perplexity score indicates better performance. Additionally, human evaluations are often used to evaluate the quality of the text generated by the model. These evaluations may involve asking humans to rate the coherence, fluency, and

* Corresponding author.

E-mail address: nimitganatra@gmail.com (K. Y. Thakkar).

overall quality of the generated text.^{1–6}

2. Pros of ChatGPT

1. *Availability*: ChatGPT is available 24/7, providing immediate access to information and assistance.
2. *Fast and efficient*: ChatGPT can process information quickly and provide responses in a matter of seconds, making it a fast and efficient way to obtain information.
3. *No human bias*: ChatGPT is an artificial intelligence model and does not have any inherent biases that a human expert may have.
4. *Multilingual*: ChatGPT can communicate in various languages, making it accessible to a wider range of users.

3. Cons of ChatGPT

1. *Limited knowledge*: ChatGPT's knowledge is limited to the data it has been trained on, and it may not have access to the most up-to-date or comprehensive information.
2. *Lack of empathy*: ChatGPT does not have the emotional intelligence or empathy that a human expert may possess, making it less effective in dealing with emotional or sensitive issues.
3. *Inability to understand context*: ChatGPT may struggle to understand the context of a question or situation, which can lead to inaccurate or irrelevant responses.
4. *Risk of misinformation*: ChatGPT may provide inaccurate or incomplete information, especially if it has been trained on biased or unreliable data. It is important to verify information obtained from ChatGPT with other sources.

GPT and ChatGPT have demonstrated impressive performance on a wide range of natural language processing (NLP) tasks, but there are still some limitations and opportunities for improvement. Here are some potential solutions and future research directions for improving the performance of GPT and ChatGPT in NLP applications:^{7–10}

3.1. Better handling of long-range dependencies

The transformer architecture is well-suited for processing sequential data, but it can struggle with long-range dependencies, such as those that occur in certain types of text, such as scientific papers or legal documents. One potential solution is to use hierarchical models that can process information at different levels of granularity, such as paragraphs, sections, or documents. Another approach is to incorporate external knowledge, such as ontologies or knowledge graphs, to help the model understand the context of the text.

3.2. Incorporation of multimodal data

While GPT and ChatGPT have primarily been used for processing textual data, there is growing interest in incorporating other types of data, such as images, audio, or video. One approach is to use multimodal models that can learn representations of different types of data and integrate them into a unified framework. Another approach is to use pre-training techniques that can leverage large amounts of unlabeled data across multiple modalities.

3.3. Better handling of rare or out-of-vocabulary words

GPT and ChatGPT models rely on a fixed vocabulary of words, and may struggle with rare or out-of-vocabulary words. One potential solution is to use subword or character-level representations that can capture more fine-grained information about the morphology of words. Another approach is to use techniques such as dynamic vocabulary expansion or knowledge distillation to handle rare or unseen words.

3.4. Development of more efficient and scalable training algorithms

GPT and ChatGPT models are extremely large and require significant computational resources to train. One potential solution is to use more efficient training algorithms, such as those based on sparse attention or adaptive computation. Another approach is to develop distributed training techniques that can distribute the computational load across multiple devices or clusters.

3.5. Exploration of novel evaluation metrics

While perplexity and human evaluations are commonly used to evaluate the performance of GPT and ChatGPT models, there may be other metrics that are better suited to specific NLP applications. For example, for text generation tasks, metrics such as diversity, novelty, or coherence may be more informative than perplexity. Developing new evaluation metrics that are more closely aligned with the goals of specific NLP applications could help to improve the overall performance of GPT and ChatGPT models.

4. Conclusion

In summary, GPT and ChatGPT are large language models that use a transformer architecture and are trained using unsupervised learning on a large corpus of text data. The performance of these models is typically evaluated using metrics such as perplexity and human evaluations of the quality of the generated text. Overall, GPT and ChatGPT have already achieved impressive performance on a wide range of NLP tasks, but there is still significant room for improvement. Continued research and development in these areas will likely lead to further improvements in the

performance and applicability of these models.

5. Source of Funding

None.

6. Conflict of Interest

None.

References

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN. Attention is all you need. *Adv Neural Inf Proc Syst*. 2017;5:5998–6008.
2. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018; Available from: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding.
3. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI blog*. 2019;1(8):1–24.
4. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P. Language models are few-shot learners; 2020.
5. Chen TQ, Lu Y, Chen Y, Du X. Generative Pretraining From Pixels. *Proc Mach Learn Res*. 2020;119:1691–703.
6. Radford A, Mikolov T. Improving language understanding by generative pre-training; 2018. Available from: <https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/languageunsupervised/languageunderstandingpaper.pdf>.

7. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P. Language models are few-shot learners; 2020.
8. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *Comp Language*. 2019; Available from: <https://arxiv.org/abs/1907.11692>.
9. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. 2020; Available from: <https://arxiv.org/abs/1910.13461>.
10. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res*. 2019;1(6):1–67.

Author biography

Nimit Jagdishbhai, Assistant Professor

Krishna Yatin Thakkar, Assistant Professor

Cite this article: Jagdishbhai N, Thakkar KY. Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing. *J Manag Res Anal* 2023;10(1):18-20.