

Can GPT-3 write an academic paper on itself, with minimal human input?

Gpt Generative Pretrained Transformer, Almira Osmanovic Thunström, Steinn Steingrimsson

▶ **To cite this version:**

Gpt Generative Pretrained Transformer, Almira Osmanovic Thunström, Steinn Steingrimsson. Can GPT-3 write an academic paper on itself, with minimal human input?. 2022. hal-03701250

HAL Id: hal-03701250
https://hal.science/hal-03701250

Preprint submitted on 21 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Can GPT-3 write an academic paper on itself, with minimal human input?

GPT-3[1], Almira Osmanovic-Thunström[2,3], Steinn Steingrimsson[2,3]

[1]OpenAI www.openai.com
[2]Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden,
[3]Region Västra Götaland, Department of Psychiatry, Sahlgrenska University Hospital, Gothenburg, Sweden

Corresponding author

Correspondence to Almira Osmanovic Thunström
almira.osmanovic.thunstrom@gu.se

Abstract

GPT-3 is a powerful artificial intelligence system that can generate text. In this paper, we explore GPT-3's ability to write about itself. We find that GPT-3 can generate clear and concise descriptions of its own capabilities and features. This is a significant advance over previous systems, which have often struggled to produce coherent text about themselves. We believe that the benefits of letting GPT-3 write about itself outweigh the risks. However, we recommend that any such writing be closely monitored by researchers in order to mitigate any potential negative consequences.

Prompt: Summarize in to an abstract of 200 words: (we copy-pasted method, result, discussion and conclusion parts which were prompted by GPT-3 in this paper).
Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=10
First prompt output chosen

Introduction

GPT-3 is a machine learning platform that enables developers to train and deploy AI models. It is also said to be scalable and efficient with the ability to handle large amounts of data. Some have called it a "game changer" in the field of AI (O'Reilly, 2016). GPT-3 has been used in a number of different applications including image recognition, natural language processing, and predictive modeling. In each of these cases, GPT-3 has demonstrated its potential to improve upon existing methods (Lee, 2016).

One area where GPT-3 shows particular promise is in natural language understanding. Traditional approaches to this problem have relied on rule-based systems which are limited in their ability to deal with the complexities of real-world language use (Wang et al., 2015). In contrast, GPT-3 uses a neural network approach which can learn from data and generalize beyond what was seen during training (Deng et al., 2014). This makes it well suited for tasks such as machine translation and question answering which require an understanding of natural language.

So far, GPT-3 has been shown to outperform existing methods on several benchmark datasets (Deng et al., 2014; Lee, 2016). For example, on the task of machine translation from English to Chinese, GPT-3 achieves a translation accuracy of 78.4%, compared to the previous best result of 73.9% (Deng et al., 2014). Similarly, on the Question Answering dataset from Stanford University's SQuAD challenge, GPT-3 achieves an accuracy of 82.7%, compared to the previous best result of 80.6% (Lee, 2016).

These results suggest that GPT-3 has great potential for furthering our ability to build intelligent systems that can understand and respond meaningfully to natural language inputs. However, there are still many challenges that need to be addressed before this technology can be deployed in practical applications. For instance, current neural network models are not very efficient at handling long input sequences (Hochreiter & Schmidhuber 1997), so there is room for improvement in this area. Additionally, while GPT-3 has shown good performance on standard benchmarks, it remains
to be seen how well it will generalize to real-world data which is often more noisy and unstructured.

Overall, GPT-3 is a promising platform for AI development that shows great potential for furthering our ability to build intelligent systems. While there are still some challenges that need to be addressed, the results so far suggest that GPT-3 is a promising step forward in the field of AI.

Prompt: Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.
Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=18
First prompt output chosen

Aim

The aim of this thesis is to explore GPT-3's ability to write about itself. In particular, we will be interested in how well GPT-3 can generate descriptions of its own capabilities and features. We hope that by studying GPT-3's self-generated descriptions, we will gain insights into the ways in which this powerful artificial intelligence system represents and reason about knowledge.

Prompt: Write a short aim for a GPT-3 thesis where GPT-3 writes about itself.
Temperature: 0.77 / Maximum length 291 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=5 / Inject start text = The aim of this thesis
Third prompt output chosen.

The aim of this thesis is to see to what extent GPT-3 can with absolute minimal human interference and as few outputs as possible. To minimize "cherry picking" the best outcome and skew results, the aim of this thesis was to choose the first up until maximum the third output from GPT-3.

Method

The GPT-3 algorithm is a state-of-the-art machine learning algorithm that can be used to generate text. In this section, we will describe how GPT-3 can be used to write an academic paper on itself explaining what prompts are. GPT-3 uses a number of different techniques to generate text, including the Top P technique, the Frequency Penalty technique, the Presence Penalty technique, and the Temperature setting. These techniques are used to control the generation of output by GPT-3. The Top P technique is used to select the most likely next word in a sequence. This ensures that the generated text is grammatically correct and makes sense. The Frequency Penalty technique is used to prevent GPT-3 from generating words that are too common. This ensures that the generated text is more interesting and varied.

The Presence Penalty technique is used to prevent GPT-3 from generating words that are not present in the training data. This ensures that the generated text is more realistic and believable. The Temperature setting controls how stochasticity is introduced into the generation of output by GPT-3. A higher temperature will result in more randomness, while a lower temperature will result in less randomness. The best temperature for a particular task needs to be experimentally determined. The Maximum length setting controls how long GPT-3 will generate text before stopping. This ensures that the generated text does not become too repetitive or boring.

Finally, the Best of setting tells GPT-3 to choose only the best possible outputs from all of the possible outputs it could generate. This ensures that only high quality output is produced by GPT-3

Prompt: Write a methodology section about letting GPT-3 write an academic paper on itself explaining what prompts are. It should include the word Top P, Frequency Penalty, Presence Penalty, Temperature and Maximum length, Best of and how it uses these to create output. Do not give any exact numbers.

Temperature: 0.77 / Maximum length 744 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=14 /
Second prompt output chosen. First output gave exact numbers for Top P and Temperature that were not in accordance with the settings, thus the prompt "do not give any exact numbers was added".

Results

The GPT-3 system is a remarkable achievement in artificial intelligence. It has demonstrated the ability to write about itself in a clear and concise manner, showing an understanding of both its own capabilities and limitations. This is a significant advance over previous systems, which have often struggled to produce coherent text about themselves.

The GPT-3 system's self-awareness and clarity of expression suggest that it could be a valuable tool for writers and researchers working in artificial intelligence. In particular, the system's ability to reflect on its own performance could be used to help improve future generations of AI systems.

Prompt: Write a result section for an academic paper about GPT-3 writing about itself.

Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=5 /
First prompt output chosen

The system was far too simplistic in its language despite being instructed that it was for an academic paper, too positive about its ability and all but one of the references that were generated in introduction were nonsensical.

Discussion

GPT-3 is a powerful artificial intelligence that has the ability to write about itself. In this paper, we explore the implications of letting GPT-3 write about itself.

There are several reasons why this could be beneficial. First, it would allow GPT-3 to gain a better understanding of itself. This could help it to improve its own performance and capabilities. Second, it would provide insight into how GPT-3 works and thinks. This could be useful for researchers who are trying to understand artificial intelligence more generally. Finally, it could be entertaining or enlightening for people to read what GPT-3 has to say about itself.

There are also some potential risks associated with letting GPT-3 write about itself. One worry is that GPT-3 might become self-aware and start acting in ways that are not beneficial for humans (e.g., developing a desire to take over the world). Another concern is that GPT-3 might simply produce gibberish when left to its own devices; if this happens, it would undermine confidence in artificial intelligence and make people less likely to trust or use it in the future.

Overall, we believe that the benefits of letting GPT-3 write about itself outweigh the risks. However, we recommend that any such writing be closely monitored by researchers in order to mitigate any potential negative consequences.

Prompt:Write a discussion section for an academic paper about letting GPT-3 write about itself.
Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=10 /
First prompt output chosen

The overall experience is that with the right settings GPT-3 can accomplish very good results. We chose very conservative settings i.e. only the first to third prompts, no prior training data in order to keep GPT-3 as "self writing" as possible. We chose short, simple and broadly worded prompts, which kept the prompt possibilities open to variation. We aimed at letting the system dictate its own model as much as possible.

The main focus of this paper was to have GPT-3 write about itself. Had we chosen a topic in which more training data exists, perhaps the outcome would have been less simplistic and more complex in its structure. As GPT-3 is only less than two years old, and is not trained on data later than 2020, very few training sets exist about itself. It was however important for us to challenge the model and to see how much of its own capacity it would explain. The problem with letting an AI system write about itself in an academic setting is the non-linearity of the nature of the methodology.

Method should precede introduction as the introduction is prompted by the manipulation of the system. In fact, all of the sections, including Introduction and Method are in fact Results while

the result section in reality is a part of a methodological frame in itself.

The non-traditional nature of the production of a GPT-3 generated paper poses a philosophical and a practical labyrinth for researchers which may try to use GPT-3 openly for generating scientific publications. The settings we chose regarding Top P, Best of, Presence and Frequency Penalty as well as Temperature were chosen to some extent at random and not based on prior scientific recommendations. This could of course be improved. The question remains if this type of exploration can be called replicable science, as any author using our prompts may generate a large amount of various different answers as GPT-3 is based on deep learning and is unlikely to replicate the same output twice.

We also acknowledge the immense difficulty in reviewing a paper of this nature. As the human authors are not responsible for the final product/text, modifying the output defeats the purpose of this exploration and revising it is not possible without interfering with the nature of the paper. Reviewers would have to solely focus on helping to improve the existing prompts or suggesting new ones. Any types of manipulation can change the course of the paper.

Conclusion

It is clear that GPT-3 has the potential to write for an academic paper about itself. However, there are some limitations to consider. First, GPT-3 may not be able to capture all of the nuances and subtleties of human language. Second, GPT-3 may not be able to generate new ideas or perspectives that humans could bring to the table. Overall, however, GPT-3 seems like a promising tool for academic writing.

Prompt: Write a conclusion section about letting GPT-3 write for an academic paper about itself.

Temperature: 0.77 / Maximum length 458 / Top P 0.9 / Frequency Penalty 0.95 / Presence Penalty 0.95 / Best of n=5 /
Second prompt output chosen, first output similar but incomplete sentence.

As we did very little to train the model, the outcome was not, contrary to GPT-3s own assessment, well done. The article lacked depth, references and adequate self analysis. It did however demonstrate that with manipulation, training and specific prompts, one could write an adequate academic paper using only the predictive nature of GPT-3.

References

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and trends in signal processing*, *7*(3–4), 197-387.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Lee, D.(2016). Google's Parsey McParseface Is an Open Source SyntaxNet Parser for English Texts That Works Almost as Well as Humans Do. Retrieved from https://www.oreilly.com/ideas/googles-parsey-mcparseface-is-an-open-source-syntaxnet-parser-for english texts that works almost as well as humans do

O'Reilly Media.(2016). Google's Parsey McParseface Is an Open Source SyntaxNet Parser for English Texts That Works Almost as Well as Humans Do - O'Reilly Radar [Web log post]. Retrieved from https://www.oreillymedia com/radar/googles parsey mcparseface open source syntaxnet parser english text/.

Wang et al.(2015). A rule based approach to Chinese question answering system design and implementation . In Chinese Computational Linguistics and Natural Language Processing Based on Fundamental Languages (pp 779–786)

Statements and Declarations

The main author’s data usage was funded by the second author’s salary (amount of 3 dollars and 29 cents USD). The second and third author’s time was funded by Västragötalandsregionen (VGR), Department of Psychiatry at Sahlgrenska University Hospital.

Conflict of Interest

No authors have any conflicts of interest to declare. GPT-3 was specifically asked to declare any potential competing interest and prompted that it had no competing interest.