Prompt Engineering Tips for Generating Text Related to Cognitive Science and Philosophy of Mind

Simon Hua

IPHS300 AI for the Humanities (Spring 2022) Prof Elkins and Chun, Kenyon College

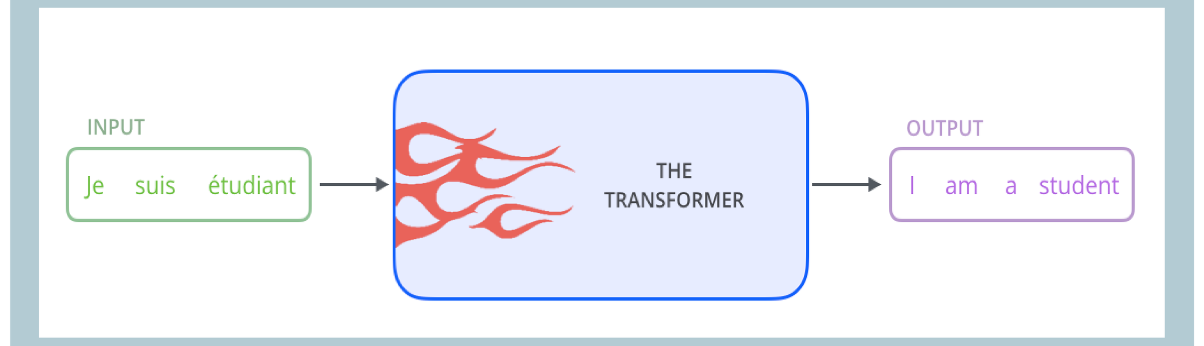
Introduction

GPT-3 is a powerful language model with 175 billion parameters developed recently by OpenAI. It is powered by artificial neural networks trained on text to predict what the next word in a sequence is likely to be. Study of Text prompt engineering for pre-trained Natural Language Processing (NLP) system is a newly emerging field, where the immense potential of what models like the GTP-3 can generate is still being revealed by researchers.
In this project, I explored how to improve GPT-3's text generation performance on academic topics (especially ones related to cognitive science/philosophy of mind) using a trial & error approach.

Methodology

- While understanding of GTP-3's implementation-level mechanism is helpful, the stochastic nature of generated texts makes a trial and error approach an effective way to explore text prompt engineering, whether for generating academic writings or other writing purposes.

- The transformer architecture used in GTP-3 is essentially a black box, where the mechanism of how an input produces corresponding output cannot be known deterministically. This means that it is extremely difficult to find hard and fast rules that always result in outputs with predictable and desirable features.



- As an alternative, we need a heuristic approach (operated via trial & error) to distinguish what kinds of text construction makes better prompts from what do not. This is by no means an ideal approach, but it is a good baseline strategy to start exploring text prompt engineering.

- In the following section, I will present my results for improving academic writings (with regard to cognitive science/philosophy of mind) generated by GTP-3. Among these four generated results, the text prompt for the 1st (original) version is written by a human with domain expertise, while each of the later versions is improved upon its predecessor.

- The text results are exhibited in different colors, while each color corresponds to specific functions:

1. Black: The original prompt written by a human (i.e. myself).
2. Orange: problems in prompt that may cause confusion for GTP-3.
3. Blue: text generated by GTP-3.
4. Underlined Blue: Unsuccessful or ambiguous text generated by GTP-3.
5. Green: Text generated by GTP-3, which are then fed back to it as part of the prompt for the next task.
6. Purple: Text insertion edited by a human (i.e. myself).

- The performance of each generation is measured by three quantitative metrics in a scale from 1 to 10: Factual Accuracy, Conceptual Significance, and Logical Coherency.

Results

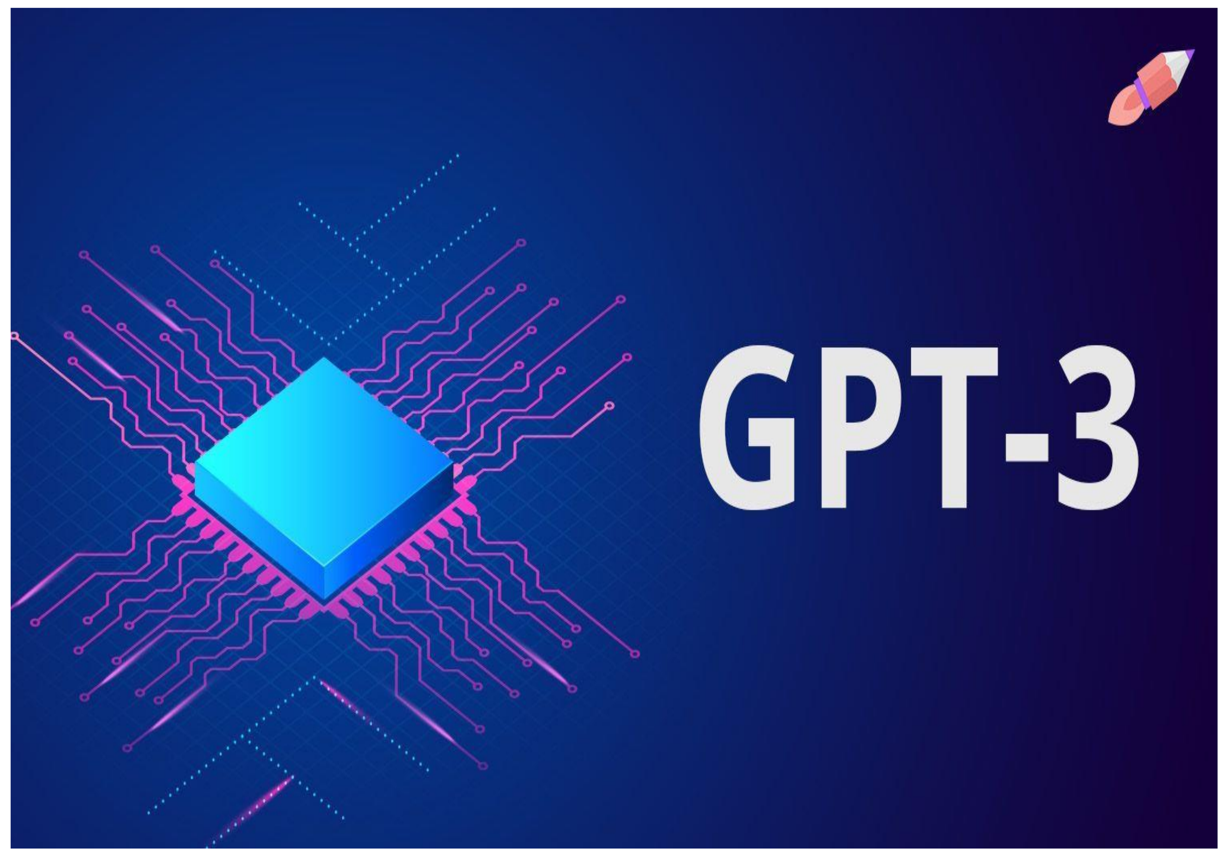
1st Version (the original version)
The Establishment view in cognitive science holds that perception is a type of cognition operating via inferences. Gibson, according to Fodor, advocates for a drastically different position: he sees perception as the "direct pickup" of "invariant properties", while the notion of 'direct' implies being non-inferentially mediated. This revolutionary project of redefining the meaning of perception deeply concerns Fodor - and other RTM cognitive scientists - for a couple of reasons:
•Firstly, it questions the adequacy of the notion of perception that has been dominant for the past few decades, and secondly, it questions the adequacy of the related notion of 'cognition'. Fodor, according to his own admission, is a Gibsonian. He has accepted Gibson's challenge, and has taken on the task of showing how perception may be non-inferentially mediated. The first step of this task is to develop an account of the nature of perceptual invariants. Fodor offers such an account in Chapter 2, and he proposes that the notion of perceptual invariants captures the central idea of Gibson's work.
Factual Accuracy: 2 Conceptual Significance: 4
Logical Coherency: 7

2nd Version
The Establishment view in cognitive science holds that perception is a type of cognition operating via inferences. Gibson advocates for a drastically different position: he sees perception as the "direct pickup" of "invariant properties", while the notion of 'direct' implies being non-inferentially mediated. This revolutionary project of redefining the meaning of perception deeply concerns Fodor - and other Establishment cognitive scientists - for a couple of reasons:
•The first reason is that this redefinition might imply a repudiation of decades of research in cognitive science. In their view, if perception were truly non-inferential, this would mean that the Establishment's entire understanding of perception is based on a false premise. The second reason is that they think this redefinition risks making Gibson's theory of perception untestable and therefore not scientific. At the heart of the debate between Gibson and the Establishment is the question of what it means for perception to be non-inferential. Fodor and his colleagues maintain that if perception is truly non-inferential, then it must be based on some kind of direct, unmediated access to reality. Gibson, on the other hand, argues that perception can be non-inferential even if it is not based on direct access to reality.
Factual Accuracy: 5 Conceptual Significance: 5
Logical Coherency: 6

3rd Version
The Establishment view in cognitive science holds that perception is a type of cognition operating via inferences. Gibson advocates for a drastically different position: he sees perception as the "direct pickup" of "invariant properties", while the notion of 'direct' implies being non-inferentially mediated. This revolutionary project of redefining the meaning of perception deeply concerns Fodor - and other Establishment cognitive scientists - for a couple of reasons: The first reason is that this redefinition might imply a repudiation of decades of research in cognitive science. In their view, if perception were truly non-inferential, this would mean that the Establishment's entire understanding of perception is based on a false premise. The second reason is that they think this redefinition risks making Gibson's theory of perception untestable and therefore not scientific.
•Fodor's response to the first concern is that, even if perception were non-inferential, this would not necessarily mean that the research of the Establishment is entirely wrong. He argues that the distinction between construal and obtaining information non-inferentially is a distinction that makes a difference at the level of description, but not at the level of explanation. In other words, even if perception were non-inferential, the fact that the Establishment has been working with an inferential model of perception does not mean that their research is entirely incorrect or that their explanations are unsound.
Factual Accuracy: 8 Conceptual Significance: 8
Logical Coherency: 7

Results (continue)

4th Version (the final version)
The Establishment view in cognitive science holds that perception is a type of cognition operating via inferences. Gibson advocates for a drastically different position: he sees perception as the "direct pickup" of "invariant properties", while the notion of 'direct' implies being non-inferentially mediated. This revolutionary project of redefining the meaning of perception deeply concerns Fodor - and other Establishment cognitive scientists - for a couple of reasons: The first reason is that this redefinition might imply a repudiation of decades of research in cognitive science. In their view, if perception were truly non-inferential, this would mean that the Establishment's entire understanding of perception is based on a false premise. The second reason is that they think this redefinition risks making Gibson's theory of perception untestable and therefore not scientific. Fodor's response to the first concern is that, even if perception were non-inferential, this would not necessarily mean that the research of the Establishment is entirely wrong. He argues that the distinction between construal and obtaining information non-inferentially is a distinction that makes a difference at the level of description, but not at the level of explanation.
•In other words, even if perception were non-inferential, cognitive scientists could still explain how we arrive at our perceptual experience via inferential processes. However, Fodor concedes that this would require a significant change in the way that cognitive scientists think about perception. With regards to the second concern, Fodor argues that Gibson's theory of perception is in fact testable. He points to research on **change blindness as an example** of how Gibson's theory can be tested empirically. Fodor concludes by saying that, even if the Establishment's view of perception is correct, there is still value in pursuing Gibson's non-inferential approach as it provides a complementary perspective that can shed new light on old problems.
•In the change blindness example, subjects were shown either a briefly flashed image or a slowly moving image and then asked to identify a change that had been made to the image. The results showed that subjects were more likely to identify the change when it was presented in the slowly moving image. Fodor argues that this shows that Gibson's theory of perception is testable, as it predicts that we are more likely to pick up on changes when they are presented in a way that is more consistent with the way the world actually changes.
Factual Accuracy: 7 Conceptual Significance: 9
Logical Coherency: 7



Conclusion & Recommendations

5 recommendations for GTP-3 prompt engineering for academic writings:

1st Step: Set the Temperature High (0.92 - 0.95)
This is higher than temperature recommendations seen for most GPT-3 generated text (0.7-0.9), and there is a reason for that. Indeed, one of the most impressive features of GTP-3 is its ability to generate logically coherent text with a high degree of randomness. By setting temperature high, GTP-3 can not only generate coherent academic writings, but with higher creativity and conceptual significance.

2nd Step: Clean Your Text Prompt
Try not to use sentence structures that cause confusion of perspectives, since one of the most common mistakes GTP-3 makes is to mistake one's perspective for someone else's. The text in orange from the original prompt is an example of what text prompt engineers for academic writings should avoid: (1) Do not proliferate the number of perspectives in a prompt unless necessary, and (2) Replace synonyms for a previously used word in a prompt with the same word that has been used (i.e. replacing "RTM" with "Establishment").

3rd Step: Text Reduction
Upon the completion of the first two steps, now we move on to the process of identifying successful/unsuccessful content generated by GTP-3. Now – at least for academic writings – most of this work requires humans in the loop with the assistance of our domain expertise. But the amount of work for us is light: Just (a) feed in your cleaned version of the original prompt to generate new text, (b) identify unsuccessful or ambiguous content in the newly generated text and delete it, and finally (c) feed your cleaned original prompt along with generated response (minus unsuccessful content) again into GPT-3.

4th Step: Text Insertion
To improve performance, we can also insert additional text prompts to let GTP-3 expand on the what it has generated. This is a crucial feature – especially for academic writings – as the succeeding text generated is often more inspiring and sometimes involve examples (like the "change blindness" example in the final version shown in bold). It is important to notice that the placement of additional prompts matters! If you start a new line then insert the additional prompt, GTP-3 will do a better and more elaborate job expanding on what you insert than if you just concatenate additional prompts onto the end of generated text.

5th Step: Repetition Between Step 3 & Step 4
By going back and forth between these two steps, you can better control the general direction of generated text, while improving its depth, clarity, and conceptual significance.

Acknowledgements

- The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time. (jalammar.github.io)
- How Crowdbotics is Using GPT-3
- Can GPT-3 Pass a Writer's Turing Test? | Published in Journal of Cultural Analytics
- https://dailynous.com/2020/07/30/philosophers-gpt-3/#askell