Task1:

Based on the outputs I got, let's analyze the generated text from each GPT-2 model and choose the most interesting one to report in the write-up.

DistilGPT-2:

Generated Text: "I am going to be a one-woman league, right? Maybe I will. (laughs)"

Comment: The generated text seems to be playful and implies that the speaker might excel in a particular domain. The addition of laughter adds a touch of humour to the text.

GPT-2:

Generated Text: "I am going to be a doctor," says her father. "And you have nothing to do with it."

Comment: This generated text showcases a conversation where the speaker's father expresses their career aspirations. The statement made by the father creates a sense of determination and independence.

GPT-2 Large:

Generated Text: "I am going to be a president very different from the alternative Donald Trump is offering!"

Comment: This generated text reflects a strong political ambition, emphasizing a desire to bring about change and be distinct from the previous political leader.

In this as we know that Donald trump is not the president now and the gpt model is not trained on today’s knowledge.

GPT-2 Medium:

Generated Text: "I am going to be a little over optimistic for a minute as to how serious the allegations could be."

Comment: This generated text displays a cautious optimism in response to some allegations, suggesting that the speaker might be considering the gravity of the situation while maintaining a positive outlook.

Based on these analyses, the generated text from the GPT-2 Large model stands out as the most interesting. The ambition to become a different kind of president and the comparison to a previous political figure sparks curiosity and opens up possibilities for further exploration.

Task2:

Based on the outputs, let's analyze the generated text from each GPT-2 model and search/sampling method:

For the GPT2LMHeadModel:

Greedy Search:

Generated Text: "I am not feeling well and I am going to be fine. I am not feeling well and I am going to be fine. I am not feeling"

Comment: The generated text seems to repeat the same phrase without providing further context or variation.

Beam Search:

Generated Text: "I am not feeling well and I am going to be fine. I am not feeling well and I am going to be fine. I am not feeling"

Comment: Similar to the greedy search, the generated text repeats the same phrase without much variation.

Top-k Sampling:

Generated Text: "I am not feeling well and I am going to be fine. I am not feeling well and I am going to be fine. I am not feeling"

Comment: Once again, the generated text repeats the same phrase without introducing any new information or creative elements.

Top-p Sampling:

Generated Text: "I am not feeling well and I am going to be fine. I am not feeling well and I am going to be fine. I am not feeling"

Comment: Similar to the previous methods, the generated text lacks variety and repeats the same phrase.

For the AutoModelForCausalLM (GPT-2):

Greedy Search:

Generated Text: "I am not feeling well and I am going to go to bed. I am going to go to bed. I am going to go to bed."

Comment: The generated text provides a clear response to the initial statement by expressing a plan to go to bed.

Beam Search:

Generated Text: "I am not feeling well and I am going to go to bed. I am going to go to bed. I am going to go to bed."

Comment: Similar to the previous method, the generated text repeats the same phrase without much variation.

Top-k Sampling:

Generated Text: "I am not feeling well and I am going to go to bed. I am going to go to bed. I am going to go to bed."

Comment: Once again, the generated text lacks creativity and repeats the same phrase without adding any new information.

Top-p Sampling:

Generated Text: "I am not feeling well and I am going to go to bed. I am going to go to bed. I am going to go to bed."

Comment: Similar to the previous methods, the generated text lacks diversity and repeats the same phrase.

Based on these analyses, the generated text from the GPT-2 model using the greedy search stands out as the most coherent and meaningful response. However, overall, the generated text from both models and search/sampling methods lacks variety and tends to repeat the same phrases, indicating a limitation in generating diverse and engaging text.

Task3:

Model 1:

Architecture: GPT-2 (base model)

Pretrained model used: "gpt2"

Tokenizer used: GPT2Tokenizer from the "gpt2" model

Parameters: Smaller model size compared to Model 2

Model 2:

Architecture: GPT-2 (large model)

Pretrained model used: "gpt2-large"

Tokenizer used: GPT2Tokenizer from the "gpt2-large" model

Parameters: Larger model size compared to Model 1

These models are part of the GPT-2 architecture, where Model 2 (gpt2-large) is a larger and more powerful variant compared to Model 1 (gpt2). The perplexity values calculated indicate the performance of these models on the given test text from "War and Peace."

Here's an analysis of the output:

Perplexity for model 1: 1.786506175994873

Perplexity for model 2: 1.6305392980575562

Perplexity is a metric commonly used to evaluate the performance of language models. It measures how well a language model predicts a given sequence of words. Lower perplexity values indicate that the model is better at predicting the next word in a sequence.

In this case, both models have relatively low perplexity values, which suggests that they perform well in predicting the next word in the given test text from "War and Peace." However, it's important to note that the perplexity values may vary depending on the specific dataset and language used for evaluation.

Comparing the perplexity values of the two models, we can see that model 2 has a slightly lower perplexity (1.6305392980575562) compared to model 1 (1.786506175994873). This indicates that model 2 performs slightly better in terms of predicting the next word in the test text.

It's worth noting that perplexity alone may not provide a comprehensive evaluation of a language model's performance. Other factors, such as the context and quality of generated text, should also be considered for a more complete assessment.

Overall, based on the perplexity values, it can be inferred that both models perform well in predicting the next word in the test text, with model 2 demonstrating a slightly better performance compared to model 1.

General Reflections on the Assignment:

This assignment provided hands-on experience with text generation using language models and comparing different models based on perplexity scores. It allowed us to observe the varying quality of generated texts and understand the influence of model size on language modeling tasks.

What I Learned from this Assignment:

Through this assignment, I learned how to use the Transformers library to generate text with GPT-2 models and compare their performance using perplexity scores. I gained insights into the capabilities and limitations of language models in generating coherent and relevant text.

Difficulty Level of the Assignment:

Overall, the assignment had a moderate difficulty level. The provided code snippets and documentation were helpful in understanding the necessary steps. However, handling the intricacies of text generation and model comparison required some familiarity with the Transformers library and language modeling concepts.

Approach for Future Assignments:

In future assignments or similar tasks, I would approach them with a clearer understanding of the data and model requirements. I would carefully review the provided code and documentation to ensure all necessary dependencies are installed and imported correctly. Additionally, I would focus on thorough analysis and documentation of the results to provide a comprehensive evaluation of the models' performance.

Final Thoughts:

This assignment was valuable in expanding my knowledge of text generation and model comparison using language models. It highlighted the creative potential and challenges associated with generating coherent text. Furthermore, it underscored the importance of evaluating models using appropriate metrics like perplexity to assess their language modeling capabilities. Overall, the assignment provided a practical and insightful learning experience.