In [3]:

```python
from transformers import GPT2Tokenizer,GPT2LMHeadModel
model_name="gpt2"
tokenizer=GPT2Tokenizer.from_pretrained(model_name)
model=GPT2LMHeadModel.from_pretrained(model_name)
#When you call the encode method, it returns a dictionary containing the tokenized input IDs, attention masks, and other information.
#By default, the encode method returns a Python dictionary with NumPy arrays as values.
#However, when you specify return_tensors='pt', the encode method returns a PyTorch tensor instead of a NumPy array.
#This is useful when you want to use the tokenized data as input to a PyTorch model.
```

```
model.safetensors:   0%|          | 0.00/548M [00:00<?, ?B/s]
generation_config.json:   0%|          | 0.00/124 [00:00<?, ?B/s]
```

In [31]:

```python
import torch
def autoregressive_generation(prompt, max_length=200):
    input_ids = tokenizer.encode(prompt, return_tensors="pt")

    with torch.no_grad():
        output = model.generate(
            input_ids,
            max_length=max_length,
            pad_token_id=tokenizer.eos_token_id,
            do_sample=True,
            num_return_sequences=2,
        )

    generated_text = tokenizer.decode(output[0], skip_special_tokens=True)
    return generated_text

prompt = "i like cricket very much"
print("Autoregressive Generation:")
print(autoregressive_generation(prompt ))
```

```
Autoregressive Generation:
i like cricket very much but has been getting better, he believes that's been the best of his career

"The first part is when I moved back, it was in the first match I played and it looked cool. I'm excited for it but it didn't feel as good with all the players."

The 19-year-old had spent the pre-season at the Rovers and went on loan at the weekend.

Beswick said: "He got an opportunity to score for the England Under-19s.

"But it is not going to be a quick task. He has done things that no other kid does – getting through three quarters with five yards out and I still think people are very young and he still ha
s a good mind of his own. He's good at his job and I am really proud and delighted that I got my chance. "

England will be without international players for three days from Sunday's semi-final against
```

In [ ]:

```
"""
GPT2LMHeadModel:
   GPT2LMHeadModel is a PyTorch model that combines the GPT-2 language model with a language modeling head.
   It's a variant of the original GPT-2 model, which is a transformer-based language model trained on a large corpus of text data.
model.generate():
   The model.generate() function is a method provided by the Hugging Face Transformers library,
   which allows you to generate text using a pre-trained language model, such as GPT-2, BERT, or RoBERTa.
   pad_token_id:The ID of the padding token.
   eos_token_id:tokenizer.eos_token_id refers to the ID of the End of Sequence (EOS) token.
                The EOS token is a special token that indicates the end of a sequence or a sentence. It's used to signal to the model that it should stop generating text.
   do_sample:When do_sample is set to True, the model will use sampling to generate text.
             This means that at each step, the model will randomly select the next token from the probability distribution over the vocabulary.
             This allows the model to generate diverse and creative text.
   num_return_sequences:num_return_sequences is an integer argument that controls the number of generated sequences returned by the model.
                        When num_return_sequences is set to n, the model generates n separate sequences, each starting from the same input prompt or context.
                        The model uses the same generation process for each sequence, but with different random initializations or sampling strategies
sequence meaning:
                A generated sequence is a sequence of tokens that the model predicts as a possible output, given the input prompt or context.
                A generated sequence typically consists of the following components:
                  1)Input tokens: The initial tokens provided as input to the model, which serve as a prompt or context for generation.
                  2)Generated tokens: The tokens predicted by the model as a continuation of the input tokens.
                  3)EOS token (optional): A special token that indicates the end of the sequence.
```

In [ ]:

```
```