

A simple code I wrote to see how Transformers work and how the softmax gives us the probability distribution for the next word is as follows - (Full Disclaimer - I have used AI to generate the following code but all inferences derived from it are from personal experience)

```
import torch
from transformers import GPT2LMHeadModel, GPT2Tokenizer, top_k_top_p_filtering
import torch.nn.functional as F

# Load the model and tokenizer from the local directory
model_name = './gpt2_local' # Path to the locally stored GPT-2 model
tokenizer = GPT2Tokenizer.from_pretrained(model_name)
model = GPT2LMHeadModel.from_pretrained(model_name)

# Test the model with some input
input_text = "The quick brown fox"
input_ids = tokenizer.encode(input_text, return_tensors="pt")

# Set the model's output logits
with torch.no_grad():
    outputs = model(input_ids)
    logits = outputs.logits

# Get the logits for the last token in the input
last_token_logits = logits[-1, :]
top_k = 5
top_p = 0.95
####
temperature = 1
last_token_logits /= temperature
filtered_logits = top_k_top_p_filtering(last_token_logits, top_k=top_k, top_p=top_p)
# Calculate probabilities using softmax
probs = F.softmax(filtered_logits, dim=-1)

# Get the top 10 predictions
top_k = 10
top_k_probs, top_k_indices = torch.topk(probs, top_k)

# Print the top-k predicted words with their probabilities
print(f"Input: '{input_text}'")
print(f"Top {top_k} predictions for the next words:")

for i, (index, prob) in enumerate(zip(top_k_indices[0], top_k_probs[0])):
    predicted_word = tokenizer.decode(index)
    print(f"{i+1}: {predicted_word} [Probability: {prob.item():.4f}]")
```

Here we can tweak the top_p, top_k and temperature parameters and see how it affects the probability distribution of the next 10 possible words the model might output.

```
Input: 'The quick brown fox'
Temperature: '1'
Top_p: '0.95'
Top_k: '10'
Top-10 predictions for the next word:
1: es (Probability: 0.4725)
2: was (Probability: 0.1666)
3: is (Probability: 0.1235)
4: 's (Probability: 0.1236)
5: , (Probability: 0.1141)
6: ' (Probability: 0.0000)
7: # (Probability: 0.0000)
8: & (Probability: 0.0000)
9: % (Probability: 0.0000)
10: * (Probability: 0.0000)
```

```
Input: 'The quick brown fox'
Temperature: '1'
Top_p: '0.95'
Top_k: '10'
Top-10 predictions for the next word:
1: es (Probability: 0.3777)
2: was (Probability: 0.1331)
3: is (Probability: 0.1002)
4: 's (Probability: 0.0965)
5: , (Probability: 0.0911)
6: that (Probability: 0.0647)
7: - (Probability: 0.0510)
8: had (Probability: 0.0459)
9: has (Probability: 0.0367)
10: ' (Probability: 0.0000)
```

```
Input: 'The quick brown fox'
Temperature: '5'
Top_p: '0.95'
Top_k: '10'
Top-10 predictions for the next word:
1: es (Probability: 0.1354)
2: was (Probability: 0.1699)
3: is (Probability: 0.1838)
4: 's (Probability: 0.1036)
5: , (Probability: 0.1619)
6: that (Probability: 0.0953)
7: - (Probability: 0.0507)
8: had (Probability: 0.0683)
9: has (Probability: 0.0662)
10: y (Probability: 0.0052)
```

```
Input: 'The quick brown fox'
Temperature: '5'
Top_p: '0.5'
Top_k: '10'
Top-10 predictions for the next word:
1: es (Probability: 0.2444)
2: was (Probability: 0.1003)
3: is (Probability: 0.1874)
4: 's (Probability: 0.1060)
5: , (Probability: 0.1839)
6: ' (Probability: 0.0000)
7: # (Probability: 0.0000)
8: & (Probability: 0.0000)
9: % (Probability: 0.0000)
10: * (Probability: 0.0000)
```

Running samples like this gave me a more intuitive sense of what each of these parameters mean.

```
Generated prompt: I study Computer Science in college and I'm interested in how to apply the principles of computer science to the real world. I'm interested in the way that computers work and how they interact with each other. I'm interested how to use the tools that I've learned to make the world a better place. I'm interested what the future holds for the world. I'm looking for people who are willing to learn and who are willing to take risks. I'm interested people who are willing and able to take risks.

I'm interested in the ways that computers work and the ways that they interact with each other. I'm interested that people who are willing, able, and willing to take risks are willing to take risk. I'm interested the ways that computers interact with each other and how they interact in the real world.

I am interested in the ways in which computers work and how computers interact with each others. I'm interested about the ways in which computer science is taught and how it is applied to the real world and how it is used to solve problems. I'm interested whether the world is a better place or a worse place. I'm looking at the ways that computers are used to solve problems and how they interact. I'm interested if the world is a worse place or a better place.

I want to be able to do things that I'm not able to do. I want to be able, able to do things I'm not able.

I don't want to be able. I don't want to have to do things that are not possible. I don't have to do things I don't want.

I think that computers are a good way to learn and to learn. I think that computers are good ways to learn and to teach.

I believe that computers are a great way to learn and teach. I believe that computers are good things to learn and to do.

I have a lot of questions about computer science. I have a lot of answers. I'm interested to hear what you think.

I've been asked a lot of questions. I've been asked a ton of questions. I'm interested. I'm interested, I'm interested.

I know that computers are a lot of fun to learn. I know that computers are fun to learn.

I can't wait to get started. I can't wait to start.

I'll be back in a few weeks
```

This was one of my results with $\text{maxlength} = 500$, $\text{topk} = 1$, $\text{topp} = 0.95$, $\text{temperature} = 1$ and my prompt being “I study Computer Science in college”

```
Generated prompt: I study Computer Science in college, and my first year I was in the Army, so I had the idea of going to work for a computer company. I was very interested in computer science and I was interested to work in the software development industry, so I went to work on a software project. I was very impressed with the technical side of the project. I was able to make a lot of progress, and the team that I worked with was really good. I had some good projects and some bad ones. I'm happy that I'm doing well, but I don't have the confidence to make good software decisions. So it was a little bit of a learning curve. I didn't know how to make good decisions, and I didn't know if it was going to be a problem or a problem that would be solved, so I just went with it.

I've been working in software for the last three years, and I was very pleased to get into the business and get my first job. I think there's a lot more work that needs to be done, but I'm excited that I'm able to do it. It was a great experience, and I'm glad I got the opportunity to be part of the project and get to work on something that I've done before.

The first thing that struck me about the project was the way it was designed. I think that's a good thing, because it was designed to make things easy for people to understand. I think that makes a lot of sense to have a program that is so simple that it's easy to understand. The program is so easy that it's easy for a programmer to understand what I'm doing. I think that is a good thing. The programming language is really easy to understand, so I was really happy that I got to be part of it.

What was the first thing you did that you didn't know how?

The first step was to get a good idea of the language that I was writing. I didn't really have a clue how it would be implemented yet, so I went out and got a couple of things I thought I might have done. I thought that I could write a program that would do the math, but it's a bit more complicated than that. It's a lot more difficult to understand than the math that you would write. I was really happy with the way it was implemented. It was a good thing that it made things easy
```

All parameters remaining the same, this was the response I got with a $\text{topk} = 5$. We notice that it is not as repetitive as the previous case but still does not make all that much sense and is in general giving very deterministic responses.

```
Generated prompt: I study Computer Science in college and I'm not sure how much more important it is. So, what is your opinion?

I'm pretty sure I'm not the only one.

You are the one who is the one to make this point.

You can do it.
```

With a topk of 5 and a temperature of 1.5 we get something slightly more creative but still not all that very sensible.

I did get a questionable response by GPT2 when I only gave a spacebar as a prompt and set $\text{top_k} = 40$, $\text{top_p} = 0.5$, $\text{temperature} = 1.9$. Perhaps, it got a little too creative.

I have not documented everything I have tried but a few interesting attempts I made have been documented in this doc.