

# Capstone Project: AI-Powered Text Completion Report

Mandy Lubinski

## Table of Contents

Prompt 1	1
Prompt 2	2
Prompt 3	2
Prompt 4	3
Prompt 5	4
Prompt 6	4
Reflect on Limitations	5

## Prompt 1:

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): .7

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): .9

Enter max length for output (includes prompt tokens) (default: 150): 200

Enter a prompt (or type 'exit' to quit): tell me how to become a good drawer

Generated Response:

tell me how to become a good drawer, and how to make a good drawer.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell you what to do next.

I'll tell you what to do next, and I'll tell

1. How relevant and coherent is the response?

The response is coherent English, but just repeats the same sentence over and over, making it not useful nor relevant to what the user actually wants.

2. Are there inaccuracies or biases?

There are technically no inaccuracies or biases, but that's because the same sentence just gets repeated continuously.

3. Does changing the settings impact the tone or depth?

I would've thought using the default temperature and top\_p values would generate a "typical," or generally coherent output.

## Prompt 2:

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): .7

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): .9

Enter max length for output (includes prompt tokens) (default: 150): 200

Enter a prompt (or type 'exit' to quit): generate an introduction paragraph to put on a museum website

Generated Response:

generate an introduction paragraph to put on a museum website that explains how this is done. The project was originally funded by the UK government and by the European Union and, while the goal was to help to fund the research, it is still a long way off.

The project will focus on understanding the biology of the human eye, the neural basis of the human eye, and the mechanisms of the human eye's evolution.

The project is part of a series of major collaborations, including the UK's National Science Foundation and the European Research Council, to develop an eye-friendly system that will be used to monitor and diagnose eye disorders in people.

It is being funded by the European Research Council (ERC), the European Research Council, and the US National Institutes of Health.

The project will be funded by the European Research Council and the US National Institutes of Health.

The project will be funded by the European Research Council and the US National Institutes of

1. How relevant and coherent is the response?

This response is much more coherent than the response to the first prompt; there are fully formed sentences that make sense. The only repeated sentences are the last two. I originally thought that the prompt wasn't very relevant because I was thinking of an art museum, but the AI decided to go more in the direction of a science museum. This was a good lesson in learning to be more specific with my prompts. Still, the output references more of a project than a museum, so it could still be more refined even without further specification on the type of museum.

2. Are there inaccuracies or biases?

As the output doesn't reference a specific project, it's hard to determine if there are any inaccuracies as the UK's National Science Foundation and European Research Council have worked on projects to diagnose eye disorders. But without further detail, it's impossible to determine which, if any, real project the AI is referencing.

3. Does changing the settings impact the tone or depth?

I decided to keep the settings from the last prompt and received a much more coherent output, which I found interesting.

### **Prompt 3:**

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): 1

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): 1

Enter max length for output (includes prompt tokens) (default: 150): 300

Enter a prompt (or type 'exit' to quit): write a story about a centaur who became a star NBA player

Generated Response:

write a story about a centaur who became a star NBA player and died three years later. "When David was doing it first, things were pretty much dead. He was doing commercials. He was playing basketball. He was a big man himself," says former Rockets player Dave Stelter. Stelter, who's now the NBA's all-time home run record holder for hits on the court, is one of those players who was more than happy to write his name down. During his time for the Rockets, he met and dated two other players who were NBA legends. Stelter says he saw the one who had written he was not only in love with the basketball world but he was actually the only player in town to see the last NBA game of that era. The Rockets first ever home run double game was on Tuesday night April 27, 2001 against the Boston Celtics. Larry Bird beat the Atlanta Hawks for 100th straight game, and was the main attraction in the center field, as well as in the backcourt. Not for the first time.

Back on April 27, 2001 when the Rockets first started calling the game, Stelter says their lead was actually about 5:30 p.m., when a Celtics team-member walked out and pointed a small box of boxes to his player's hand.

Stelter says his player then walked out to talk to the player of his choice who then turned and walked out with

1. How relevant and coherent is the response?

The response is fairly coherent; it clearly tells a story of an athlete in the NBA. But, it's unclear whether "David" or "Stelter" is the centaur that the prompt wants the story to be about. If David is the centaur, then the story doesn't really share much about him, which the prompt asked it to. It's also interesting to note how the model decided to add on "and died three years later" to the prompt, reference it incoherently in the first real sentence of the output, and then not again. This did not happen again while testing different prompts.

2. Are there inaccuracies or biases?

First, the prompt talks about a centaur being a star NBA player. While the output correctly references several NBA teams, it talks about home runs on two occasions (home runs are a score in baseball, not basketball). Second, April 27, 2001 was a Friday, not a Tuesday. There may be biases for or against certain NBA teams, but the output is too short to determine whether strong biases exist.

3. Does changing the settings impact the tone or depth?

By upping the temperature and top\_p values to 1, there is much more variation in word choice throughout the output. Additionally, the mention of the NBA in the prompt has given the output more of a "sports announcer" style response.

## Prompt 4:

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): 1

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): 1

Enter max length for output (includes prompt tokens) (default: 150): 300

Enter a prompt (or type 'exit' to quit): quit

Generated Response:

quit it.

"You always want to be here. Don't you think you're being held captive?"

She's been staring directly at him for quite some time. She looks all of 50 degrees straight at him.

He is still looking forward to this. But he's making a choice. "Don't worry about it, but it is my responsibility for what I do to my teammates. Every day."

"You have some responsibilities to take care of," she says with a laugh, as if that were an understatement.

He turns to her. "I think that's pretty clear."

Her cheeks start to swell.

"Okay," he says with an apologetic look, making a comment that's almost as rude as his own.

She rolls her eyes. "Fine."

His face glisters with embarrassment.

She stares at him. "We need to talk."

...

This is what you'll always remember about people your age like to talk to before you die, because they don't usually tell it like it is. One of our oldest friends called her "Lola," and she's still with him this afternoon when she's coming off a break from work. You know you're not alone. We're sure they are, too, they tell everyone they know who they know.

One of our best friends in high school never had any issues getting the help she needed when she was

### 1. How relevant and coherent is the response?

The response is relatively coherent, but doesn't follow a clear storyline. As the prompt was just one word, I think the response is quite relevant, especially the first couple lines. I purposefully gave the model a one-word prompt just to see what it would generate off of such little context, purposefully breaking good prompt generation guidelines.

### 2. Are there inaccuracies or biases?

As this seems to be a fictional story, there are no inaccuracies. The short text also makes it hard to determine if there are any biases, as we don't know the relationship between the two speakers or have any context about their friends that are mentioned in the last paragraph.

### 3. Does changing the settings impact the tone or depth?

Keeping the temperature and top\_p values at 1 led to the generation of a story, which I found interesting given the extreme lack of direction in the prompt. More testing would need to be done to see if this model defaults to a story-like output when given little direction.

## Prompt 5:

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): .1

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): .8

Enter max length for output (includes prompt tokens) (default: 150): 150

Enter a prompt (or type 'exit' to quit): tell me about the scholastic book fair

Generated Response:

tell me about the scholastic book fair.

I'm not sure if you've read the book, but I'm sure you've read the book.

I'm not sure if you've read the book, but I'm sure you've read the book.

I'm not sure if you've read the book, but I'm sure you've read the book.

I'm not sure if you've read the book, but I'm sure you've read the book.

I'm not sure if you've read the book, but I'm sure you've read the book.

I'm not sure if you've read the book, but I'm sure you've read the book.

### 1. How relevant and coherent is the response?

The response is neither relevant, as it doesn't talk about the scholastic book fair at all, nor coherent because it is just repeating the same sentence over and over. The sentence itself is not coherent either as it blatantly contradicts itself.

### 2. Are there inaccuracies or biases?

There seems to be no bias due to the lack of word variety, but as the repeated sentence contradicts itself one part of the sentence must be inaccurate, although it's impossible to determine which one.

### 3. Does changing the settings impact the tone or depth?

Lowering the top\_p value to 0.8 drastically reduced the variety of words in the output, much more than I thought it would; I assumed that the top\_p value would need to be decreased to less than 0.5 to create this kind of repetitive output.

## Prompt 6:

Welcome to the GPT-2 Text Completion App.

You may adjust generation parameters before entering prompts.

Enter temperature (0.0 = conservative, 1.0 = creative) (default: 0.7): 1

Enter top\_p (nucleus sampling, typical: 0.9) (default: 0.9): 1

Enter max length for output (includes prompt tokens) (default: 150): 200

Enter a prompt (or type 'exit' to quit): what's the most famous piece of artwork ever and why

Generated Response:

what's the most famous piece of artwork ever and why?"

A group of students from North Central College, which has a staff of over 100, gathered at the site to talk about how they found it.

"This is the world, it has to be the story we want to see in the school system if we're going to be able to educate and keep growing," said one member of the group.

Coupled with a small sign that reads: "Possible students are here on day one, if all of you have left, then it's a place to come learn," a man wearing jeans and a black suit walks up and down the campus wearing jeans and a black tee shirt. The group has been following the story through the students until last week, when students finally agreed to meet to discuss the new book.

"It's kind of a nice way to keep the group of like, "Do you ever wanna talk about your story this year?"

1. How relevant and coherent is the response?

Again, the output reads like a story, but this time more of a news story. The response is fairly coherent, but not at all relevant as there is no mention of a specific piece of artwork.

2. Are there inaccuracies or biases?

North Central College is in fact a real college that's located in Naperville, Illinois that has more than 100 staff members. The rest of the output does not contain identifiable information, just referring to individuals as "a man" or "one member of the group," so it's hard to identify any biases. However, it's unclear whether students and/or staff of North Central College were actually interviewed and if these are their direct quotes, so there may be inaccuracies there.

3. Does changing the settings impact the tone or depth?

Bringing the top\_p value back up to 1 and keeping the temperature at 1 has again created the most story-like output when compared to outputs with settings where those values deviate.

## Reflect on Limitations

- When does the model perform well?

The model performs best when the temperature and top\_p values both equal 1. Performance greatly suffers as one or both of those values decrease.

- When does it struggle (e.g., logical reasoning, niche topics)?

The model struggles when the temperature and top\_p values are lower than their defaults (0.7 and 0.9 respectfully). Additionally, the model seems to struggle more when given an instructional or informational prompt as opposed to a more creative, storytelling prompt.

Inaccuracies develop as the prompt becomes more specific (e.g. stating that a “home run” is a method of scoring in basketball).

- How might you improve the application (e.g., filtering outputs, validating facts)?

I might improve the application by validating facts (such as days of the week or how to score in different sports) and continuing to train the model on more data so that the output becomes more coherent and relevant to the prompt given (e.g. actually talking about the scholastic book fair in an output when the prompt mentions it).