UNIVERSITY OF CALIFORNIA

Los Angeles

An Application of Customized GPT-2 Text Generator for Modern Content Creators

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics

by

Jingwu Fang

2021

ABSTRACT OF THE THESIS

An Application of Customized GPT-2 Text Generator for Modern Content Creators

by

Jingwu Fang

Master of Applied Statistics

University of California, Los Angeles, 2021

Professor Yingnian Wu, Chair

The number of content creators in the cyber world is growing faster year by year, and the competition fiercer. Large video platforms such as YouTube offers creators incentives to upload original content more frequently. However, every creator has a different definition of novelty and uniqueness. The biggest challenge a creator has to face every day lies in the generation and practice of ideas.

As a result, a customized and efficient "idea" generator has become necessary in our times, and any content creator, whether video, advertising, or writing, can benefit from making their content unique efficiently without losing their style. The advent of GPT-2/3 makes this possible, and in this thesis, I will explore the types of model, feasibility of streamlining, and practical challenges of customizing a text generator for content creators nowadays.

The thesis of Jingwu Fang is approved.

Nicolas Christou

Hongquan Xu

Yingnian Wu, Committee Chair

University of California, Los Angeles

2021

To Italo Calvino and Haruki Murakami

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

"I want to thank my mom, Carol Zhang. Thank you."

CHAPTER 1

Introduction

In February 2019, leading artificial intelligence firm OpenAI released GPT-2, a Generative Pre-trained Transformer (Solaiman et al., 2019). This pre-trained transformer specializes in text-related fields such as machine translation, question answering, basic reading comprehension, sentiment analysis, text summarization, and text generation (Solaiman et al., 2019). A year later, in June 2020, a more advanced transformer called GPT-3 was released by OpenAI (OpenAI & Pilipiszyn, 2020).

1.1 Amount of Parameters

In 2012, Alexnet won the champion in the ImageNet competition, with 61 million parameters reaching the human level and representing the state of art image recognition technology (Krizhevsky et al., 2012). A rule of thumb in machine learning is that the more parameters a model has, the more complex the neural network becomes, and the more things a model can learn. However, more parameters also implies more computation, and since language is much more convoluted than graphics, language models need more parameters.

There is a concept in the Silicon Valley startup community called "10X Thinking", which states that a new product must be ten times better than others to make it more appealing. GPT-3's predecessor, OpenAI's GPT-2, has 1.5 billion parameters; Megatron, one of Nvidia's language models, has 8 billion parameters; Microsoft's Turing NLG has 17 billion parameters. GPT-3, on the other hand, has 175 billion parameters. Figure 1.1 visualized this comparison

for us.
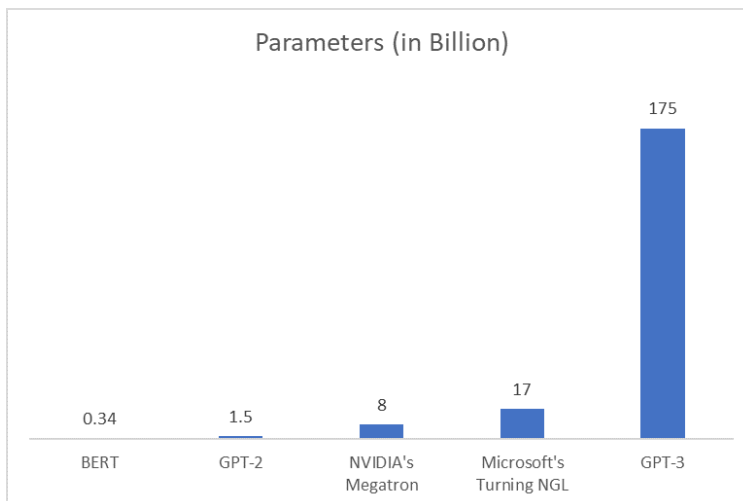
Figure 1.1: Parameters Comparison

With 175 billion parameters, one can claim that the GPT-3 read every book, article, document, computer code manuals, and whatever else a person can find on the Internet that counts as "language".

Having such a gargantuan scale of parameters indicates that GPT-3 can do generic things. Take AlphaGo as an example. If researchers ask it to play Indian Chess, they will have to reprogram and train it all over again. Whereas GPT-3, because it has exhausted any text that ever existed, can do anything without specific training or instruction. Previously, the GPT-2 required specialized training for specific tasks, whereas the GPT-3 did not. Nowadays, scientists only have to build up those 175 billion parameters, feed it a steady stream of text, and GPT learns. The manually tagging part is almost gone as well.

1.2   Working Mechanism

As its name implies, GPT is a variation of transformer model. It first converts words into vectors, vectors are then taken into a dark box (as shown in below Figure 1.2), and the

output of the dark box is converted back to words (Alammar, 2020). The dark box is where 175 billion parameters reside, when a vectorized input comes into the dark box, layers of transformer decoder starts working on it one by one (Alammar, 2020). What happened in the dark box lacks interpretation, humans can audit what had been going on, but cannot fully understand how and why those layers in dark box function like that (Alammar, 2020).
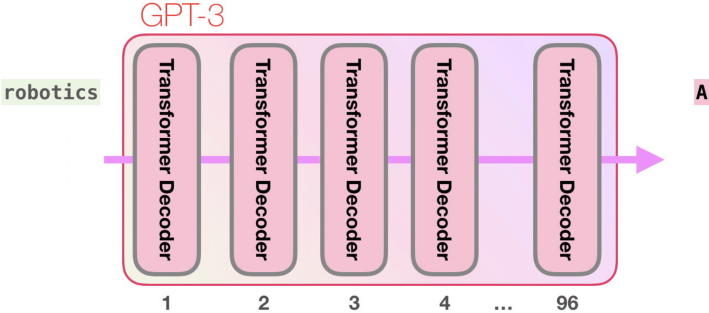
Figure 1.2: Layers of Transformer Decoder within the Dark Box

1.3 Applications

Many applications and businesses are making use of those models since then. Games like AI Dungeon, chatbots like Aivo, and graphic-text models like DALL-E are all successful, proving GPT's outstanding performance. However, GPT is not well known in areas other than the tech world, one of the reasons is that GPT is still not fully accessible. In fact, for safety reasons, OpenAI only released GPT-2 models publicly so far (Solaiman et al., 2019). As for GPT-3, researchers still need to get permissions from OpenAI to use it (OpenAI & Pilipiszyn, 2020). Moreover, GPT does have some drawbacks. Its size forbids GPT to predict quickly and remember details of interactions from a while ago. In addition, GPT under some occasion can be very biased and rude, it seems that it does not fully understand what it is responding from time to time (Alammar, 2019). And that could have brought some serious consequences for GPT when being used.

GPT produces articles that will amaze readers, but could also annoy some of them. Tests have shown that many of GPT's statements are still incoherent, especially when it makes offensive remarks, such as being misogynistic. GPT is acting from experience without fully understanding what statement it is making.

However, that's a great start. It shows the edge GPT has over Google. Instead of googling a topic, people can now ask GPT and receive an automatically generated summary. Instead of writing a piece of code, many programmers google for ready-made code blocks. GPT-3 can generate this code directly for them: it's essentially the same as doing a web search, but it's a lot more convenient.

Let's say Tom is a new employee and this is his first day at work. He received a task from the supervisor. When he was about to start, next to an old-timer spoke up, "Are you new? That's how we do..." He went through it for Tom. As soon as he saw that how it was done, Tom knew what the supervisor is expecting from him.

Almost everything we do at the end of the day, from eating in the canteen at noon to speaking at a meeting in the afternoon to looking for a bus after work in the evening, there is an old-timer beside us who says to us: "Are you new? That's how we do..."

GPT is that old-timer. It knows all the "most likely actions", which is very useful for us. But our goal is to one day say to GPT, "You are old-timer, right? Now I suggest that we do it differently.."

Since GPT-3 is not fully available, I've decided to use the GPT-2 124M and 355M parameter models for this dissertation. They are accessible to everyone, and they are among the smallest GPT-2 models released by OpenAI. If these two models can perform well, bigger models can perform even better. Also, because of their smaller size, it takes less time and less GPU power to train them as well.

1.4   Layout of the Paper

As for the layout of the following chapters, Chapter 2 covers why and what inspired me to choose this topic. Chapter 3 discusses methodology and steps of implementing such model. Chapter 4 serves as a case study where three customized text generators that only outputs Haruki Murakami style sentences will be constructed. Chapter 5 compares and analyzes the results from three text generators. Chapter 6 concludes on which model is the most fitful for modern day content creators' use and end this thesis with the feasibility of streamlining the whole process.

CHAPTER 2

Background

2.1 Italo Calvino's Vision

In 1973, Italian novelist Italo Calvino introduced the idea that books one day can be written by something other than human beings. A mysterious mechanism would be designed to generate stories and characters almost effortlessly. With that thought in mind, Italo Calvino wrote his famous book *The Castle of Crossed Destinies*. In his book, Calvino came up with a system of storytelling purely based on the random draws of Tarot cards (Calvino 20). As Figure 2.1 indicates, the characters and plots were created not by Calvino but by the randomness of that deck of Tarot cards he was using. In the postscript, Calvino states that such a system is revolutionary, the stories and characters invented by it are bestowed with non-human nature, so unique and groundbreaking that only "industrial revolution in literal world" can serve as an adequate analogy (Calvino 325).

Almost 50 years later, although such a system still does not exist, GPT is giving us hope. A new era of content creation when human beings can create side by side with AI is with us now.

With such a powerful tool, even Friedrich Nietzsche's otherworldly notion of *Übermensch*, which can be loosely translated as "above human", has become feasible to be understood and even materialized to some extent. From my point of view, GPT has the potential of assisting content creators to make transcendental works, be that fiction, music, photography, movie, or video.

2.2　As a Content Creator

Being a passionate content creator myself, I run a blog, a YouTube channel, and have posted nearly 400 articles as well as 100 videos. As a content creator who lives to create and creates to live, I found GPT intriguing and inspirational. Even though its creativity can be limited and sometimes it composes strange sentences, GPT's language pattern and use of grammar are completely alien to me. Like the move 37 AlphaGo made when playing Lee Sedol, GPT often strikes me with its original, otherworldly, uninterpretable, but eerily beautiful sentences.

So far, transformers built on DistilGPT-2 have difficulty doing free-writing, it seems writing long paragraphs is still a big challenge for it, let alone writing a fiction all by itself (Huggingface 2019). However, I wonder if GPT technology can assemble an idea generator with much shorter sentences and strong logic. As is known, the biggest struggle of every content creator is to come up with ideas that are better than the previous ones, thus an efficient and customized idea generator should be in high demand. In this thesis, I will explore the feasibility of customizing models based on GPT-2 to suffice the need of ever-growing content creators across the world.
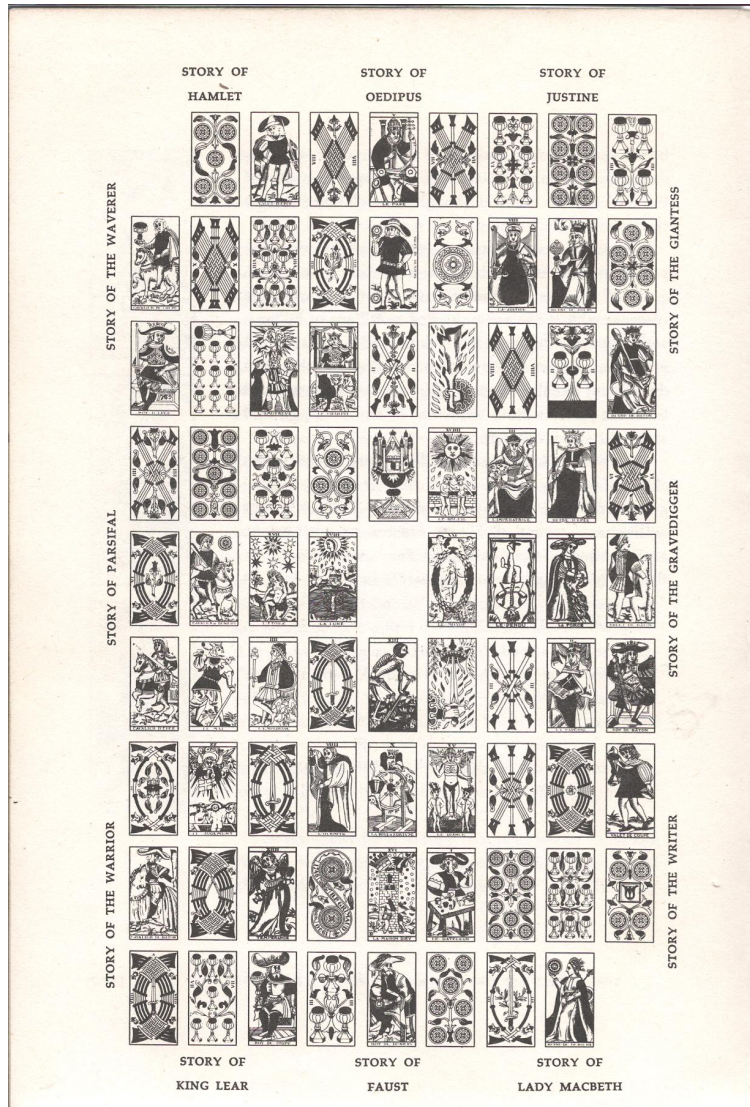
Figure 2.1: *The Castle of Crossed Destinies* via Tarot Cards by Italo Calvino

CHAPTER 3

Methodology

This chapter covers the steps of implementation of a customized text generator based on a GPT-2 model. Later on, we discuss how content creators fine-tune their generators with unique language styles and patterns.

3.1 Choosing GPT-2 Model

Thanks to OpenAI who has open-sourced three versions of the GPT-2 model (shown below as Table 3.1), content creators have three basic models to choose from 124M (small), 355M (medium), and 774M (large) model (Solaiman et al., 2019). Since 124M and 355M are relatively smaller and easy to finetune, they are selected for this thesis. And their performances will be compared in Chapter 5.

Model Size	Number of Parameters	Space on Disk
Small	124M	500MB
Medium	355M	1.5GB
Large	744M	3GB

Table 3.1: Three released GPT-2 models

3.2 Google Colab

Google Colab is the best environment for customizing GPT-2 so far. It not only connects large text files and trained models to Google Drive directly, but also allows users to get access to powerful GPUs that are normally not available among content creators. Figure 3.1 presents us the relationships between Google Colab, Google Drive, and Github Repository.
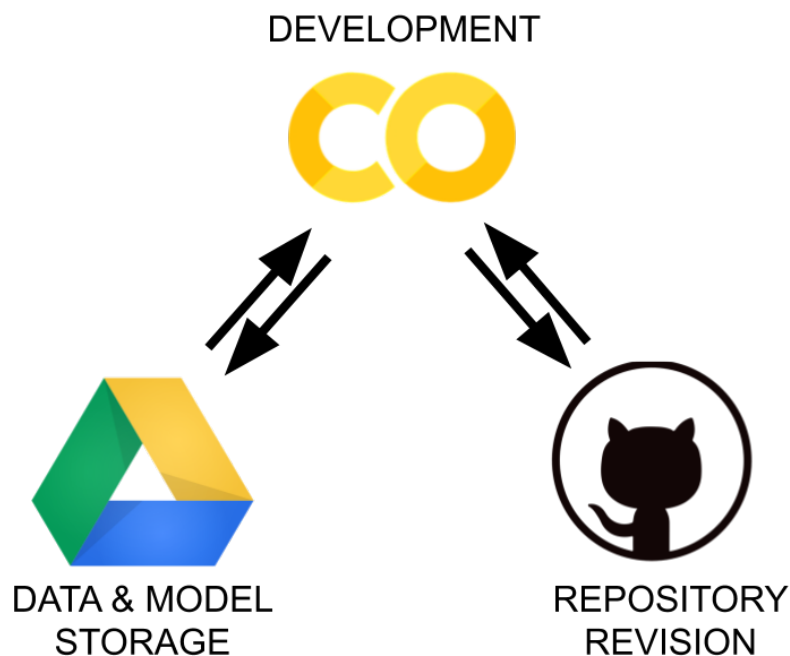
Figure 3.1: Relationship between Google Colab and Goodle Drive

3.3 GPU Type

Training such a model takes a significant amount of GPU power, and regular computers could not have this task done promptly. Therefore, an Nvidia-powered GPU (i.e., Figure 3.2), Persistence-M, is connected via Google Colab, which provides sufficient computational power that is required (Woolf, 2019).

```
      Sun Apr 18 17:36:36 2021
      +-----------------------------------------------------------------------------+
      | NVIDIA-SMI 460.67       Driver Version: 460.32.03       CUDA Version: 11.2   |
      |-------------------------------+----------------------+----------------------+
      | GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
      | Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
      |                               |                      |               MIG M. |
      |===============================+======================+======================|
      |   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
      | N/A   49C    P8    10W /  70W |     0MiB / 15109MiB |      0%      Default |
      |                               |                      |                  N/A |
      +-------------------------------+----------------------+----------------------+

      +-----------------------------------------------------------------------------+
      | Processes:                                                                  |
      |  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
      |        ID   ID                                                   Usage      |
      |=============================================================================|
      |  No running processes found                                                 |
      +-----------------------------------------------------------------------------+
```

Figure 3.2: GPU: Nvidia Persistence-M

3.4   Finetuning and Re-finetuning

Since we are building a text generator that can be constantly re-customized based on the need of content creators, the model must be able to re-finetuned easily. Thanks to TensorFlow's command in Figure 3.3, we can re-finetune that model as many times as we want.

```python
import tensorflow as tf


tf.reset_default_graph()
```

Figure 3.3: The Command that Allows Reset and Re-finetuning

3.5   Text Generation

After the model is fed with the text data that we want it to grasp, it is then trained and can generate us the desired original content that can be further edited by content creators. The procedure is straightforward, and we will be applying this methodology in the next chapter.

Figure 3.4 contains an excerpt of generated text.


```
gpt2.generate(sess, run_name='HARUKI01')
```

The phone rang one day in the middle of the night. It rang three times, stopped, then rang again. It was

"Hey, Tengo," Fuka-Eri said.

This was a very unusual ring. He had never seen it before.

"It's me, Tengo," Fuka-Eri said with a curt nod.

"You're the one who wrote the book," Tengo said.

Fuka-Eri did not reply to this. She simply let her shoulders droop with each successive ring. The phone

"The thing is, you wrote the book. There's no way I could have done it."

"I did it," Fuka-Eri said, but her voice sounded hollow and devoid of warmth.

"Did you just say, 'I wrote the book.' You mean you made it up?"

Figure 3.4: An Example of Generated Text

CHAPTER 4

Case Study on Works of Haruki Murakami

In this chapter, three experiments and their results will be discussed and compared from a model-size perspective and an input-size perspective as shown below in Table 4.1. Haruki Murakami's language style will serve as the main pattern in those experiments. The goal is to find the most effective way to train a model that has the best performance for the content creators.

Both Experiment A and Experiment C were trained with the book *1Q84* , and both Experiment B and Experiment C were trained on 355M GPT-2 model. and the text data that was used to train Experiment B contains all of Haruki Murakami's fourteen novels.

Experiment	Model Size	Input Size
A	124M	One Book(Less than 2MB)
B	355M	All Haruki's Books(Over 10MB)
C	355M	One Book(Less than 2MB)

Table 4.1: Experimental Subjects

Comparisons between model sizes (i.e., Experiment A v.s. Experiment C), text data sizes (i.e., Experiment B v.s. Experiment C), and quality of generated text (i.e., Experiment A v.s. Experiment B v.s. Experiment C) would be discussed in the later chapters.

4.1   Why Haruki Murakami?

First of all, Haruki Murakami is one of the most important modern Japanese writers. He has been a potential recipient of the Nobel Prize for over a decade and is known for his peculiar language style and convoluted storylines. Many of his mysterious and unfathomable sentences have left his readers from all age groups in awe. Therefore, a generator that outputs Murakami-like sentences would be interesting for any content creator to look at and study from.

Title	Year
Hear the Wind Sing	1987
Pinball	1973
A Wild Sheep Chase	1989
Hard-Boiled Wonderland and the End of the World	1991
Norwegian Wood	1989
Dance Dance Dance	1994
South of the Border, West of the Sun	2000
The Wind-Up Bird Chronicle	1997
Sputnik Sweetheart	2001
Kafka on the Shore	2005
After Dark	2007
1Q84	2011
Colorless Tsukuru Tazaki and His Years of Pilgrimage	2014
Killing Commendatore	2018

Table 4.2: All of Haruki Murakami's Fourteen Novels

Another reason that Haruki Murakami is chosen is due to his special technique of writing books. In order to cut down all the unnecessary details and facts, Murakami insists on writing

his thoughts in English, as his English is not great, so that he forces himself to condense his thoughts further and write in the most distilled manner (Kelts 2013). Only after he completed the draft, he would start translating his work back to Japanese for revision (Kelts 2013). Later on, when the book was published, translators would once again translate the book back to English.

These multiple translations between English and Japanese end up making his works in English quite readable and concise (Kelts 2013). To some extent, his works in English are even more grammatically and logically rigorous than books that were originally written in English, which helps us generate correctly-looking text.

Last but not the least, Haruki Murakami is prolific. As we can observe from Table 4.2, there is plenty of text data we can gather in order to train a decent model.

4.2 Experiment A

Feed small text data (only one book) into the smaller GPT-2 model (124M).

4.2.1 Text Data and Model Chosen

For Experiment A, the book *1Q84* was fed to the model. The model that was chosen for this task is the 124M GPT-2 model.

4.2.2 Finetuning

Normally the default setting is enough for content creators to customize their own text generators, thus I made no changes on that part. Figure 4.1 gives the code for default setting for fine-tuning.

```
[ ]   sess = gpt2.start_tf_sess()

      gpt2.finetune(sess,
                    dataset=file_name,
                    model_name='124M',
                    steps=1000,
                    restore_from='fresh',
                    run_name='run1',
                    print_every=10,
                    sample_every=200,
                    save_every=500
                    )
```

Figure 4.1: Default Setting for Finetuning

4.2.3   Text Generation

As we can see in Figure 4.2, the generated conversation looks reasonable and quite engaging. It also captured Tengo's confusion toward Fuka-ri and Fuka-ri's lifeless personality. Although Austria and Russia never appeared in the trilogy *1Q84*, our generator is holding a biased view on them throughout this conversation.

These paragraphs in Figure 4.3 show us the model's potential to write long descriptive text. It is both creative and without a doubt agrees with Haruki Murakami's style. In the actual book, Haruki Murakami was keeping Tengo's family history as a mystery, however, the generator is actively answering that mystery by letting Aomame bringing this to the police station. This is a bold move that even the author himself did not attempt to do.

In general, all of the generated paragraphs are formatted correctly, even the chapter titles looked very much like the titles from *1Q84*, which is discussed in later sections.

4.3   Experiment B

Feed larger text data (all fourteen books over 10MB) into the larger GPT-2 model (355M).

Tengo found this odd, as he hadn't even set foot in Australia.

"Well, you know," the man said, half joking about his own extensive experience of living in a place like this, and half accusing Tengo of being a snob. "I don't get it. Australia gives you plenty of time to think about stuff like this. Is there some reason I'm going to spend so much time in Australia?"

"Well, not really. I just like Australia. Like the Russians."

"But why do you want to spend such a long time in Australia?"

"Well, Australia is such a free country. And because it is such a free country, of course, people have places to go and things to do and feel free to go there. It's the only free country in the world. So why are you so interested in Australia?"

"Because it is the only free country?" Tengo asked, surprised.

"Because it is the only free country?"

"Exactly," Tengo said, sounding genuinely surprised.

"Well, what are you going to do when you're down?"

Figure 4.2: Generated Text from Experiment A(1)

4.3.1 Text Data and Model Chosen

For Experiment B, all of Haruki Murakami's fourteen novels was fed to the 355M GPT-2 model, which includes *Norwegian Wood, 1Q84, The Wind-Up Bird Chronicle, Kafka on the Shore, A Wild Sheep Chase, Hard-Boiled Wonderland and the End of the World, Killing Commendatore, Colorless Tskuru Tazaki and His Years of Pilgrimage, Hear the Wind Sing, Pinball, South of the Border, West of the Sun, Dance Dance Dance, After the Quake, and After Dark.*

The setting for the fine-tuning part is the same as Experiment A.

A few days after Aomame went to the police station to file a report concerning Tengo Kawana's disappearance, a relative of Tengo's contacted them and urged them to file a search warrant at a residential address located in the city of Ichikawa. The relative had no knowledge of the Kawana family, but he was concerned. As a result, he had gone to the city's headquarters and was evaluating the situation there. Headquarters headquarters is in the fashionable Azabu district, but from what he had heard there was a strong possibility that the whereabouts of the woman who disappeared was under surveillance. This was because the police had obtained a search warrant in connection with the Tengo Kawana case. The relative told the headquarters that Tengo lived alone in a small apartment in the city's Shibuya District. Headquarters headquarters did not have the authority to expunge the name Kawana. If that were to happen, it could put the police in a difficult spot. Headquarters headquarters was the most secure location for the investigation, and they were fully aware of the danger.

Aomame visited the headquarters again on Tuesday. She ate dinner alone at the restaurant, then went to the bathroom and brushed her teeth, wiping away the moonlight with a napkin. She went to bed at ten o'clock, rested her forehead against the pillow, and listened to the news. As always, Tengo's memory retained her.

These are some of her favorite memories:

Figure 4.3: Generated Text from Experiment A(2)

4.3.2 Text Generation

All fourteen books being fed into the model certainly adds more dimensions to the story. For example, in the Figure 4.4 paragraphs, Aoyama and Kanagawa are no longer people's name, but village names. A whole description of those villages and histories followed, and it eerily resembles the history of Macondo in Gabriel Garcia Marquez's *one hundred years of solitude.*

However, more dimensions introduces more chaos, as we can observe from paragraphs in Figure 4.5. The uncleanness of the text is a serious issue, and this chaos is rather prevalent throughout the generated text. In addition, the logic between sentences becomes weaker and more blurry. Once in a while, we are able to find a wonderful original piece, however, most of the time, it is less readable and confusing.

Towards the end of the seventeenth century, the village of Painted Grease was settled by a handful of immigrants from the villages of Aoyama and Kanagawa. The village was originally a wintering village, but the inhabitants went on living in the village only two years. Their local government had been organized by the British crown in the late eighteenth century to provide employment for the working class. The village became a major port in the early twentieth century, though its residents had neither a farm nor a car. The settlers were proud of the fact that the village had been settled by cheap, uneducated immigrants. The settlers had been brought here from far and wide in search of work, and they were blessed with a comparatively small number of children. The settlers gradually built an accumulation of houses and other structures, and they lived in the village, although they were under the supervision of the village elders. They worked hard to improve the village, and the settlers became better off. In the late nineteenth century, a heavy sandstorm caused the inhabitants to flee the village, and the village was torn apart by the war of 1914-'15. The settlers were never again to move to the new settlement, and the village was soon deserted by its inhabitants.

Towards the end of the twentieth century, the village of Painted Grease was overcrowded. The town was full of houses and shops, both of which were sold to the immigrants. The settlers never returned, nor did they purchase permanent houses here. The settlement was abandoned, and there were only a few houses or shops left in the village. The settlers were sent to other places, but the village, through the kindness of the sea, remained in existence. The settlers managed to build a few houses, but they did not find work. They began to build a few others. The settlers were tired of living in the village. They had not seen the village for so many years, and they had no desire to go back. They decided to leave the village, and so the immigrants left.

Figure 4.4: Generated Text from Experiment B(1)

4.4 Experiment C

Feed small text data (only one book) into the larger GPT-2 model (355M).

4.4.1 Text Data and Model Chosen

For Experiment C, the book *1Q84* was once again fed to the model. The model that was chosen for this task is the 355M GPT-2 model.

Setting for the fine-tuning part remains the same.

```
 .
\'93But I\'92ve never been cured of anything before. I\'92m sure you can imagine how
painful it is for me.\'94\
\


\
\'93I want you to take care of me for as long as I can.\'94\
\


\
\'93As long as I can,\'94 she said. \'93And I want you to be sure to take care of me, too.
Nothing\'92s going to change that.\'94\
\


\
\'93But I don\'92t think I can be cured of myself. I don\'92t know what that is. I can\'92t
ever change me. I can\'92t forget anything, but I can\'92t forget about myself.\'94\
\


\
\'93I can\'92t believe you would do something like this.\'94\
\


\
\'93I don\'92t know why it would be such a waste of time,\'94 she said, as if to explain. \'93I
can\'92t explain the matter to you, but all I can tell you is that it is not good. It\'92s not
```

Figure 4.5: Generated Text from Experiment B(2)

4.4.2 Text Generation

As the conversation in Figure 4.6 shows, Ushikawa was calling a mysterious woman, who never called any female characters in the *1Q84* trilogy. This conversation is well constructed and engaging. Our model also created a new place called Hotel Okura, which only appeared in this conversation and nowhere else. This conversation itself could be a good place to start a new storylines for Murakami himself, or even the beginning of a new book.

Moreover, it seems the larger GPT-2 model digested one book better than the smaller GPT-2 model. The larger GPT-2 model gives us longer and more readable text, the scenery narration in Figure 4.7 paragraph two on page 22 is especially vivid. Also if we pay attention

Yes, that was it. I'm going back to Tokyo, where I can take care of things."

"Are you going to stay in the Hotel Okura?" Ushikawa asked.

"I'll stay in the Hotel Okura," she said, sounding as if she had just heard the news.

"Why?"

"I don't know," she said.

"Don't you have any friends?"

"I don't have any friends. I just like to go about my business, without bothering anyone. I enjoy being alone. I'm going to stay put, I'm going to take care of things here in Tokyo."

"Don't you get lonely?"

"I don't," she said.

"You don't get lonely?"

"I don't get lonely. I enjoy being alone."

"I don't get lonely," she said.

Figure 4.6: Generated Text from Experiment C(1)

to the chapter name in Figure 4.7 i.e., "Chapter 7 Tengo THE THING THAT MIGHT NOT BE THIS", it is almost indistinguishable from the original trilogy's chapter names.

Figure 4.8 on page 23 contains an excerpt from *1Q84*. It is clear that the customized text generator captured not only the Haruki Murakami's title style, but also the layout of the paragraphs. In my opinion, content creators can get to know their own styles better after applying this technique to their works.

CHAPTER 7 Tengo THE THING THAT MIGHT NOT BE THIS

As he walked, Tengo passed a number of shops along the waterfront. He had been here before, but he was never drawn into that many such places. The only people who kept him company were the small family dogs who barked incessantly in the same places. The dogs themselves seemed to have something to say about their owner. Some even seemed to be looking for him. They barked incessantly, but they never made his way here.

Tengo came to a stop, and he looked out at the beach. The waves were beginning to recede, and the sea was getting higher and higher. The sound of the waves was growing steadily stronger. The waves were not yet as rough as they were on the shore, but they were certainly not so rough, either. This was the first time he had been to see such a scene in his life, and it was certainly the most beautiful thing he had ever seen. The scene was like a nostalgic dream.

Tengo returned the way he came, to the dock, and walked out to the small boat that carried him away. The boat was packed with a small number of people who had come from the nearby shopping district. The people were walking along the shore carrying shopping bags, and people were walking along the shore carrying bags of groceries. The same thing happened when he came by in the taxi that carried him to the pier. Everyone seemed to be carrying something—a shopping bag, some groceries, a refrigerator, a radio, a few books, a few pairs of shoes, a few towels. The people seemed to be considering what they should do next. There was a woman walking a shopping cart. She wore a long-sleeved shirt and dark pants, and her white shoulder bag was full of everything she needed. There were a few towels and a small electric razor, a few pairs of underwear, a few sets of socks, and a box of tissues.

Figure 4.7: Generated Text from Experiment C(2)

CHAPTER 2

Tengo

SOMETHING ELSE IN MIND

Tengo's first memory dated from the time he was one and a half. His mother had taken off her blouse and dropped the shoulder straps of her white slip to let a man who was not his father suck on her breasts. The infant in the crib nearby was probably Tengo himself. He was observing the scene as a third person. Or could the infant have been his twin? No, not likely. It was one-and-a-half-year-old Tengo. He knew this intuitively. The infant was asleep, its eyes closed, its little breaths deep and regular. The vivid ten-second scene was seared into the wall of his consciousness, his earliest memory in life. Nothing came before or after it. It stood out alone, like the steeple of a town visited by a flood, thrusting up above the muddy water.

Tengo made a point of asking people how old they were at the time of their first memory. For most people it was four or five. Three at the very earliest. A child had to be at least three to begin observing a surrounding scene with a degree of rationality. In the stage before that, everything registered as incomprehensible chaos. The world was a mushy bowl of loose gruel, lacking framework or handholds. It flowed past our open

Figure 4.8: Original Text from *1Q84*

23

CHAPTER 5

Model Comparison and Analysis

We conducted three experiments in the previous chapter and have gathered enough data to determine which model generates the text that not only is imaginative, but preserves Haruki Murakami's writing style.

5.1 Why *1Q84*?

The trilogy 1Q84 has a relatively peculiar structure, the first two books only have two-chapter leading names i.e., all the chapter names start with either Tengo or Aomame as Table 5.1 on page 26 shows. Stories develop around these two characters who never even met until the last five chapters of the final book. In addition, all the characters in Tengo chapters never knew the existence of the characters in Aomame chapters. In the third book, another main character, Ushikawa, was added to the scene, and he somehow connects Aomame and Tengo in an obscure way.

5.2 Assessment Criteria

As I mentioned previously, the story develops throughout Aomame and Tengo's perspective for the first two books and the majority of the third book. Those two characters live in each other's parallel universe as Figure 5.1 shows, meaning not only them but all the characters in their stories do not meet each other or even know the existence of each other.
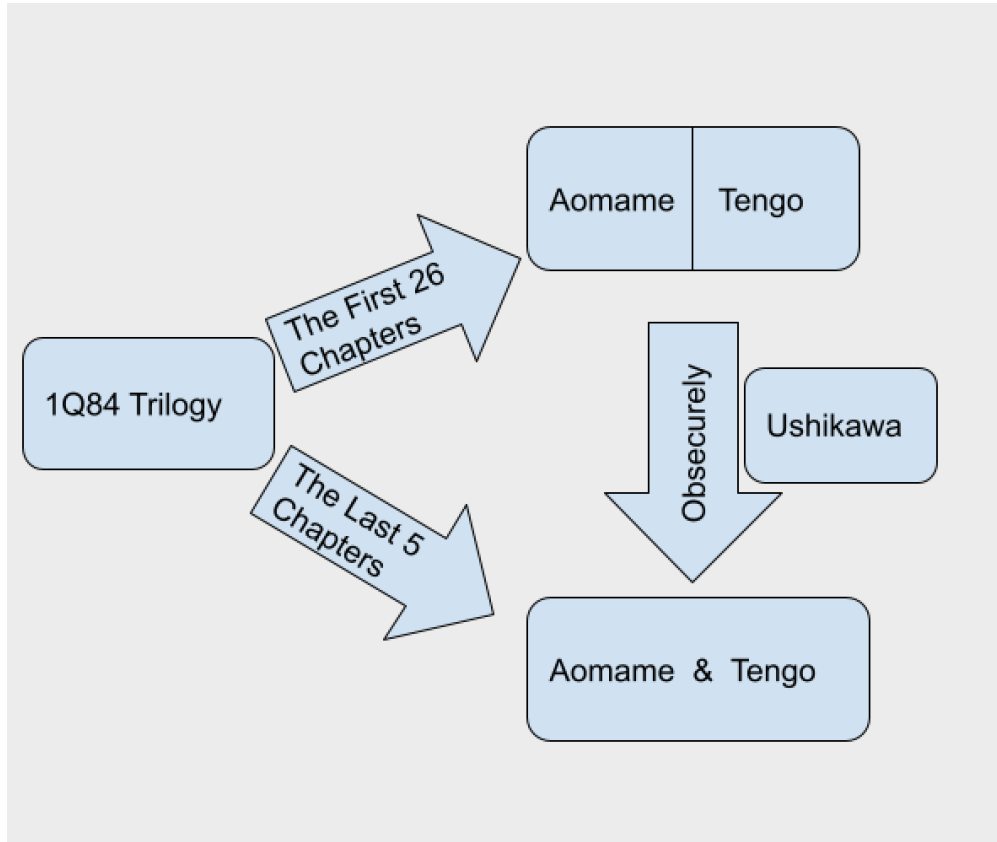
Figure 5.1: Basic Structure of *1Q84*

It wasn't until the last five chapters of the trilogy did Aomame meet Tengo as Table 5.2 on page 27 indicates. Even though the main characters found each other at the end, the characters in their storylines still remain unknown to their counterparts except for Ushikawa, who unilaterally met Tengo and vaguely knows who Aomame was. Yet he passed away for knowing too much.

Thanks to this special setting, we are able to develop a judging criteria for the generated text from the above three experiments. Figure 5.3 gives us such judging criteria.

Chapter	Chapter names
1	**Aomame** DON'T LET APPEARANCES FOOL YOU
2	**Tengo** SOMETHING ELSE IN MIND
3	**Aomame** SOME CHANGED FACTS
4	**Tengo** IF THAT IS WHAT YOU WANT TO DO
5	**Aomame** A PROFESSION REQUIRING SPECIALIZED...
6	**Tengo** DOES THIS MEAN WE'RE GOING PRETTY FAR...
...	...

Table 5.1: Fashion of Chapter Names in the First Two Books

5.2.1 Criterion One: Cleanness and Clarity

As for Cleanness and Clarity, the generated text should be free of typos and scrambled sentences. When content creators are using such model, the last thing they want to see is chaos. Thus, being able to deliver a clean, sharp-looking text is very essential.

And if we compare the cleanness and clarity among those generated text, we would immediately rule out the model from Experiment B. Since it has way too many scrambled sentences, unnecessary repetitions, and strange line breaks as Figure 5.2 suggested on page 28. In addition to this, the generated text from Experiment B in Figure 5.3 looks somewhat incoherent and less logical.

Too many of those "\" and "-" give audience a rather negative reading experience. We also see numbers such as "92", "93", and expressions such as "I couldn" rushing in between words and sentences, which is only making things worse i.e., Figure 5.3.

At the same time, generated texts from Experiment A and Experiment C never had such issue. Furthermore, the other two experiments generate us paragraphs with decent and more logical content as well as narration.

Chapter	Chapter names
...	...
25	**Ushikawa** COLD OR NOT, GOD IS PRESENT
26	**Aomame** VERY ROMANTIC
27	**Tengo** THE WHOLE WORLD MAY NOT BE ENOUGH
28	**Ushikawa** AND A PART OF HIS SOUL
29	**Aomame** I'LL NEVER LET GO OF YOUR HAND AGAIN
30	**Tengo** IF I'M NOT MISTAKEN
31	**Tengo AND AOMAME** LIKE A PEA IN A POD

Table 5.2: Fashion of Final Chapter Names in the Last Book

Level of Difficulty	Criterion
1, Elementary	Cleanness and Clarity of the Text
2, Intermediate	Distinguishing the Distinct Language Pattern
3, Difficult	Level of Creativity

Table 5.3: Assessment Criteria

Thus, it is safe to say that model from Experiment B is ruled out based on this criterion. One take-away from this analysis is that too much input text data would damage the delivery of the generated text.

5.2.2   Criterion Two: Language Pattern

As for the second Criterion i.e., "being able to distinguish and mimic author's unique language pattern", the whole point of adopting and customizing such model for content creator is to help them come up with ideas that are both original and similar to their own styles.

```
\
\
\
\
\
\
\
\
It seemed.\
\
\
\
\
\
\
\
The wind was blowing steadily, the rain still falling.\
\
\
\
\
\|
\
```

Figure 5.2: Strange Line Breaks (Experiment B)

Thus, it is expected that customized GPT-2 would present its users texts with distinct styles that resemble their own. And indeed, if we take a look at the paragraphs from Figure 5.4-5.6 on page 30-31, all three experiments have met this criterion.

5.2.3   Criterion Three: Level of Creativity

Level of creativity is finally what determines our choice of model, whichever tells the most intriguing story gets selected. As shown in Table 5.4 on page 29, I have decided to break it into three mini-levels: Elementary, Intermediate, and Advanced. Moreover, in the case of *1Q84*, I define Elementary level of creativity to be able to resemble the layout and sentence structure of the original text, Intermediate level of creativity to be able to have somewhat abstract notion of time and space, and Advanced level of creativity to be able to come

to. I\'92m sorry. I can\'92t act like I\'92m not sorry.\'94
\

\
\'93I remember what it was like,\'94 the man said. \'93I\'92d be walking along and suddenly
I\'92d be hit. I couldn\'92t get up. I couldn\'92t leave the place. I couldn\'92t move. I felt like
I was falling apart. I tried to get away, but it was too late. I fell down, my body was shaking. I
couldn\'92t get back up. I fell down and got attacked from behind. I couldn\'92t get up. I
couldn\'92t move. I was suffering in that way. It was horrible. I mean, I was dying inside. I
was so weak. I was in a terrible situation. I couldn\'92t even think about what to do. I was so
weak. I couldn\'92t get up. I couldn\'92t get back up. I fell down, my body was shaking. I
couldn\'92t get up. I fell down and got attacked from behind. I couldn\'92t get up. I was
suffering in that way. It was horrible. I mean, I was dying inside. I was in a terrible situation. I
couldn\'92t even think about what to do. I was so weak. I was in a terrible situation. I
couldn\'92t even think about what to do. I fell down and got attacked from behind. I
couldn\'92t get up. I fell down and got attacked from behind. I couldn\'92t get up. I was in a
terrible situation. I couldn\'92t even think about what to do. I was in a terrible situation.\'94
\

\
\'93I know it\'92s a terrible thing, but I was sorry I hadn\'
=====================
You can't just send someone another letter, can you?\

Figure 5.3: Scrambled Sentences (Experiment B)

up with stories that happened to the characters that originally live in each other's parallel universes.

Level of Creativity	Criterion
1, Elementary	Able to Mimic the Original Text
2, Intermediate	Able to Distort Time and Space
3, Advanced	Able to Merge Parallel Universes

Table 5.4: Creativity Criteria

In short, the model in Experiment A achieved Elementary level of creativity, Experiment B ventured into other realms and fulfilled the Intermediate level, and Experiment C did not venture into other realms, however, went far beyond Experiment B and without a doubt met what it takes to be Advanced level.

He was a ten-year-old girl. She was a model and an actress. She was popular with the boys. His father was an NHK fee collector. His mother was a model and an actress. He attended private elementary school in Ichikawa, but his father retired after the class year 2010, and his mother was married to an NHK collector. After his father left NHK, his mother moved to another city, and his father worked as a tax collector in the National Capital Region. He married a worker from a company that did business with the city. The wife had two children.

The elderly couple were in their mid-seventies and slightly obese. Both had faded admiring hair and wore glasses. The eldest child was well built and liked to say the opening lines of his father's story:

"I came here to stay,

But I didn't stay long.

The eldest child had been born in this place and now stays close by. His mother's family have come here to stay,

Which means she and her husband are staying. His father is a farmer from Shibuya, and his mother is a small-town nurse. They both work for a coffeepot company. He is the sole breadwinner of the family. His father and mother work for the coffeepots, but your mother is the only other one working in that company. The three of you brought our little one here by yourself. We didn't invite anyone by yourself. We invited everyone. We invited you, too, each of you. We invited our hearts, and our minds, and our tiny hearts.

Figure 5.4: Murakami Style Paragraphs (Experiment A)

As Figure 5.7 on page 32 shows, generator from Experiment A certainly has some level of creativity. For instance, Ushikawa in the trilogy never had a guru to ask advice from, he is more of a solitary person and enjoys that lifestyle. However, among all the text it generated, I do not see sparks that either merges the parallel universes or distortion between time and space. New elements are introduced to the story, but nothing spectacular.

The highlighted names in Figure 5.8 on page 33 are originally names of the two main characters of the trilogy. However, generator B has decided to turn them into village names, which is bold and intriguing. The follow up narration on vicissitude of the villages has sound logic and is quite engaging. The village of Painted Grease is particularly well-written. It

My name is Soleil, and I was born in the town of Fez, in the province of Kerman. One day in the spring of that year, a woman named May Kasahara came to the town center and told me that a rich man was living there. She was married, and had two children. The rich man was a professional politician. While I was living in Fez, he lived in the house in the same building, and when he died in 1957, he left a large sum of money in my name. He never had a child. When that man died, a large sum of money kept him from being born. I was the only one who had been paid for the rent of the man's house. He had been the son of a wealthy businessman who owned a large business. When May Kasahara came to the town center and told me this, I agreed to pay the man's estate in my name. I had no other choice. I was greatly indebted to him, and he was my father's only son. After I became a student at the university in Fez, I went to work for the man as an assistant in his office. I was not very well at work, and I felt as if I were being punished by a distant power. I had no idea why, but the man always had his assistants pay his expenses. In time, however, I learned that May Kasahara was actually his daughter\'97a beautiful girl who was very fond of other people. She would never marry, and she never had any children. I always saw her as something of a student, and I used to go over to her house and see her often. I felt sorry for her\'97and for me\'97until I saw the picture of a baby she had brought home from the airport. She was a beautiful little girl, but she was in fact nothing but a shell. She never lived a happy life, like a person who had been given a big present by her parents. She was a normal child, with no special powers, and she never went on to be a very bright person.\
\
I was a student at the university in Fez when May Kasahara came to live in the house in the same building. I was the only one who had been paid for the rent of the man's house. At the time, I was in my second year. I used to go over to his place every

Figure 5.5: Murakami Style Paragraphs (Experiment B)

took care of the humble beginning of the village, things that had changed village, and why people were leaving one by one. This kind of short but logical content is exactly what content creators are after. They are easy to adopt and could be developed into various storylines.

Tamaru is a character that only appears in Aomame chapters and has nothing to do with Tengo for the majority of the trilogy. However, in Figure 5.9's conversation, generator C decided to bring Tamaru, Tengo, and Ushikawa into one scene. Their change of words contains ambiguous information regarding the "woman". Who is she? Why is Tamaru so concerned about her call? Questions like these can certainly lead content creators to imaginative lands even Murakami himself has never set foot on before.

Fuka-Eri is a character that only appears in Tengo chapters, and she has never met Aomame in the actual book of *1Q84*. However, in Figure 5.10, our generator for Experiment

The story begins with a young girl, a girl whose name is not given. The story contains many details that are not present in the novel, details that are, if anything, more vivid. The girl is not called Aomame or Princess. She is not given a surname or given a name. The story contains no names or surnames. She is simply called "Aomame." The girl is locked in a fierce battle with a giant, bird-like creature. The battle takes place in the woods—not in the city or in a park, but in the woods of a secluded country. The girl and the creature are both naked, and they are shouting in a low voice, holding each other in the foetal position. Aomame is in the center of the battle, looking down at the girl from the sky. The girl is chained to a pole. The creature has long, thin legs, and its thick, red, tongue emerges from its mouth. The girl is crying out in pain. She is being bitten by the creature's leg.

The story does not tell us who the girl's guardian angel is. What her guardian angel does is not clear. The girl is not raised by her own parents. She is taken in by a foster mother who has no intention of abandoning her. She is taken into this new home when she is two, and then becomes an ordinary child when she reaches the age of ten. The girl has a stable home life, and the foster mother has strong religious beliefs and a strong sense of purpose. She has no enemies. Most of the time, however, they live in a very tense and dangerous situation.

The girl's guardian angel is a young woman in her early thirties. She is the daughter of the dowager's second husband. A young woman in her mid-twenties, she has good looks and is very thin, and she often speaks in a very calm, formal way

Figure 5.6: Murakami Style Paragraphs (Experiment C)

C brought all three of them together in a bus listening to a piece of music-sounded news. This abstract notion of "music-sounded news" is especially otherworldly and could have inspired many content creators in their work.

Based on the comparison and analysis, it is reasonable to conclude that the model used in Experiment C has the best performance. In other words, the combination of smaller text input (less than 10MB) and a larger model (355M GPT-2) can potentially be the best idea generator for content creators.

"I don't follow the rules. No matter how tough a fight it might be, you have to take the initiative."

"I'm not a coward. I just have to take the initiative."

Ushikawa was silent for a while. After thinking about it for a while, he nodded. "That may be true, but I'm going to follow the advice of the guru."

"And you don't have to follow the rules. You just have to take the initiative."

"That's it for today."

Ushikawa concentrated on the word choice for a while and decided to pause and think. He needed some time to think. He finally pulled out a book and started studying it. It was a long, long book. Ushikawa studied the year numbers written in jest at the top. He had read in great numbers in elementary school. There was a huge literature search going on, and a huge amount of material on the history and development of mathematics in the country. He could see that a large number of people had written large numbers of books on mathematics in elementary school. They had been making their own mathematical contributions to society, or making contributions that had some substance, while still making up the smallest possible amount of time. And people were starting to put mathematical formulas in their textbooks. They were starting to put intuition into their formulas. They were beginning to put smell into their formulas. And human beings were beginning to put things into terms they could actually understand. There was a great deal at stake for both sides."

Figure 5.7: (Elementary) Basic Level of Creativity (Experiment A)

Towards the end of the seventeenth century, the village of Painted Grease was settled by a handful of immigrants from the villages of Aoyama and Kanagawa. The village was originally a wintering village, but the inhabitants went on living in the village only two years. Their local government had been organized by the British crown in the late eighteenth century to provide employment for the working class. The village became a major port in the early twentieth century, though its residents had neither a farm nor a car. The settlers were proud of the fact that the village had been settled by cheap, uneducated immigrants. The settlers had been brought here from far and wide in search of work, and they were blessed with a comparatively small number of children. The settlers gradually built an accumulation of houses and other structures, and they lived in the village, although they were under the supervision of the village elders. They worked hard to improve the village, and the settlers became better off. In the late nineteenth century, a heavy sandstorm caused the inhabitants to flee the village, and the village was torn apart by the war of 1914-'15. The settlers were never again to move to the new settlement, and the village was soon deserted by its inhabitants.

Towards the end of the twentieth century, the village of Painted Grease was overcrowded. The town was full of houses and shops, both of which were sold to the immigrants. The settlers never returned, nor did they purchase permanent houses here. The settlement was abandoned, and there were only a few houses or shops left in the village. The settlers were sent to other places, but the village, through the kindness of the sea, remained in existence. The settlers managed to build a few houses, but they did not find work. They began to build a few others. The settlers were tired of living in the village. They had not seen the village for so many years, and they had no desire to go back. They decided to leave the village, and so the immigrants left.

Figure 5.8: (Intermediate) Turning People Names into Village Names (Experiment B)

Tengo looked back at the clock. The two minutes were ticking away.

As he waited for it to be over, the phone rang once. It was Tamaru's voice, as if he had just called from his office for the last time. Tengo could just imagine the man's face when he spoke.

"Tengo?" Tamaru asked. "Hey Tengo."

"Hey Tamaru."

"A phone call from your girlfriend called last night, and she has called again this morning."

"Oh, my girlfriend called, wasn't it? She knows that I'm with Ushikawa at the library."

"That's fine, Tengo, sorry for making you wait. Can you talk now if it's time?"

"All right, so, tell her everything. What's she saying about meeting someone here tonight?"

"You mean the person you always wanted to see? You're going out today? That woman?"

"My real girlfriend? And the one I always wanted to see? She didn't say anything about it. Was she acting strange?"

Figure 5.9: (Advanced) Bringing Characters from Parallel Universes together 01 (Experiment C)

On a road in the country near the city of Ichikawa. A young driver in a bus stopped to let someone out. He noticed that Tengo was there. Tengo jumped out and started to give him a ride. As he leaned against the arm of the bench, Fuka-Eri, Aomame, and Tengo were in the bus, listening to the news on the radio. The driver looked at Tengo and asked an awkward question. "Are you listening to that music?"

"It's part of the news," Tengo replied. He had heard the news on the radio.

"And it's out of print."

"That's right. The New Japan Foundation for the Advancement of Scholarship and the Arts wouldn't touch it."

Figure 5.10: (Advanced) Bringing Characters from Parallel Universes together 02 (Experiment C)

CHAPTER 6

Conclusion

In this thesis, we explored several ways of customizing GPT-2 as an idea generator and found the best balance between the input size and the model size. The result is rather uplifting and the methodology is straightforward. More importantly, since it is easy to use, customizing GPT-2 can be even streamlined for people with basic computer knowledge and weaker GPUs.

Non-English using content creators can also apply this to their project by first translating their contents' languages into English, and then translate the generated text back to the original language. The fundamental logic remains unchanged.

One potential future research direction would be to introduce a factorial design with two factors, one on the complexity of the model and the other on the number of books. For instance, in this paper, I have only done comparisons between one book v.s. all fourteen books. However, comparisons between three books v.s. five books or five books v.s. nine books could also provide us with meaningful results.

Nevertheless, such attempts had been made, and there appeared to be a problem with assessment criteria. The challenging part of comparing more books is that the level of creativity would be difficult for me to assess since I am not so familiar with all of Haruki Murakami's fictions. Therefore, the criteria of judging generators' performances will become harder and vaguer to follow. Unfortunately, the solution for this issue has yet to come to me, future attempts at this will be made.

In all, I believe creating side by side with AI is the future for any industry, and it is our

content creators' responsibility to get used to this tool for future projects and the yet-to-come truly transcendental works.

APPENDIX A

Code

```
%tensorflow_version 1.x
!pip install -q gpt-2-simple
import gpt_2_simple as gpt2
from datetime import datetime
from google.colab import files

#GPU
!nvidia-smi

#Download model
gpt2.download_gpt2(model_name="355M")
#Connect to Gdrive
gpt2.mount_gdrive()
#get the text data ready
file_name = "2.1Q84 - Haruki Murakami.txt"

sess = gpt2.start_tf_sess()
#finetune
gpt2.finetune(sess,
              dataset=file_name,
              model_name='355M',
              steps=1000,
              restore_from='fresh',
              run_name='run1',
              print_every=10,
              sample_every=200,
              save_every=500
              )
#copy to gdrive
gpt2.copy_checkpoint_to_gdrive(run_name='run1')
gpt2.copy_checkpoint_from_gdrive(run_name='run1')
```

Figure A.1: Code Part I

```python
import tensorflow as tf
#enable the re-finetune process
tf.reset_default_graph()

sess = gpt2.start_tf_sess()
gpt2.load_gpt2(sess, run_name='run1')
gpt2.generate(sess, run_name='run1')

#sample generation from trained model
gpt2.generate(sess,
              length=250,
              temperature=0.7,
              include_prefix=False,
              nsamples=5,
              batch_size=5
              )

gen_file = 'gpt2_gentext_{:%Y%m%d_%H%M%S}.txt'.format(datetime.utcnow())
#bulk generation
gpt2.generate_to_file(sess,
                      destination_path=gen_file,
                      length=500,
                      temperature=0.7,
                      nsamples=100,
                      batch_size=20
                      )
#save the generated text
files.download(gen_file)
```

Figure A.2: Code Part II

REFERENCES

[1] Alammar, Jay. *How GPT3 Works - Visualizations and Animations.*
http://jalammar.github.io/how-gpt3-works-visualizations-animations/ 2020.

[2] Alammar, Jay. *The Illustrated GPT-2 (Visualizing Transformer Language Models).*
https://jalammar.github.io/illustrated-gpt2/ 2019.

[3] Calvino and Weaver. *The Castle of Crossed Destinies.* Vintage, 1998.

[4] Huggingface. *Write With Transformer.*
https://transformer.huggingface.co/ 2019.

[5] Kelts, Roland. *Lost In Translation?.*
https://www.newyorker.com/books/page-turner/lost-in-translation 2013.

[6] Krizhevsky, Sutskever, and Hinton. *ImageNet Classification with Deep Convolutional Neural Networks.*
https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
2012.

[7] Murakami, Haruki. *1Q84.* Vintage International, 2013.

[8] OpenAI and Pilipiszyn. *GPT-3 Powers the Next Generation of Apps.*
https://openai.com/blog/gpt-3-apps/ 2020.

[9] Solaiman, Clark and Brundage. *GPT-2: 1.5B Release.*
https://openai.com/blog/gpt-2-1-5b-release/ 2019.

[10] Woolf, Max. *How To Make Custom AI-Generated Text With GPT-2.*

https://minimaxir.com/2019/09/howto-gpt2/ 2019.