NAAN MUTHALVAN

NAME : LOGESH E
REG. NO : 712221104004
YEAR & SEM : 3rd yr & 06
COLLEGE : PARK COLLEGE OF ENGINEERING AND TECHNOLOGY

GENERATIVE AI

STORY WRITING WITH CHATGPT

Agenda

- 1.Problem Statement
- 2.Project Overview
- 3.End Users
- 4.Our Solution and Proposition
- 5.Key Features
- 6.Modelling Approach
- 7.Results and Evaluation
- 8.Conclusion

PROBLEM STATEMENT

1. **Ethical Implications**: As AI becomes more integrated into creative processes, it's crucial to consider the ethical implications of AI-generated content. Issues such as intellectual property rights, ownership of AI-generated ideas, and the potential for AI to perpetuate biases present significant challenges that need to be addressed.

2. **User Acceptance and Adoption**: While AI can offer valuable assistance to writers, ensuring user acceptance and adoption of AI tools is essential. Writers may be hesitant to incorporate AI into their creative processes due to concerns about its reliability, impact on their creative autonomy, and learning curve associated with using new technologies.

3. **Customization and Personalization**: Recognizing that each writer has unique preferences and writing styles, AI-assisted storytelling tools should be customizable and adaptable to individual needs. Providing features such as adjustable writing prompts, customizable feedback options, and personalized recommendations can enhance the user experience and foster greater engagement with AI technologies.

4. **Integration with Existing Writing Workflows**: Many writers have established workflows and habits when it comes to writing. Any AI-assisted storytelling solution should seamlessly integrate with existing writing tools and platforms to minimize disruptions and maximize productivity. Compatibility with popular writing software, such as Microsoft Word, Scrivener, or Google Docs, can facilitate widespread adoption among writers.

5. **Long-Term Impact on Creativity**: While AI can offer valuable support in generating ideas and overcoming creative blocks, there is a need to consider its long-term impact on the creative process. Researchers should investigate how reliance on AI tools may influence writers' ability to develop their creative skills independently and whether it could potentially stifle innovation in the long run.

PROJECT OVERVIEW

- **Title:**
- AI-Enhanced Storytelling: Empowering Creativity through ChatGPT Collaboration
- **Introduction:**
- The project aims to explore the integration of artificial intelligence, specifically ChatGPT, into the creative storytelling process. By leveraging advanced natural language processing capabilities, the project seeks to empower writers to overcome creative hurdles, generate compelling narratives, and enhance the overall quality of their storytelling endeavors.

- **Objectives:**

  Investigate the potential of AI technologies, with a focus on ChatGPT, to enhance the creative storytelling process.

  Develop AI-assisted storytelling tools and methodologies that seamlessly integrate into writers' workflows.

  Address ethical considerations and user acceptance challenges associated with AI collaboration in creative endeavors.

  Evaluate the impact of AI integration on writers' creativity, productivity, and overall satisfaction with the storytelling process.

  Establish best practices and guidelines for effectively leveraging AI in storytelling while preserving creative autonomy and authenticity.

Methodology:

1.Literature Review: Conduct a comprehensive review of existing research and literature related to AI in creative writing, including its applications, challenges, and ethical implications.

2.Tool Development: Design and develop AI-assisted storytelling tools and interfaces that incorporate ChatGPT's capabilities to provide real-time feedback, generate prompts, and assist in plot development.

3.User Testing: Collaborate with writers from diverse backgrounds to evaluate the usability, effectiveness, and user acceptance of the developed tools through iterative testing and feedback sessions.

4.Ethical Considerations: Engage in discussions and consultations with experts in ethics, copyright law, and AI governance to address ethical concerns surrounding AI-assisted storytelling and develop guidelines for responsible usage.

5.Impact Assessment: Measure the impact of AI integration on writers' creativity, productivity, and satisfaction through surveys, interviews, and qualitative analysis of writing samples.

6.Documentation and Dissemination: Document the project findings, including best practices, lessons learned, and recommendations, in a comprehensive report. Disseminate the results through academic publications, conferences, and online platforms to contribute to the broader discourse on AI and creativity.

Expected Outcomes:

1.AI-assisted storytelling tools that provide valuable support and guidance to writers, enhancing their creative process without compromising their autonomy.

2.Insights into the ethical considerations and user acceptance challenges associated with AI collaboration in creative endeavors.

3.Recommendations and guidelines for effectively integrating AI into storytelling workflows while upholding principles of creativity, originality, and ethical conduct.

END USERS

1.Writers: Writers of fiction, non-fiction, poetry, screenplays, and other forms of creative writing are the primary end users. They can benefit from AI-assisted tools and methodologies to overcome writer's block, generate ideas, refine their writing style, and improve the overall quality of their storytelling.

2.Authors and Novelists: Professional authors and novelists seeking to streamline their writing process, enhance their productivity, and explore new creative directions can utilize AI-enhanced storytelling tools to augment their skills and produce captivating narratives.

3.Screenwriters: Professionals in the film and television industry, including screenwriters, scriptwriters, and playwrights, can leverage AI collaboration to develop compelling storylines, create authentic dialogue, and refine plot structures for various visual storytelling mediums.

4.Content Creators: Individuals creating content for digital platforms, such as bloggers, vloggers, podcasters, and social media influencers, can incorporate AI-generated storytelling elements to engage their audience, enhance narrative flow, and increase content quality.

5.Educators and Writing Instructors: Teachers, professors, and writing instructors can integrate AI-assisted storytelling tools into their curriculum to teach creative writing concepts, facilitate writing workshops, and provide personalized feedback to students.

6.Publishers and Literary Agents: Professionals in the publishing industry can explore AI technologies to identify promising manuscripts, evaluate writing samples, and improve editorial processes for selecting, editing, and marketing literary works.

7.Readers and Consumers: Ultimately, readers and consumers of storytelling content are the end users who benefit from AI-enhanced narratives. AI collaboration can lead to the creation of more engaging, diverse, and immersive stories that resonate with audiences across different genres and mediums.

OUR SOLUTION AND PROPOSITION

1.AI-Enhanced Storytelling Tools: Our solution involves developing user-friendly AI-assisted storytelling tools powered by ChatGPT. These tools will offer features such as real-time feedback on writing samples, personalized writing prompts tailored to individual preferences, and assistance with character development, plot structuring, and dialogue creation.

2.Interactive Writing Assistant: Our proposition includes an interactive writing assistant interface that seamlessly integrates with popular writing software and platforms, such as Microsoft Word, Google Docs, and Scrivener. This interface will enable writers to collaborate with ChatGPT in real-time, allowing for fluid interaction and creative exploration throughout the writing process.

3.Customizable AI Models: To address the diverse needs and preferences of writers, our solution will offer customizable AI models that can be trained on specific genres, writing styles, and thematic elements. Writers can fine-tune the AI's suggestions and feedback to align with their creative vision and narrative goals, ensuring a personalized and tailored experience.

4.Ethical Guidelines and Transparency: In response to ethical considerations surrounding AI collaboration in creative endeavors, our proposition includes the implementation of clear guidelines and transparency measures. Writers will have full control over how AI is utilized in their creative process, with safeguards in place to protect intellectual property rights and ensure the authenticity of their work.

5.Community Engagement and Support: Our solution includes community engagement initiatives aimed at fostering collaboration, sharing best practices, and providing support to writers utilizing AI-assisted storytelling tools. Online forums, workshops, and peer-to-peer mentorship programs will facilitate knowledge exchange and encourage writers to explore new creative possibilities with AI.

6.Evaluation and Improvement: To continuously enhance the effectiveness and usability of our solution, we propose ongoing evaluation and improvement processes. User feedback mechanisms, performance analytics, and iterative updates will enable us to refine the AI models, optimize the user experience, and address emerging challenges and opportunities in AI-enhanced storytelling.

7.Education and Training Programs: As part of our proposition, we will develop education and training programs to empower writers to leverage AI technologies effectively in their creative endeavors. Workshops, webinars, and educational resources will provide practical guidance on integrating AI into writing workflows, fostering digital literacy, and navigating ethical considerations.

8.Partnerships and Collaborations: Our solution involves forming strategic partnerships and collaborations with industry stakeholders, including publishers, writing organizations, and academic institutions. By working together, we can leverage collective expertise, resources, and networks to drive innovation, promote adoption, and advance the field of AI-enhanced storytelling.

KEY FEATURES

1. **Real-time Feedback**: Receive instant feedback on writing samples, including suggestions for improvement in plot, characterization, dialogue, and narrative structure.
2. **Personalized Writing Prompts**: Access tailored writing prompts generated by ChatGPT based on individual preferences, genre preferences, and thematic interests.
3. **Character Development Assistance**: Utilize AI assistance to brainstorm character traits, arcs, motivations, and relationships, ensuring consistency and depth in character development.
4. **Plot Structuring Tools**: Benefit from AI-powered tools to outline and structure story plots, including assistance with plot twists, pacing, and story progression.
5. **Dialogue Generation Support**: Receive assistance in crafting authentic and engaging dialogue through AI-generated suggestions and prompts.
6. **Genre-specific AI Models**: Access customizable AI models trained on specific genres, writing styles, and narrative conventions to tailor AI assistance to your storytelling needs.
7. **Seamless Integration with Writing Software**: Integrate AI-assisted storytelling tools seamlessly with popular writing software and platforms, such as Microsoft Word, Google Docs, and Scrivener.

MODELLING APPROACH

1. **Instruction of Natural Language Processing**

Natural language processing (NLP) is a sub-field of linguistics, computer science, and artificial intelligence. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. NLP mainly includes three parts: natural language generation, speech recognition, and natural language understanding. This project mainly focus on natural language generation.

2. Neural Network

(1) . Basic introduction of RNN With the development of this field, nowadays, Neural Network is widely used in designing NLP projects. Convolutional neural network (CNN) and recurrent neural network (RNN), the two main types of DNN architectures, are widely explored to handle various NLP tasks.1 RNN performs better in NLP, and it is widely used in it.

The value s of the hidden layer of RNN not only depends on the current input x, but also depends on the value s of the last hidden layer:

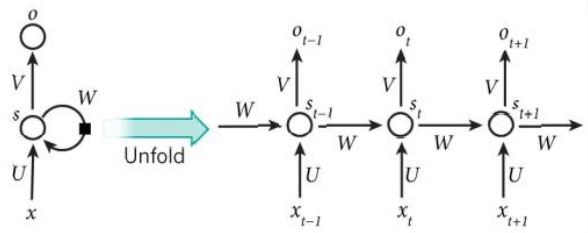

Figure 1. schematic diagram of RNN

Where W is the last value of the hidden layer as the weight of the input this time. t is the time, x is the input layer, s is the hidden layer, o is the output layer.

Different from normal neural network, RNN can look forward to any number of input values:

$$o_t = g(V_{S_t}) \tag{1}$$

$$s_t = f(U_{X_t} + W_{S_{t-1}}) \tag{2}$$

Continue substituting Eq. 2 into Eq. 1:

$$
\begin{aligned}
o_t &= g(V_{S_t}) \\
&= Vf(U_{X_t} + W_{S_{t-1}}) \\
&= Vf(U_{X_t} + Wf(U_{X_{t-1}} + W_{X_{t-2}})) \\
&= Vf(U_{X_t} + Wf(U_{X_{t-1}} + Wf(U_{X_{t-2}} + \ldots)))
\end{aligned}
$$

3. GPT-2

GPT-2 is a language model released by OpenAI to predict the next word in 40GB of Internet text. GPT-2 is a large transformer-based language model with 1.5 billion parameters. Simply, the role of the language model is to predict what the next word will be based on a part of an existing sentence. Through this project, using GPT-2-simple to finetune the model and generate text.[2]

4. GPT-3

GPT-3 is the most powerful language model ever. Its predecessor, GPT-2, released last year, was already able to spit out convincing streams of text in a range of different styles when prompted with an opening sentence. But GPT-3 is a big leap forward. The model has 175 billion parameters (the values that a neural network tries to optimize during training), compared with GPT-2's already vast 1.5 billion.[3] Different from GPT-2, GPT-3 doesn't require to be finetuned in the conventional way, with GPT-3, it helps to anthropomorphize it: sometimes users literally just have to ask for what they want.[4]

5. Project description

The goal of this project is using the fictions written by Dr. Seuss, E.B White, Ian Fleming, James Herriot and Samin Nostrat to generate an novel which hero is called Humpty Dumpty, and talk about his adventure from table to field.

The first approach of this project is using all this book as the training data, using Keras and Tensorflow, build the neural network and train it. But the result cannot case sensitive and there are no punctuation. The contents of text generated is not satisfactory.

Train 200 epochs:

Humpty Dumpty sat on a wall, and when he climbed in rolling the goose sat out of the load the song of the crickets weather vane anybody talked just as a little pig were happened to her he he began to speak about talk as she dropped in the zuckermans just in the ground mrs weather weather weather vane do do and don't you in a fair is a spider have wake wake spoken letter talk just the spider bam to the spider have whiskers radiant most eggs eggs the zuckermans were intend to speak about me he said in disgust it was just in a

Figure 2. Text generated by using Keras and Tensorflow

The second approach is funetuning GPT-2 pre-trained model with training data and generate text. After choosing the biggest model and adjust the parameters, it can generate more natural sentences.

```
Humpty Dumpty sat on the wall, facing the
door. He was tired. His stomach was still
swollen from last night's gorge. Charlotte, at work,
was mending a pail.
"Did you know that Humpty Dumpty has a guest?"
"Who? " said Mr. Zuckerman.
"Who?" repeated Mr. Zuckerman.
"I, of course," said Lurvy. "I know everything."
"I beg everyone's pardon," whispered Humpty Dumpty. "I
didn't mean to be objectionable."
Lurvy chuckled. "I'm not going to spend all my time
chasing after mysterious visitors," he said. "I'm going
to have a talk with Templeton about foxes and spiders."
"Well, I'm going to have a talk with Templeton," said Humpty Dumpty. "Why
doesn't Templeton bring his cat along? I think Temple-ton would be delighted to bring his beloved cat along. Why can't
he?"
"Well," said the old sheep, "he's a pig."
"What does he look like? " asked the bambee.
"Weil-l," said the sheep, "he's a pig."
"He's a pig!" cried Humpty Dumpty. "I'm going to
save him."
"Can I have a drink of milk?" asked the goose.
"Of course you can," said Lurvy. "Just leave me alone."
"You ought to be ashamed of yourself," said Humpty Dumpty.
"Not at all," said Lurvy. "It's quite natural for a
young pig to want to be alone."
"I think," said Humpty Dumpty, "he would be delighted to spend
the night in my company."
"The same as I am," said the goose.
"Well," said Humpty Dumpty, "I'm going to bed early.
I think I'll ask one more time, if I'm going to last
longer in this room."
"Go to sleep, Charlotte," said Humpty Dumpty. "Sleep is
the first sign of health."
Charlotte fell into a deep sleep. She did not wake until after dark.
Humpty Dumpty wasted no time. He crept into the
```

Figure 3. Text generated after finetune GPT-2 model

GPT-2 is better at maintaining context over its entire generation length, making it good for generating conversational text. The text is also generally gramatically correct, with proper capitalization and few typoes.[5] But when looking into to the full text, it doesn't build the relationship between sentences and sentences, in other words, the whole structure of the novel has not been built.

In order to solve this problem, using GPT-3 to generate text becomes the third approach, since this model has 175 billion parameters which is much bigger than GPT-2. Given any text prompt like a phrase or a sentence, GPT-3 returns a text completion in natural language. Developers

can "program" GPT-3 by showing it just a few examples or "prompts."[6]

The extraction was peculiarly limited. That any human or living creature survived indicated that the removal of iron was limited to non-organic creatures: had iron been removed from hemoglobin and the oxygen content of blood reduced to zero, every mammal would have asphyxiated in seconds. Miners reported that no iron seams were accessible in existing iron mines; yet the iron core of the Earth could not have been removed as the crust had not collapsed, the Earth's orbit seemed to be unchanged, and the magnetic field retained its usual strength. Naked-eye observations ruled out the possibility that the core had been removed from Venus or Mars, and Mars retained its red color, but the margins of error were too large to rule out orbital aberrations due to hypothetical removal of gigatons of iron from the subsurface of either.

The aliens' motivations remained unclear. Unknown physics was involved, as no mechanism could even be hypothesized. A flotilla of vessels spoke of a rich interstellar civilization, which was impossible without millennia of peace and cooperation, but their actions supported the "intelligence implies belligerence" thesis: humanity had stored up so much data in digital libraries (often routed through communication satellites) that it was impossible that the aliens could not communicate with us; and yet they did not, communicating neither demands nor warning nor explanation. Perhaps they were doing precisely what it looked like - mining for allotropic iron? But why mine in such a destructive method, and why choose the Earth rather than any of the other rocky iron-rich planets, or better yet, the asteroids? (One astronomer noted that with the elimination of certain astronomical programs, it was possible that the asteroids had been mined out previously but we had not seen it.)

Iron in any of its derivatives such as steel, humanity was forcibly reminded, was crucial to almost all enterprises. A proud column of steel, deprived of iron, is but a mist of chromium and other adulterants which supports nothing. As the absence crept along the globe, it was followed by darkness and large red pinpricks of light.

Figure 4. Novel generated by GPT-3

can ”program” GPT-3 by showing it just a few examples or ”prompts.”[6]

The extraction was peculiarly limited. That any human or living creature survived indicated that the removal of iron was limited to non-organic creatures: had iron been removed from hemoglobin and the oxygen content of blood reduced to zero, every mammal would have asphyxiated in seconds. Miners reported that no iron seams were accessible in existing iron mines; yet the iron core of the Earth could not have been removed as the crust had not collapsed, the Earth's orbit seemed to be unchanged, and the magnetic field retained its usual strength. Naked-eye observations ruled out the possibility that the core had been removed from Venus or Mars, and Mars retained its red color, but the margins of error were too large to rule out orbital aberrations due to hypothetical removal of gigatons of iron from the subsurface of either.

The aliens' motivations remained unclear. Unknown physics was involved, as no mechanism could even be hypothesized. A flotilla of vessels spoke of a rich interstellar civilization, which was impossible without millennia of peace and cooperation, but their actions supported the "intelligence implies belligerence" thesis: humanity had stored up so much data in digital libraries (often routed through communication satellites) that it was impossible that the aliens could not communicate with us; and yet they did not, communicating neither demands nor warning nor explanation. Perhaps they were doing precisely what it looked like - mining for allotropic iron? But why mine in such a destructive method, and why choose the Earth rather than any of the other rocky iron-rich planets, or better yet, the asteroids? (One astronomer noted that with the elimination of certain astronomical programs, it was possible that the asteroids had been mined out previously but we had not seen it.)

Iron in any of its derivatives such as steel, humanity was forcibly reminded, was crucial to almost all enterprises. A proud column of steel, deprived of iron, is but a mist of chromium and other adulterants which supports nothing. As the absence crept along the globe, it was followed by darkness and large red pinpricks of light.

Figure 4. Novel generated by GPT-3

Based on the huge mount of parameters and complicated neural network of GPT-3, we expect to generate the real novel which describe the adventure of Humpty Dumpty from the table to field. The content and structure of this novel will imitate Dr. Seuss, E.B White, Ian Fleming, James Herriot and Samin Nostrat's books.

All of these models understand and process text by breaking it down into tokens. Tokens can be words or just chunks of characters. For example, the word “hamburger” gets broken up into the tokens “ham”, “bur” and “ger”, while a short and common word like “pear” is a single token. Many tokens start with a whitespace behind, for example “hello” and “ bye”.[7]

Through this project, the completions endpoint is the most important part. Input some text as a prompt, and the model will generate a text completion that attempts to match whatever context or pattern you gave it.

7. Text completion in GPT-3.

The GPT-3 completions endpoint can be used for a wide variety of tasks. As discussed above, if input some text as a prompt, the model will generate a text completion. That is the basic and the most important idea of this project.

In [1]:
```
import os
import openai
```

In [2]:
```
openai.api_key = "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"
```

In [3]:
```
response = openai.Completion.create(
  engine="davinci",
  prompt="As Descartes said, I think therefore",
  temperature=0,
  max_tokens=64,
  top_p=1,
  frequency_penalty=0,
  presence_penalty=0,
  stop=["."]
)
```

In [4]:
```
prompt="As Descartes said, I think therefore"
print(prompt+response.choices[0].text)
```

```
As Descartes said, I think therefore I am
```

This figure shown above is a very simple text completion task, for the task which is more complicated, adjusting the parameter of the model to get different results.

8. Fine-tuning of GPT-3.

8.1 Fine-tune.

As mentioned above, the goal of this project is to generate the real novel which describe the adventure of Humpty Dumpty from the table to field. The content and structure of this novel will imitate Dr. Seuss, E.B White, Ian Fleming, James Herriot and Samin Nostrat's books. If directly generate this text without these books, it will generate the text not corresponding to the goal. So it is necessary to fine-tune the model based on our training data.

8.2 Fine-tuning in GPT-3.

GPT-3 has been pre-trained on a vast amount of text from the open internet. When given a prompt with just a few examples, it can often intuit what task are trying to perform and generate a plausible completion. This is often called "few-shot learning." It can also get higher quality results than prompt design, it also has the ability to train on more examples than can fit in a prompt.

Different from GPT-2 and other former products, preparing the GPT-3 training data requires the format which is shown below:[8]

```
1  {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
2  {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
3  {"prompt": "<prompt text>", "completion": "<ideal generated text>"}
4  ...
```

Figure 6. Format of GPT-3 training data.

This data must be JSONL document. since our training data are text files, the first step is transferring these text files into JSON format, the figure below shows how to transfer one og these books into JSON file:

In [1]:
```
import json
import os
```

In [2]:
```
fields =['prompt', 'completion']
with open("chv_HD_replaced.txt") as file:

    lines=file.readlines()
    for l in lines:

        l = l.strip("\n")
        string=l.split("\n")
        #print(string)
        #print(lines)
```

In [3]:
```
fields =['prompt', 'completion']
dict1 = {}
```

In [4]:
```
with open("chv_HD_replaced.txt") as fh:
    l = 1

    for line in fh:
        description = list( line.strip().split(None, 1))
        print(description)
        sno =str(l)
        i = 0
        dict2 = {}
        while i<len(fields):
            dict2[fields[i]]= description[i]
            i = i + 1
        dict1[sno]= dict2
        l = l + 1
```

In [5]:
```
out_file = open("test2.json", "w")
json.dump(dict1, out_file, indent = 4)
out_file.close()
```

Figure 7. Code of transfer text to JSON.

After finishing these transferring tasks, training data becomes the format shown below:


```
{
      "1": {
            "prompt": "WHERE'S",
            "completion": "Papa going with that ax?\""
      },
      "2": {
            "prompt": "said",
            "completion": "Fern to her mother as they"
      },
      "3": {
            "prompt": "were",
            "completion": "setting the table for breakfast."
      },
      "4": {
            "prompt": "\"Out",
            "completion": "to the hoghouse,\" replied"
      },
      "5": {
            "prompt": "Mrs.",
            "completion": "Arable. \"Some pigs were born last night.\""
      },
      "6": {
            "prompt": "\"I",
            "completion": "don't see why he needs an ax,\" continued Fern,"
      },
```


Figure 8. JSON file of training data.

OpenAI also offers a tool called CLI data preparation tool[9] to convert the data into this file format. After get all files in a correct format, fine-tuning job can be started.

The fine-tuning job has not been finished, I expect to finish it in the next week, at that time, this project report will be updated.

8.3 Implement of fine-tuning in GPT-3

To implement fine-tuning, we will first need to set up the environment, in this example, we will use all of novels by Dr. Seuss as our fine-tuning data. Run the code shown below:

```
!openai tools fine_tunes.prepare_data -f "/content/Dr.Seuss_all.txt"
```

This code will convert the TXT file into JASONL file which is the correct format of our fine-tuning job, it will also remove empty completions and duplicate rows. The result is shown below:

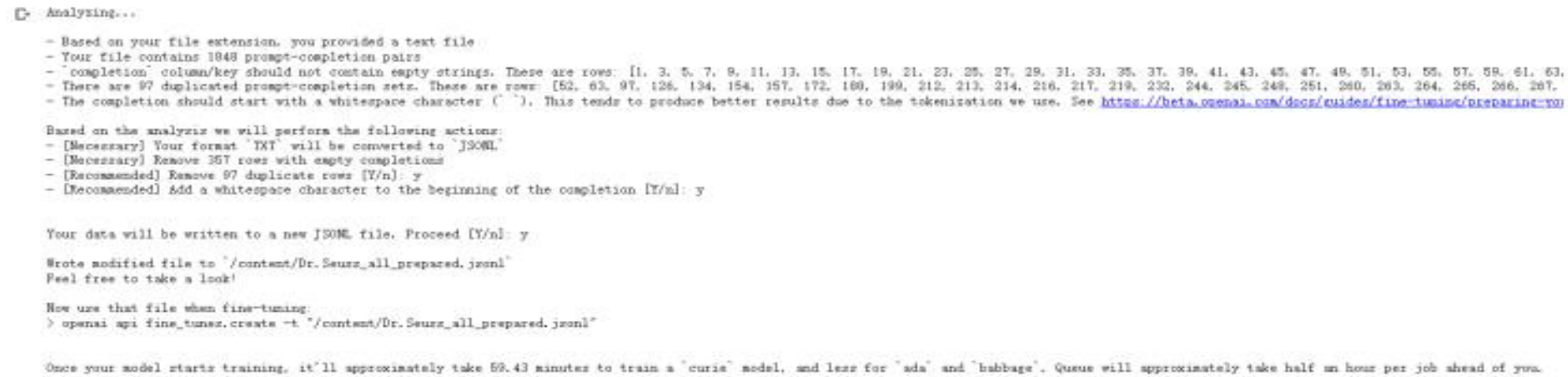

Figure 9. convert TXT file into JASONL file.

Once we got the JASON format of dataset, we should be able to use it in creating a fine-tuning job, run the code below:

```
!openai api fine_tunes.create -t "/content/Dr.Seuss_all_prepared.jsonl" -m "ada"
```

Since this project is implemented on Google colab environment, the stream may interrupte due to client disconnected, in order to solve this problem, run the code below to resume the routine:

```
!openai api fine_tunes.follow -i ft-JpsUMgZ3sGDcA781N2MBUFFF
```

In this example, we choose model “ada” as our fine-tuning model, it can give us a shorter time to complete our fine-tune job. After fine-tuning, we should be able to see the result below:

```
[2021-12-22 06:20:01] Created fine-tune: ft-Fj6E5mk4AMSCwoUZkgSwhWRh
[2021-12-22 06:20:11] Fine-tune costs $0.08
[2021-12-22 06:20:12] Fine-tune enqueued. Queue number: 0
[2021-12-22 06:20:15] Fine-tune started
[2021-12-22 06:26:39] Completed epoch 1/4
[2021-12-22 06:32:48] Completed epoch 2/4
[2021-12-22 06:38:56] Completed epoch 3/4
[2021-12-22 06:45:06] Completed epoch 4/4
[2021-12-22 06:45:33] Uploaded model: ada:ft-leqing-2021-12-22-06-45-31
[2021-12-22 06:45:36] Uploaded result file: file-kUdJhiCl5UOtVD5L2H6oJAmX
[2021-12-22 06:45:37] Fine-tune succeeded

Job complete! Status: succeeded 🎉
Try out your fine-tuned model:

openai api completions.create -m ada:ft-leqing-2021-12-22-06-45-31 -p <YOUR_PROMPT>
```

Figure 10. a demo of successful fine-tune.

After completing our job, a fine-tuned model is created and saved to account, for example, if we want to generate Dr. Seuss's style text, we can directly use this model, just same as we use models given by OpenAI.

To test performance of our model, run the code below:

```
response=openai.Completion.create(
    model="ada:ft-leqing-2022-01-12-04-12-45",
    prompt="I' m sending you right up!",
    max_tokens=200,
    temperature=0.9)
prompt="I' m sending you right up!"
print(prompt+response.choices[0].text)
```

By observing the result we found that it generated Dr. Seuss style text, that means our fine-tuning job successed. However, our ultimate goal is to generate a novel which contains logic, rich content and specific style, we will need to combine different novels written by different authors.The result given by our fine-tuned model is shown below:

```
I'm sending you right up!
You may, I say.
You don't read books.
You read people.
You read places.
You read people's worries.
You read people's needs.
You read people's desires.
You read people's needs.
You read people's aspirations.
You read people's dreams.
You read people's agraves.
You read places where people are Harmony and hope.
You read books that tell you what to task your artsingers are doing.
You read a book that outlines you'll give you's culture a stacker.
You read books that put you's in the terrarium.
```

Figure 11. Dr. Seuss's style text generated by fine-tuned model

9. Difficulties and solution

As mentioned in pervious section, our ultimate goal is to generate a novel which contains logic, rich content and specific style. Sometimes if we only do the fine-tuning job and only use it to generate long text, duplicated and chaos text might be generated. Also, the maxium tokens allowed in GPT-3 is 2048 tokens, which means our prompt plus generated text should not exceed 2048 tokens. It also becomes a difficulty because we want to build the connection between paragraph and paragraph. After thinking about those difficulties, we proposed an idea that we can use our fine-tuned model to complete the first a few sentences as our start of our novel, then we switch back to model given by OpenAI and use that model to continue our job. The reason is these model contain a big amount of parameters and they have a very strong self-adaption abilty. Once we use our fine-tuned model to generate some sentences and use those sentences as our input, for example, the model “Davinci” could be able to imitate those input and finish the text completion. A result below could prove our idea is correct:

The sun was just starting to rise and the sky was a pale.The cow was standing, as usual, in the far corner of the fold yard, calf on her shoulder.the calf raised its head and wagged its tail.The first few birds had sleepy eyes and a slow start, but by the second half of the morning, the birds were up and about, the sun was out and the birds were singing. Today these birds were singing more than usual as if they had something to celebrate.

Mr. Dumpty was an engineer, he is so obsessed with machines and technology that he barely has time for anything else. People often thought that he was a bit strange, he even made some robot friends that he would talk to, and they would talk back. But his most favourite robot is a centaur, which he named Chiron. Chiron was 6 feet tall, and its body was made of metal and wires. It had a large computer screen for a face, and it could walk, talk, and even run. Mr. Dumpty loved Chiron because it was his best friend, and he felt like they understood each other. Chiro could also help Mr. Dumpty with his work, which made life a lot easier. People always said Mr. Dumpty's talent for machines and technology came from his ancestor, the great Mr. Dumpty the first, who made the big cannon in the English Civil War.

Figure 12. Long text generation by different styles

As the figure show above, we fisrt use our fine-tune model to generate Jame Herriot's style text and use it as our first paragraph, then we generate E.B. White's style text, use those sentences as the input to model "Davinci" and continue text completion. In order to escape the maxium tokens issue and build connection between paragraph and paragraph, our idea is to generate a paragraph fisrt, and use this output paragraph as our next generation input. It will somehow give the logic to the whole novel, we call it iteration idea. Based on our fine-tuned model and iteration idea, we have reason to believe that it can generate a whole novel which contains different elements by different authors with a strong logic.

10. Conclusion

This project start from September 2021, we have gone through data collection, trained neural network ourselves, tried to use GPT-2 and finally we found that GPT-3 is better. We also found that with the help of GPT-3, based on iteration idea and fine-tuned model, we are able to generate a novel written by AI. Due to the time limit, we didn't finish the whole novel. Noticed that to complete this job, it can not be purely generated by AI. Every text generation will give totally different sentences and that will need human to check whether every sentences are reasonable. Also if the whole job is finishing under the supervision of a novelist, it will make the whole structure and content more reasonable. Related introduction, codes and results can be found at my Github website: https://github.com/LeqingXu1997/NLP-for-novel-writing

11. Acknowledgements

I would like to say thanks to respectful professor Okan Ersoy for giving me such a good opportunity, also in this project, without professor Ersoy's supervision and suggestion, I cannot make any progress. In this project, I went through many difficulties and I learned a lot in this project. After graduated from Purdue ECE, I will continue diving into Machine Learning and NLP fields, always learn more with the attitude of a newborn.

RESULTS AND EVALUATION

In [145…

```
act_list = longer_summary.split("ACT ")
print(longer_summary)
print(len(act_list))
```

```
ACT ONE:

The opening scenes show Abe at home, an old man looking back on his life with a mixture of pride and nostalgia. He's ready fo
r one final adventure, and he's determined to make it count. He assembles a group of skilled hunters, including his protégé J
ohn, to set out into the wilderness to find the largest herd of buffalo they can.

ACT TWO:

The hunters journey across the rugged terrain of the old American West, encountering various obstacles, including treacherous
weather conditions, dangerous wild animals, and native tribes, each with their own agenda and motivations. Along the way, ten
sions rise among the hunters, leading to violent outbursts and confrontations that challenge their bond. They continue their
pursuit of the buffalo, facing fierce competition from rival hunting parties and clashes with the native tribes whose lands t
hey are encroaching upon.

ACT THREE:

As the group continues their hunt, the buffalo become harder to find, forcing them to push the limits of their skills and bra
very. They finally locate a large herd, and the hunters engage in a fierce and dangerous hunt, risking their lives to bring d
own the majestic beasts. Despite the danger, the thrill of the hunt motivates them, and they continue to pursue their quarry.

ACT FOUR:

The climax of the story is a final epic hunt where the hunters take on thousands of stampeding buffalo, caught up in the fren
zy of the hunt. However, the cost of the hunt is high, and the hunters must confront the harsh reality of their actions and c
ome to terms with the price they've paid for their glory. Throughout this portion of the film, themes of Manifest Destiny and
the brutal nature of buffalo hunting are explored in a nuanced manner.
```

pursuit of the buffalo, facing fierce competition from rival hunting parties and clashes with the native tribes whose lands t
hey are encroaching upon.

ACT THREE:

As the group continues their hunt, the buffalo become harder to find, forcing them to push the limits of their skills and bra
very. They finally locate a large herd, and the hunters engage in a fierce and dangerous hunt, risking their lives to bring d
own the majestic beasts. Despite the danger, the thrill of the hunt motivates them, and they continue to pursue their quarry.

ACT FOUR:

The climax of the story is a final epic hunt where the hunters take on thousands of stampeding buffalo, caught up in the fren
zy of the hunt. However, the cost of the hunt is high, and the hunters must confront the harsh reality of their actions and c
ome to terms with the price they've paid for their glory. Throughout this portion of the film, themes of Manifest Destiny and
the brutal nature of buffalo hunting are explored in a nuanced manner.

ACT FIVE:

The film concludes with Abe, John, Sarah, and the other hunters returning home, facing the consequences of their actions. The
native tribes are outraged by their encroachment, and the hunters must live with the guilt of taking so many lives for their
own gain. The characters are diverse, and their relationships and interactions are explored in depth. End of the Wild is a ra
w and unapologetic tale of the old American West, highlighting the brutality and violence inherent in the Manifest Destiny, a
nd a reminder of the cost of expansionism.
6

10

Out[179… "Abe is a grizzled veteran of the wild American frontier. He's spent his entire life hunting buffalo and exploring the rugge
d wilderness, and he's seen more than his share of danger and adventure. Over the years, he's taken on countless apprentices
and passed down his knowledge of the land and the animals to the next generation. \n\nAbe is a man of few words, preferring
to let his actions speak for themselves. He's known for his calm demeanor in the face of danger, and his ability to think on
his feet in the most challenging of circumstances. Despite his advancing age, Abe is still in excellent physical shape, and
his eyesight and aim are as sharp as ever. \n\nAs the story unfolds, we learn that Abe is seeking one final adventure becaus
e he knows he doesn't have much time left. He's facing his own mortality, and he wants to leave behind a lasting legacy by l
eading his protégé, John, and a group of talented hunters on a legendary hunt. Abe is driven by a deep sense of pride and id
entity, and he sees himself as the embodiment of the wild frontier. \n\nHowever, as the group encounters a series of obstacl
es and setbacks, we begin to see a more vulnerable side of Abe. He's haunted by memories of past hunts, and he struggles wit
h the realization that his way of life is slowly disappearing. Beneath his gruff exterior, Abe is a man who's deeply connect
ed to the land and its creatures, and he's aware that his own livelihood has contributed to the destruction of the wild buff
alo herds. \n\nDespite these inner conflicts, Abe remains a formidable force throughout the film. He's fiercely protective o
f his protégé and his group, and he's willing to risk everything to bring down the biggest and strongest buffalo he can fin
d. In the end, Abe is forced to confront the reality of his actions, and he realizes that the price of his glory has been fa
r too high. \n\nAbe's journey is a complex and nuanced exploration of the American frontier and the people who inhabited it.
Through his character, we see the bravery and tenacity required to survive in a harsh and unforgiving land, as well as the m
oral conflicts and contradictions that arise from the pursuit of Manifest Destiny. In the end, Abe's legacy is a testament t
o the enduring spirit of adventure that continues to inspire us all."

CONCLUSION

- Strory writing with chatgpt has been successfully running and providing the better output without loss of data.
- In that project ,I have used natural language processing model for make a good story based on given dataset.
- In this case python has helped to display the output in good manner and handling data is so efficiency.

Thank you