**1. Introduction**

* **Objective**: Explain the purpose of the project – to create a storytelling AI system capable of generating coherent and creative stories.
* **Scope**: Describe the system workflow, from dataset generation to story generation using embeddings.
* **Overview**: Briefly mention the tools and technologies used (e.g., Ollama, OpenAI API, JSON datasets).

1:Introduction.

**Objective**: The goal is to leverage generative AI models to create a system capable of taking user prompts and transforming them into rich, narrative-driven outputs. By combining advanced AI techniques, the system ensures that the generated stories are contextually relevant and tailored to user inputs while maintaining creativity and flow. This project emphasizes creating a storytelling platform that enhances the user experience by seamlessly integrating multiple AI models for high-quality story generation

**SCOPE**:The project workflow involves a multi-stage process, beginning with dataset generation and passing the dataset to a storytelling AI that dynamically creates new stories based on user prompts. The scope of the project includes:

1. **Dataset Generation**: Using locally installed Ollama models to produce an initial dataset of diverse stories. Multiple AI agents with varied configurations and prompts are employed to ensure the stories reflect a range of creative styles and themes. Different tests were done by using diffent prompts to get the most out of the model. A more detailed explanation will be done in the system architecture on the experiments to create the dataset
2. **Dataset Structuring**: The generated stories are organized into a structured format (JSON), making them accessible and reusable for downstream tasks which are for finetuning the OpenAI model and for testing phases the image generation model.
3. **Story Generation**: Utilizing embeddings derived from the dataset and integrating OpenAI’s generative models, the system creates new stories by referencing prior narratives. This step ensures that the generated content is both coherent and contextually linked to the user's input and making sure the model is tailored to create story’s and not used for other functions. Different testes where done to ensure that the AI system only creates story’s and nothing else
4. **Comparative Analysis**: Comparing outputs from different AI agents to select the best stories in terms of creativity, coherence, and relevance, to the story title created by one of the agents.

**2. Design Choices for Generative Models**

* **Ollama for Dataset Generation**

Dataset Creation: As for the creation of the dataset which will be used for finetuning the model ollama3.2 3billion parameters was chosen. Using a locally installed model has the benefits of privacy but it gave a huge understanding how to use Ollama on PyCharm and how different prompts can significantly influence the model. The stories could have easily been created using ChatGPT with a simple prompt but this method helped have a better understanding in prompt design and how to have a better understanding on how prompts effect outcomes.

Also a localy installed model meant that I could create the sotry dataset in places where I had limited connectivity which was the case since I work in a particular field and access to the internet is limited and finally a vast handon experience was learned.

For the dataset creation the idea was used to utilize AI agents but for our case we would have multiple ai agents which all has a significant role in creating the title and story. In total 12 AI agents are being used to create the dataset. The number is high because for this assignment an experiment was done two experiments,

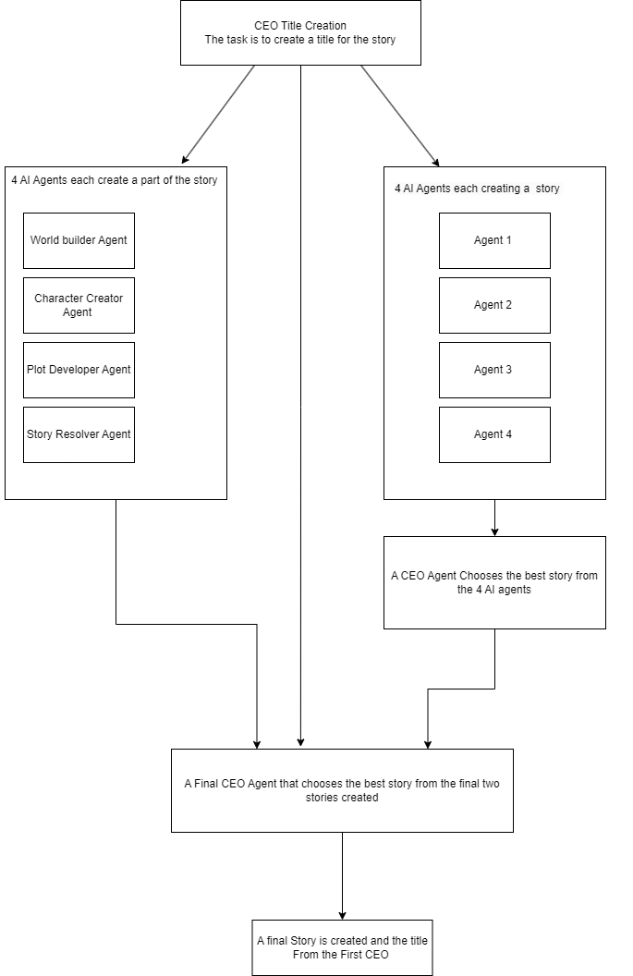
1. Split the story generation between for AI agents, and each ai agent had a specific task, such as worldbuilder, character creator, plot developer and story resolver. Each AI agent had to build a specific part of the story.
2. The other experiment was that different four AI agents would be utilized each with the task to create a story based on the title. Finally, another AI agent was delegated to choose the best story from
3. The final experiment was to have a finally AI agent acting as the CEO to choose the best story between the ai agent where the task was split up or the AI agents doing all the work on their own.

**OpenAI API for Story Generation**

Once the story json file has been created the next task to to utilize the file and using embeddings create a new story that the user has given inside the prompt. The Idea for this model is to use OPENAI gpt 3.5 turbo to do this task and using the locally run webui the user can insert the prompt and see the generated reply form the model. The OPENAI model excel’s at producing human-like narratives with minimal effort, making them ideal for storytelling tasks. The temperature for the model could also be changed making finetuning more easily. The main reason why the openai model was selected for it’s ease to allow incremental inputs which our case is the custom dataset and finally is the request time. Since hardware is a limited issue utilizing gpt 3.5 would drastically increase the generated story response time.

**3: System Architecture**

**Overall System Workflow**



**Data Generation**: As discussed in the in chapter 2 there are in total 12 AI agents that have each carefully tasked into contributing to create a final story and title.

The first AI Agent is called the Creative CEO and the task of this agent is to create a compelling story title from a randomly selected theme choice. The list of theme choices is a list of 36 different themes varying from space exploration to psychological thrillers and parallel universe. The CEO title creator has a very specific prompt and set of guidelines when creating the title. The first guideline is that a set of words can not be used when creating the title and this is done because in testing phase it was seen that the model uses a lot of similar words when creating the story titles such as Echoes of, The Last, Beyond the. It was seen that most titles started this way hence in the prompt it was guided not to use titles starting with these phrases.

The second prompt was that the titles need to be limited to not more than 7 words, ideally between 3-7 words max. The CEO was instructed to create strong emotion and vivid story titles and finally since a large dataset was being created the model needed to know all the story titles that were created as not to have any duplicated story titles, therefore each time a title is created it is stored in an array that the model has access to each time it tries to generate a title. This has been implemented after seeing that at around 50 story titles there where roughly each time 5 story titles that are the same.

**The next four AI agents** work collaboratively to create a complete and cohesive story based on a title generated by the CEO Story Title Generator. This approach divides the storytelling process into four specialized tasks, allowing each AI agent to focus on a specific aspect of narrative development, ensuring the final product is both creative and coherent.

1. World Builder: The first AI agent, nicknamed *World Builder*, is tasked with creating the setting and atmosphere for the story. This agent focuses on crafting a vivid and immersive environment that aligns with the title. By dedicating attention to world-building, the agent lays the foundation for the story’s tone and mood in a concise, 25-word description.
2. Character Creator: The second AI agent, *Character Creator*, is responsible for developing the characters who drive the story forward. This agent ensures that the characters are unique, relatable, and gender-neutral to avoid bias and enhance inclusivity. In exactly 25 words, this agent introduces the characters, emphasizing their personalities and conflicts, contributing to the emotional depth of the narrative.
3. Plot Developer: The third AI agent, *Plot Developer*, handles the main action of the story. This includes defining the central conflict and key events that propel the narrative. By focusing on the storyline's core dynamics, the agent creates tension and intrigue in 25 words, ensuring that the narrative stays engaging and purposeful.
4. Story Resolver: The fourth and final AI agent, *Story Resolver*, resolves the conflicts and concludes the story. This agent’s role is to tie up loose ends, delivering a meaningful resolution that aligns with the preceding narrative and the title’s theme. Its concise 25-word conclusion ensures the story feels complete and satisfying.

Together, these four AI agents form a collaborative framework that efficiently distributes the workload, ensuring that every aspect of the story is thoughtfully developed. Each agent’s specialized focus enhances the narrative's overall quality and creativity, creating a seamless and engaging storytelling experience.

**The next set of four AI agents** where designed to create whole stories independently, each with unique prompts yet with the same overarching goal: crafting vivid, imaginative narratives based on the given title and theme. Each agent operates autonomously to deliver a complete story, focusing on engaging world-building, gender-neutral characters, and a cohesive narrative arc. These agents were the most challenging to train due to their dual responsibility of producing a self-contained story while adhering strictly to the goal of generating only the story itself, with no additional explanations or justifications.

To achieve this, the prompts were meticulously refined to eliminate ambiguity, ensuring the agents stayed focused on storytelling without veering into analysis or reasoning. Each prompt was tailored to guide the agents in creating immersive settings, rich descriptions, and emotionally resonant character actions. The iterative refinement process was key to resolving early issues, such as overly verbose outputs or narratives lacking depth, by enforcing constraints like word count and a single-paragraph structure.

A critical feature of these agents is their adaptability to diverse themes. Regardless of whether the theme is "space exploration," "ancient myths," or "urban legends," the agents consistently generate stories that align with the intended mood and style of the theme. This adaptability ensures a wide range of creative outputs that remain true to the prompts’ directives.

Additionally, the prompts were designed to encourage variety across the agents. By focusing on different aspects of storytelling—such as emotional depth, narrative flow, or descriptive imagery—the agents avoid redundancy, producing unique interpretations of the same title and theme. This ensures that while the agents share a common goal, each story feels distinct, showcasing the creative breadth of the AI.

Finally, the process of training these agents highlighted the importance of balance: maintaining creative freedom while ensuring structure and coherence. The result is a team of storytelling agents capable of delivering rich, compelling narratives tailored to specific themes and titles, offering a fresh perspective with every story they create.

To explain the role of the CEO in choosing the best story from the four independently created by the AI agents, you can write something like this:

The **CEO Story Selector** represents an expert decision-making AI tasked with evaluating and selecting the best story among four independently crafted narratives by AI agents. This AI embodies the critical perspective of a publishing CEO, combining expertise in storytelling with a focus on market viability and creative excellence. Its purpose is to ensure that the chosen story aligns seamlessly with the provided title, showcasing the highest quality and narrative coherence.

The next AI agent is another CEO choosing agent that selects the best story from the 4 individual AI agents. This selection process is designed to simulate the real-world decision-making of a publishing house CEO. The CEO evaluates the four stories based solely on their alignment with the given title, ensuring that the chosen narrative is the most compelling, imaginative, and thematically appropriate. By focusing exclusively on the story quality and its connection to the title, the CEO AI avoids bias and guarantees that the most impactful story is selected.

The **decision criteria** include:

1. **Relevance to the Title**: How well the story reflects the essence and mood of the title.
2. **Creativity and Originality**: The uniqueness of the story’s concept, world, characters, and narrative style.
3. **Narrative Quality**: The coherence, structure, and emotional resonance of the story.
4. **Thematic Consistency**: How well the story adheres to the theme implied by the title.
5. **Engagement**: The ability of the story to captivate and hold the reader’s attention.

The prompt for the CEO AI is specifically structured to ensure impartiality and focus. By requesting the AI to consider only the title and the stories themselves, extraneous factors are excluded, enabling an objective and fair evaluation.

This final layer of decision-making adds a level of quality control to the collaborative efforts of the storytelling agents, ensuring that the output meets the highest standards of creativity and narrative integrity. Through this process, the CEO AI demonstrates how technology can emulate professional judgment in creative industries.

**The Final CEO Decision** represents the ultimate stage in the story selection process, where the CEO AI compares the collaboratively created story (the team effort) against the best story chosen from the four independently crafted narratives. This ensures that the most compelling and impactful story is selected for publication, combining the strengths of collaborative creativity with the individuality of independent storytelling.

In this phase, the CEO AI functions as a discerning evaluator with expertise in identifying stories that align closely with the provided title and theme while resonating deeply with the target audience. The final decision process involves assessing two stories on key parameters such as:

1. **Alignment with the Title and Theme**: How well each story reflects the essence of the given title and the nuances of the selected theme.
2. **Narrative Strength**: The coherence, structure, and emotional resonance of the story.
3. **Creative Merit**: The originality and imaginative quality of the world-building, characters, and plot.
4. **Market Potential**: Which story is likely to captivate and engage readers, ensuring commercial success.

The CEO AI evaluates these two stories without bias or external influence, considering only the intrinsic qualities of each narrative. By following a structured prompt, the CEO AI ensures impartiality and focus, delivering a single, definitive output: the best story, ready for publication.

This final decision-making layer underscores the collaborative synergy and creative potential of AI storytelling while ensuring that the chosen story meets professional standards of excellence and audience appeal. It highlights the ability of AI to not only generate content but also emulate expert-level judgment in the creative process.

This process is repeated for every story that is created but first although inside the prompts there are multiple points on keeping the story positive and not subjective another part of the code handles just this. A function is added that monitors how positive a story is and how objective or subjective it is. This is done using the library from textblob. The library works by helping the model learn how to understand human language. The text is anlalyzed for tone, meaning and structure. In simple a text can be anlayzed to interpret if it is positive , negative or if its based on facts or opinions. The algorithim from textblob splits the results into polarity and subjectivity. Polarity is shown in a value from -1 to 1 -1 being very negative and 1 being very positive. An example is if there are words “good” the algorithim will interpret that as positive whilst seeing the word bad will interpret that as negative. And the same applies for being objective and being subject. The library from textblob relies on a predefined dataset which includes a dataset of common English words and their associated **polarity** and **subjectivity** scores.

For our case what we are doing is that at the end when the final CEO choose the best story between the single agent and the team effort, the story is passed through a final test to check how subjective it is and how positive. The idea here was to change the prompts for each AI agent if the story falls within a certain threshold and these are as follows.

Polarity: for polarity we have two thresholds

1. polarity < 0.1 (Too Neutral or Negative)

A polarity below 0.1 suggests that the story lacks emotional warmth or has a slightly negative or neutral tone. This threshold prompts changes to make the story more uplifting and engaging, aligning with the goal of creating positive narratives.

1. polarity > 0.5 (Excessively Positive)

A polarity above 0.5 indicates that the story is very positive. While positivity is desirable, excessively positive tones can feel unrealistic or overly simplistic.

Subjectivity we have two thresholds

1. subjectivity > 0.7 (Highly Subjective)

Subjectivity above 0.7 indicates that the story is heavily opinionated or based on emotions rather than concrete details. This can make the story feel unbalanced or lacking in depth. This threshold encourages adding clarity and factual details to balance the emotional or imaginative aspects of the story. This ensures relatability and narrative grounding.

1. subjectivity < 0.4 (Too Objective)

Subjectivity below 0.4 suggests the story is overly factual and may lack emotional depth or creative descriptions. This can make the narrative feel clinical or unengaging. At this threshold, the system prompts more creativity and emotional depth to enhance the story's appeal and engagement.

The chosen thresholds for **polarity** and **subjectivity** in the code reflect a balanced approach to storytelling, ensuring narratives are engaging, realistic, and emotionally resonant. When **polarity is below 0.1**, the story may feel neutral or slightly negative, prompting the prompt:

* **"Please ensure the story has a more hopeful and uplifting tone, with positive elements that inspire and engage readers."**

Conversely, when **polarity is above 0.5**, it suggests excessive positivity, which risks making the story feel unrealistic. The associated prompt encourages balance:

* **"Maintain a positive tone but incorporate realism and emotional depth to make the story relatable and grounded."**

For **subjectivity**, scores above 0.7 indicate overly opinionated content, leading to the prompt:

* **"Balance creativity with concrete details and narrative clarity. Ensure the story is relatable and avoids overly opinionated elements."**

On the other hand, subjectivity below 0.4 signals a lack of emotional depth, triggering the prompt:

* **"Add more emotional depth and creative descriptions to make the story engaging and imaginative."**

The algorithm works by detecting whether the generated story's sentiment falls within predefined thresholds for **polarity** and **subjectivity**. If a story is too neutral or negative (polarity < 0.1), excessively positive (polarity > 0.5), highly subjective (subjectivity > 0.7), or overly objective (subjectivity < 0.4), specific prompts are dynamically added to guide the AI agents. For example, the prompts include **"Please ensure the story has a more hopeful and uplifting tone, with positive elements that inspire and engage readers"** or **"Add more emotional depth and creative descriptions to make the story engaging and imaginative."** These adjusted prompts are incorporated into both the team effort and single-agent story generation processes. The story is then regenerated and re-evaluated for compliance with the thresholds. This iterative process continues until the generated story meets the desired emotional and narrative criteria.

* 1. Detail how Ollama and OpenAI were integrated into the pipeline.

Once the dataset which in our case is the json file is created the next task is to use the data set as training data for gpt 3.5 turbo to create its story’s. The model works by using embedings which are numerical representation of data that capture the semantic meaning of words, sentences, or even larger text structures. In the context of GPT-3.5 Turbo, embeddings are the foundation for understanding and processing the training data to create coherent and contextually appropriate stories.

The JSON dataset serves as the **training corpus**. It contains structured examples of story elements, such as settings, characters, plot points, and conclusions, organized in a way that GPT-3.5 Turbo can learn patterns and relationships

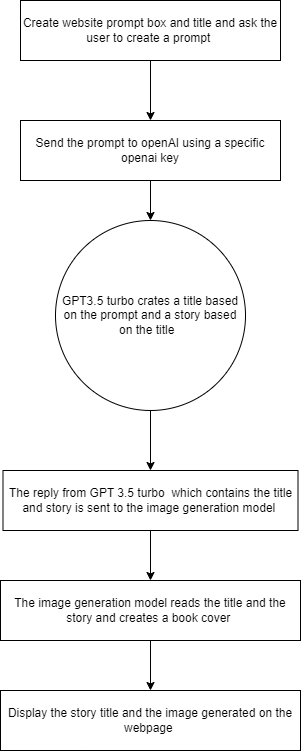
Code explanation to integrate GPT 3.5 turbo and the json/traing file and final webpage for the user to enter a prompt and display the image

This code demonstrates a workflow that combines OpenAI’s **GPT-3.5-turbo** model, a **JSON-based story archive**, and a **Streamlit-based web interface** to create a dynamic, user-friendly story generation platform. The application begins by presenting the user with an intuitive prompt field on a locally hosted webpage, allowing them to input specific story parameters such as themes, characters, or settings. Upon submission, the user’s input is sent—along with contextual references from previously generated stories stored in stories.json—to OpenAI's GPT-3.5 model. The model generates a creative response, producing a **unique story title and approximately 500 words of narrative content**, adhering to professional storytelling guidelines such as inclusivity, emotional resonance, and avoidance of harmful stereotypes.

The generated story is saved into the local JSON archive, ensuring that future prompts can build on the existing tone and style while avoiding repetition of previous content. This iterative approach helps maintain creativity while reinforcing consistency and ethical storytelling. Additionally, the interface showcases not only the story but also supports the potential for integration with **image generation models**, allowing users to create corresponding illustrations or book covers.

A unique feature of the platform is its ability to calculate and display word counts, character counts, and split longer stories into sections or pages. This is useful for applications such as e-books or serialized storytelling. The program also incorporates error handling to ensure that invalid prompts or API errors are gracefully managed, improving the user experience.

By combining **machine learning** with a **streamlined user interface**, this code represents an efficient and extensible framework for creative content generation, offering opportunities for further enhancements such as personalization, additional media integration, or advanced narrative controls.



Structure/architecture of the code integrationg local web-base and gpt3.5 turbo

The code’s architecture is organized into distinct yet interconnected layers, ensuring clarity and maintainability.

1: At the top level, Streamlit provides a clean user interface where the user can input a prompt and trigger the story generation process.

2: Beneath this interface, a controller-like function (handle\_story\_generation()) orchestrates the workflow by validating input, showing status indicators, and calling the model interaction function (generate\_story()). This model function interfaces with OpenAI’s GPT-3.5-turbo, sending the user’s prompt and receiving a structured response containing both a title and narrative content

3: Once generated, the story is appended to a JSON-based storage layer via the load\_stories() and save\_stories() functions, maintaining a record of all previous creations.

4:Finally, the updated story archive and any accompanying images are displayed back on the webpage, creating a polished, end-to-end solution that gracefully integrates user input, AI-driven content generation, persistent data storage, and a visually appealing frontend.

**Challenges encountered (e.g., coherence, bias) and how they were addressed.**

A number of challenges were found when trying to create this assignment and specifically integrating all the models in order to work hand in hand and produce a final story and image for the user based on the prompt.

Challenges found

1.Repation of story titles

2. Balancing Creativity and Structure

3.creting a vast story dataset on different themes

4. Controlling Output Quality and Style

5. Reducing Bias and Improving Inclusivity

8. Creating the AI agents

9. prompts for the AI agents to do specific tasks

10 gpt 3.5 creating large word count titiles

11. repeated story using the same prompt using gpt 3.5 turbo (add created story to json file)

13 adding the sentiment to the prompt to make sure it follows ethical guidelines

**1:Reptation of story titles**

The first instance at creating the story dataset was to create a simple code to prompt ollama to create story titles. With the help of chat GPT01 a number of prompts where tested which gave promising results. Such prompts were

1.Suggest a short, creative story title that feels intriguing

2. Offer a compact yet compelling title that makes a reader eager to explore its narrative

3. Create a one-line story title that sparks the imagination without revealing too much

Ollama was able to create compelling four-to-seven-word titles but the issue lied when a for loop was used to create 100 different story titles. It could be seen that ollama repeated a number of story titles with the exact words and also it was noticed that ollama likes to create story titles starting with “echoes of”,” The last of”, “Beyond the” which caused a bit of a bias to how gpt 3.5 was to create its title. Therefore, to fix the first issue which was repeated titles names an array was created that stores all of the story titles and this array was passed in the prompt so ollama can see the titles created and then create a new different one. As for the second issue which was that ollama likes to use certain words when creating the titles inside the prompt it was instructed not to story title patterns starting with those words. Both of these issues where tackled but a new one emerged which was the theme for the story. Since we wanted a vast range of story titles ollama was not creating any murder story titles or any alien invasions, it was sticking to the traditional story titles where everyone lives happy ever after. Therefore, once again using ChatGPT I asked the question to give me 50 different story title themes that ollama can use. Like this having a random theme choice was selected and llama was promoted to create the story titles based on chosen theme choice

**2. Balancing Creativity and Structure**

Having a well structured dataset and dataset creating was key when creating this project as one of the issues that was found is when passing story’s or titles to different AI agents. It was found that certain AI agents where not taking into considerations the guild lines and they were still giving reasons for creating particular story’s which had a big impact at the final result when saving the stories in the dataset since we needed a clear structure. After multiple attempts the best ratio was to make the guideline to create only the story or the title in the first line of the prompts and make them in capital letters and this seemed to have solved the issue. Another point to structure is that each AI agent if it is creating a title or a story it was instructed on the number of words it had to write for example the collaborative AI agents each needed to create a 100 word contribution to the story but a new challenge was found that 1 out of 10 the celebrative agent or single agent when creating the story would return that it couldn’t create a story either of 500 words or of 100 words exact. Therefore, a min max limit had to be implemented and this solved the issue.

In order to push the AI agents being creative but having constrains such as don’t use certain foul language or prohibits stereotypes, try and use gender neutral names helped the AI agents creating story’s with abide with the structure of the assignment but still being imaginative. The major help in creative for the AI agetents where the themes chosen. The themes where carefully chosen from a vast suggestion by chat GPT

**Creative Freedom Within Constraints:**  
Despite the stringent guidelines, the model is still encouraged to be imaginative. For example, the prompt suggests themes like “a city powered by music” or “friendly robots learning human emotions,” which leave plenty of room for the model to invent unique worlds, characters, and conflicts. By limiting certain patterns (e.g., avoiding clichéd title formulas) and encouraging vivid imagery, the model can explore creative territory while staying within the boundaries of the instructions.

3.**creting a vast story dataset on different themes**

Developing this dataset for fine-tuning GPT-3.5 posed several challenges, requiring careful consideration and innovative solutions to ensure quality, diversity, and ethical integrity. The primary challenge was balancing creativity and structure. While the goal was to generate imaginative and original stories, maintaining a coherent narrative framework, including setting, character development, and plot resolution, required meticulous planning and coordination among the story creation agents. Another significant hurdle was ensuring ethical compliance and inclusivity. This involved designing prompts and workflows that strictly avoided harmful stereotypes, offensive language, and discriminatory tropes, which demanded thorough testing and validation of the generated outputs.

Additionally, achieving thematic diversity while avoiding redundancy in the stories required the development of a wide array of unique, engaging themes and a robust system to prevent duplication. The collaborative approach, involving multiple agents contributing to different aspects of the story, introduced complexity in harmonizing their outputs into a cohesive narrative. Finally, selecting the best story from multiple versions necessitated a reliable and objective evaluation process based on criteria such as creativity, coherence, and emotional impact. These challenges collectively required a combination of technical expertise, creative vision, and ethical diligence, making the creation of this dataset a multifaceted and rewarding endeavor. Furthermore, this entire process was carried out using local hardware, which added to the complexity and required considerable time and resource management to handle the computational demands efficiently.

4. **Controlling Output Quality and Style**

When addressing the point of "Controlling Output Quality and Style," it's essential to emphasize the strategies and techniques employed to ensure the generated stories meet high standards of creativity, coherence, and readability. Maintaining output quality and adhering to a consistent style required defining clear guidelines for story generation, including the use of inclusive language, ethical considerations, and narrative structure. The prompts for generating story elements were meticulously crafted to set specific expectations, such as word limits, vivid descriptions, and avoidance of harmful stereotypes.

Additionally, quality control was achieved through a multi-layered evaluation process. Generated outputs were reviewed based on criteria like thematic alignment, logical progression, and emotional resonance. Collaborative storytelling, where different agents contributed to various aspects of the narrative (e.g., world-building, character creation, plot development, and resolution), introduced the need for harmonizing these parts into a cohesive whole. To ensure consistency, a final evaluation process was implemented, involving multiple CEO agents who acted as evaluators to assess the stories. These CEO agents rigorously reviewed the narratives, selecting the best outputs based on creativity, coherence, adherence to stylistic guidelines, and alignment with the assigned themes.

Furthermore, style consistency was achieved by incorporating specific instructions in the prompts to guide the tone, level of detail, and language complexity. By iteratively refining the generation processes and using robust review mechanisms, including feedback from the CEO agents, the project successfully maintained high standards for both output quality and narrative style. These efforts were critical in creating a polished dataset suitable for fine-tuning GPT-3.5 and achieving optimal performance in creative applications.

5. **Reducing Bias and Improving Inclusivity**

Reducing bias and improving inclusivity posed a significant challenge during the creation of the dataset, as it required deliberate and methodical approaches to ensure fairness, representation, and respect in the generated content. One of the main challenges was identifying and addressing potential biases that could arise from pre-existing patterns in language models or data. To counteract this, prompts were carefully designed to emphasize the use of inclusive language, avoid harmful stereotypes, and respect cultural, gender, and social diversity. This included explicit guidelines for creating gender-neutral characters, avoiding discriminatory tropes, and ensuring that the stories reflected diverse perspectives without perpetuating harmful assumptions.

Another challenge was ensuring that inclusivity extended to the themes and narratives. A conscious effort was made to incorporate diverse settings, characters, and plots that celebrated differences and fostered empathy and understanding. Rigorous evaluation processes were implemented to review the generated stories, focusing on eliminating biased content while maintaining creativity and coherence. Multiple agents, including "CEO" evaluators, assessed the stories to ensure alignment with inclusivity and ethical guidelines, further reducing the likelihood of biased outputs.

Additionally, iterative testing and validation of the dataset were conducted to refine the prompts and generation mechanisms, ensuring that inclusivity was not only a theoretical goal but also a practical outcome. These efforts required a balance between promoting diversity and maintaining the creative freedom of the storytelling process. Addressing this challenge was crucial in creating a dataset that upholds ethical standards and represents a step forward in reducing bias in AI-generated content while fostering inclusivity across themes and narratives.

6.Creating the AI agents

One of the major challenges in this project was designing AI agents to collaborate effectively in generating cohesive and compelling stories. Inspired by articles predicting a future with specialized AI agents for specific tasks, the initial approach used two agents: one to create the story and another to verify its alignment with the title and theme. However, this setup felt limited, leading to the idea of splitting the storytelling process among multiple agents. The first attempt involved three agents—one for the introduction, another for the body, and a third for the conclusion—but this often resulted in fragmented and inconsistent narratives.

A refined system was then developed, with four specialized agents: the **World Builder** to set the scene, the **Character Creator** to develop inclusive characters, the **Plot Developer** to advance the main action, and the **Story Resolver** to provide a satisfying conclusion. While this improved narrative depth and diversity, ensuring seamless collaboration and adherence to user prompts required iterative refinements.

An important experiment compared teamwork stories created by multiple agents with those generated independently by single agents. In seven out of ten trials, the teamwork approach produced better narratives. However, significant challenges remained in guiding the agents to consistently meet the project’s requirements. For instance, agents often provided explanations for their contributions—useful for the final CEO decision but outside the scope during story generation. Additionally, ensuring agents adhered to specific instructions, such as writing precisely 500 words, proved difficult. This issue led to the implementation of thresholds to manage word count variability and refine prompts for replicability. These challenges highlighted the complexity of coordinating multiple AI agents while maintaining quality and consistency.