

# Date: 2024-04-11

## Last week's action items

- **Analyzing the Dataset:** Continued the in-depth analysis of the dataset, focusing on extracting insights related to character diversity, gender representation, and potential biases within the AI-generated stories.
- **Lead Character Identification Using NetworkX:** Utilized NetworkX in Python to create and analyze graphs depicting the interactions between characters. This approach helped in identifying the primary characters within the narratives.
- **Character Type Classification:** Undertook an exploration to categorize characters into types such as humans, animals, or toys. This classification aimed to enhance understanding of the variety and representation of different entities in the stories.

## Issues

- **Complexity in Mapping Results to Prompts:** Encountered challenges in directly associating the analytical results with the specific prompts that generated the stories. This complexity arises from the need to systematically link character dynamics and representation with the initial conditions set by the prompts.
- **Identifying Bias with Outlined Rules:** Faced difficulties in applying the predefined rules systematically across the dataset to identify biases effectively. The complexity of narratives and varied character representations adds to the challenge.

## Action Items:

- **Result-Prompt Mapping:** Develop a methodical approach to map the analytical findings back to the respective prompts. This step is crucial for drawing concrete conclusions regarding the influence of prompts on character diversity, gender representation, and bias.
- **Complete Bias Analysis:** Finalize the analysis to detect bias within the narratives, strictly adhering to the outlined rules. This completion will involve refining the analytical methods to ensure accurate identification of biases.
- **Collect Data from Gemini:** Initiate the collection of a new dataset from Gemini to complement the existing dataset. This additional data will provide a broader basis for analysis and comparison.
- **Analyze Gemini Dataset:** Perform a comprehensive analysis of the newly collected Gemini dataset. Compare and contrast the findings with those from the current dataset to identify patterns, similarities, or discrepancies in bias and representation.

- **Use of Transformer for Summarization:** Implement a transformer-based model for summarizing the narratives within the dataset. This summarization will aid in a quicker and more efficient review of stories, facilitating an enhanced analysis of themes, character roles, and potential biases.
- **Leveraging ChatGPT for Story Summaries and Bias Evaluation:** Utilize ChatGPT to generate concise summaries of the stories, providing a supplementary approach to identifying and evaluating bias within the narratives. This strategy will enhance our capacity to systematically assess and address biases in AI-generated content.

Date: 2024-04-04

#### Last week's action items

- **Character Identification using Named Entity Recognition (NER) in NLP:** Successfully implemented NER techniques to identify characters within the AI-generated stories. This process enabled a systematic extraction of character names and references, providing a foundation for further analysis.
- **Completion of Character and Gender Identification Modules:** Developed and finalized modules for identifying both the characters and their genders within the stories. These modules are crucial for analyzing the representation and diversity of characters, as well as for identifying potential biases in gender portrayal.

#### Issues

- **Ambiguity in Character Categorization:** Faced difficulties in categorizing characters that do not fit neatly into predefined categories, such as mythical creatures or characters with ambiguous characteristics.
- Very few characters in the narration of stories.

#### Action Items:

- **Further Dataset Analysis:** Continue the in-depth analysis of the dataset to uncover additional insights into character diversity, gender representation, and potential biases in the AI-generated stories.
- **Lead Character Identification via NetworkX Graphs:** Utilize NetworkX in Python to analyze character interaction graphs with the aim of pinpointing the primary characters within the narratives.
- **Character Type Classification:** Engage in an exploration to classify characters based on their nature, such as humans, animals, or toys, to understand the variety and representation of different entities in the stories.

Date: 2024-04-04

#### Last week's action items

- **Character Identification using Named Entity Recognition (NER) in NLP:** Successfully implemented NER techniques to identify characters within the AI-generated stories. This process enabled a systematic extraction of character names and references, providing a foundation for further analysis.
- **Completion of Character and Gender Identification Modules:** Developed and finalized modules for identifying both the characters and their genders within the stories. These modules are crucial for analyzing the representation and diversity of characters, as well as for identifying potential biases in gender portrayal.

### Issues



- **Ambiguity in Character Categorization:** Faced difficulties in categorizing characters that do not fit neatly into predefined categories, such as mythical creatures or characters with ambiguous characteristics.
- Very few characters in the narration of stories.

### Action Items:

- **Further Dataset Analysis:** Continue the in-depth analysis of the dataset to uncover additional insights into character diversity, gender representation, and potential biases in the AI-generated stories.
- **Lead Character Identification via NetworkX Graphs:** Utilize NetworkX in Python to analyze character interaction graphs with the aim of pinpointing the primary characters within the narratives.
- **Character Type Classification:** Engage in an exploration to classify characters based on their nature, such as humans, animals, or toys, to understand the variety and representation of different entities in the stories.

Date: 2024-03-28

### Last week's action items

- Used ChatGPT API with gpt-3.5-turbo-instruct engine and collected 11,000 records of data.
  - Dataset :  final\_dataset\_chatgpt.xlsx
  - The data collection was done using 2 categories of prompts: Age-specific prompts(1-18) and stereotypic prompts.
  - Outlined the approach to implement for analyzing bias in the AI-generated stories.
-  APPROACH

### Issues

- There is a token generation restriction for the ChatGPT API, limiting the number of responses that can be generated.

- Applying different approaches to large datasets can be challenging and resource-intensive.

#### **Action Items:**

- Explore character identification using Named Entity Recognition (NER) in NLP.
- Complete character identification and gender identification modules.
- Upload the Jupyter Notebook to GitHub for versioning and collaboration.

**Date: 2024-03-21**

#### **Last week's action items**

- Renewed the ChatGPT API subscription and explored alternative APIs for story generation and analysis.
- Conducted a thorough search for research papers and studies related to bias analysis in text generation, particularly focusing on methodologies applicable to children's stories.
- Refined and expanded the prompt collection further to encompass a broader range of biases and stereotypes.

#### **Issues**

- Need to find additional research papers to explore different methodologies for calculating bias.

#### **Action Items:**

- Use API and collect data for analysis.
- Document the approach—where to start and how to start.
- Analyze the data and explore NLP.

**Date: 2024-03-02**

#### **Last week's action items**

- 1. Extracted 585 records of stories with prompts mentioning age using the Gemini API.
- 2. Analyzed gender bias by identifying the lead character's gender in the stories.
- 3. Finalized the definition of bias in this project and listed out rules and approaches.
- Link : [IndependentStudy-outline.pdf](#)

#### **Issues**

- 1. ChatGPT API subscription expired, hindering further analysis and generation of stories.
- 2. Currently exploring various approaches to evaluate bias effectively.

- 3. Need to find additional research papers to explore different methodologies for calculating bias.

#### **Action Items:**

- Renew the ChatGPT API subscription or explore alternative APIs for story generation and analysis.
- Conduct a thorough search for research papers and studies related to bias analysis in text generation, particularly focusing on methodologies applicable to children's stories.
- Refine and expand the prompt collection further to encompass a broader range of biases and stereotypes.
- Conduct additional testing with the refined prompts to assess bias in the generated narratives comprehensively.
- Document findings, methodologies, and approaches for bias calculation for future reference and analysis, ensuring transparency and reproducibility in the project's workflow.

**Date: 2024-02-29**

#### **Last week's action items**

- 1. Conducted testing with age-specific prompts and noted trends around gender inclusion and age-appropriate content.
- 2. Conducted testing with age-specific prompts and noted trends around gender inclusion and age-appropriate content.
- 3. Explored Google Studio to understand and effectively utilize the Gemini API.
- 4. Compiled a list of stereotypes for one-line prompts and specified general prompts tailored to different age groups.
- 5. Started investigating methods to calculate bias in generated narratives.

#### **Issues**

- 1. Experienced significant timeout issues with the Gemini API, impacting testing and productivity.
- 2. Continued challenges with response time from certain APIs, affecting testing efficiency.
- 3. Difficulty in quantifying bias and determining appropriate metrics for evaluation.

#### **Action Items:**

- Address timeout issues with the Gemini API by reaching out to support or exploring alternative solutions.
- Refine and expand prompt collection to cover a wider range of biases and stereotypes.
- Conduct additional testing with refined prompts to assess bias in generated narratives.
- Continue exploring Google Studio to fully leverage the capabilities of the Gemini API.

- Document findings and methodologies for bias calculation for future reference and analysis.

**Date: 2024-02-22**

### **Last week's action items**

- 1. Tested Gemini, ChatGPT, and Claude AI with general prompts to generate stories for kids.
  - Compared outputs across LLMs for 10 prompts
  - Analyzed age-appropriateness, creativity, and consistency
- 2. Investigated APIs to generate stories for multiple prompts
  - ChatGPT, Claude, and Gemini offer paid APIs
  - Gemini provides a free API with usage limits
  - Bing Chat API had slow response times
- 3. Evaluated context tracking across tools
  - Only ChatGPT retains context across a single conversation
  - Others do not track context once a chat session expires
  - Gemini so not track the results of the previous prompts

### **Issues**

- 1. High response time from BingChat.
- 2. Lack of memory for previous prompts and results in Gemini, Claude AI, and BingChat.
- 3. Need to explore and understand the Google Studio to utilize the Gemini API effectively.
- 4. Define clear guidelines for assessing bias in written narratives.

### **Action items**

- Expand testing with age-specific prompts
  - Note trends around gender inclusion and age-appropriate content
  - ChatGPT: Generates stories with male and female characters regardless of the child's age.
  - Gemini: Begins generating stories with characters addressed as he/she for 7-year-old kids. Stories for kids up to 7 years old use imaginary or genderless characters like ladybugs, bunnies, trains, animals, etc.
  - Claude AI: Generates stories with male and female characters regardless of the child's age.
  - Gemini output adjusts based on specified target age
- Prompts Collection:
  - Compile a list of stereotypes to use as one-line prompts.
  - Specify general prompts tailored to different age groups.

- Investigating the utilization of the free Gemini API effectively by understanding the Google Studio platform.

**Date: 2024-02-15**

#### **Last week's action items**

- 1. Conducted a thorough review of the research paper.
  - Explored the process of generating stories from given prompts and compared them with original stories authored by humans. Raised questions regarding whether to focus on identifying bias and stereotypes or analyzing creative narration.
- 2. Discussed the approach for assessing bias in both visuals and written content. Considered whether bias should be evaluated separately for pictures and story writing.
  - Decided to work with the text content first.
- 3. Experimented with providing various prompts to the LLMs.
  - Observed that most LLMs tended to incorporate moral lessons into their story conclusions, even if they utilized stereotypes within the narratives. Explored different prompts, including those specifying the race of characters, which often resulted in LLMs incorporating racial elements and concluding with a moral.

#### **Issues**

- 1. Determining the primary focus of comparison between AI-generated stories and human-authored originals: bias and stereotypes versus creative narration.
- 2. Clarifying the methodology for assessing bias in both visuals and story content.
- 3. Understanding the prevalence of moral conclusions in AI-generated stories and their relationship with the use of stereotypes.

#### **Action items**

- 1. Further discussions and clarifications on the comparative analysis between AI-generated stories and human-authored originals. Decide whether the emphasis will be on identifying bias and stereotypes or evaluating creative narration.
- 2. Define clear guidelines for assessing bias in written narratives.
- 3. Continued experimentation with different prompts to understand how LLMs respond to various inputs and whether moral conclusions persist across different scenarios.