Gender Dynamics in the Marvel Cinematic Universe: A Comparative Analysis

Prakhar Gupta* Richa Muktibodh* Sahil Saraswat* Vartika*
{prakhar21550, richa21086, sahil21091, vartika21571}@iiitd.ac.in

Abstract

Movies are integral to modern-day human life and undoubtedly influence how we see and perceive the real world. Hence, films modelled on our society must not reinforce the prejudices and stereotypes of the real world and portray a fair and balanced view of society. For this, we aimed to analyse the differences in the representation of male and female characters in the Marvel movie series by examining different factors of the character dynamics and traits of each gender and the famous characters. We have thoroughly examined the differences to understand their general depiction in the Marvel Cinematic Universe (MCU). We have achieved this task by analysing primarily three tasks - sentiment analysis, emotion recognition, and Myers-Briggs personality analysis. Further, we have attempted to fine-tune a pre-trained model on our manually annotated MCU dataset, where we find a trend in general prediction for a given sentence. These methodologies have enabled us to understand the deep-rooted biases and stereotypes between genders and characters, moving to a more holistic representation of our society in movies.

1. Introduction

Movies serve as powerful reflections of society, as they mirror our biases and stereotypes while shaping our perceptions, values and attitudes. Movies are responsible for representing a diverse and inclusive view of society, free from reinforcing any unnecessary biases and prejudices. Movies must represent a balanced and honest view of society, free from biases which a collection of like-minded individuals can enforce. Gender representation is a critical segment of this narrative, and it is known that the prejudices of movies and social media have effects on society [6]. While we support the LBTQIA++ movement, most movies do not adequately represent these communities. Hence, we felt that analysing such limited data would only strengthen such prejudices. We limited ourselves to studying differences between males and females. We felt a growing need to critically examine the representational differences between males and females in a collection of related movies instead of a stand-alone movie. We considered various series for this task, but the Marvel movie series stood out from our search.

The Marvel Cinematic Universe (MCU) is a sprawling media franchise that has captivated audiences worldwide through its extensive series of superhero films and television shows. As a cultural phenomenon, the MCU offers a unique platform to explore various themes, narratives, and character dynamics. Among the many aspects that attract scholarly interest, the portrayal of different male and female characters in the MCU is of particular significance.

This report aims to delve into these dynamics by conducting a comparative study on the characteristics, emotions, and sentiments associated with characters in the MCU. Our approach involves analyzing character traits to understand how the universe's timeline influences their depiction, behaviour, and relationships.

To conduct this analysis, we employ several methodologies, including sentiment analysis, emotion analysis, and Myers-Briggs Personality analysis. These methods allow us to quantify and assess the differences in portrayal across the universe. Additionally, we fine-tune a pre-trained model on a manually annotated MCU dataset to predict gender based on textual data, providing further insights into trends and patterns.

This report aims to contribute to the broader understanding of character development in popular culture by exploring the depiction of character dynamics in the MCU. Our analysis provides a quantitative view of how the Marvel franchise's character traits, emotions, and interactions evolve. Through this examination, we can identify emerging trends and patterns that shape audience perceptions of superheroes and their relationships. These insights may inform future narrative choices within the MCU and in similar entertainment industries, guiding the portrayal of characters in new and meaningful ways.

2. Related Work

For the MCU dataset, we performed the annotations of dialogue-to-gender mapping for each movie ourselves. Due to this, analysis and work on the MCU dataset is a novel task. Hence, the results and data obtained on this methodology

*Equal contribution

are unique, and there are no baselines to compare our novel dataset. For this task, we tried to work with zero-shot sentiment analysis, zero-shot emotion recognition, and fine-tuned personality analysis for this dataset.

2.1. Sentiment Analysis

Hutto and Gilbert [5] have used qualitative and quantitative methods to produce and empirically validate a gold-standard sentiment lexicon attuned to microblog-like contexts. They combined these lexical features with consideration for five generalizable rules that embodied grammatical and syntactical conventions that humans used when expressing or emphasizing sentiment intensity. Their model, VADER, performed well in the social media domain.

2.2. Emotion Analysis

In their work on emotion recognition, Demszky and Movshovitz-Attias [4] curated a novel dataset which consisted of 27 emotion categories and the neutral category. It is the largest human-annotated dataset for emotions on text, with over 500,000 sentence-to-emotion data points. In this task, they have modelled the problem as a multi-emotion recognition task, where each text can be classified into multiple labels. On their novel dataset, they fine-tuned a BERT-base model by adding a dense output layer on top of the pre-trained model, with a sigmoid cross entropy loss function to support multi-label classification. Here, a sigmoid output is predicted for each emotion, and a threshold of 0.3 is applied before choosing the predicted emotions. Hierarchical clustering was performed on the 27 emotions, which mapped the emotions to 6 groups. They achieved an average F1-score of 0.46 on their setup.

2.3. Personality Analysis

Majumder and Poria [7] trained five different networks, all with the same architecture, to predict the Big Five Personality traits. Each network was a binary classifier that predicted the corresponding trait to be positive or negative. They used a CNN feature extractor to obtain n-gram feature vectors. These feature vectors were concatenated, and emotionally neutral sentences were discarded. The final classification was performed by a fully connected neural network with one hidden layer.

Sang and Mou [8] proposed to predict a movie character's MBTI personality based on the character's narratives. The challenges they faced were handling the long inputs(which were greater than 10k words on average), and taking into account both the descriptions and dialogues to predict the personality. Their model was an extension to the BERT classifier, to handle long and multi-view inputs. They observed better macro F1 scores by considering the description and dialogues rather than only the dialogue. Their f1 score could not cross 0.6 because of the dataset's imbalance and the task's complexity.

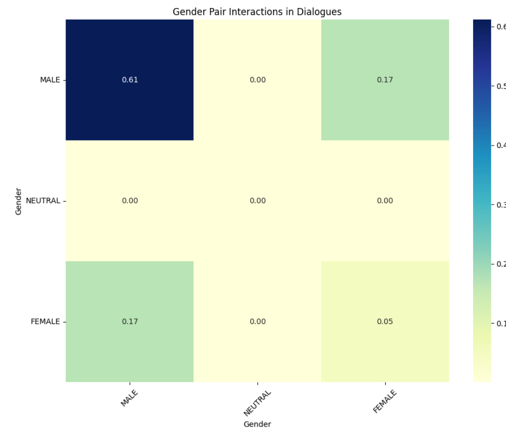


Figure 1. Percentage of Male and Female Character Dialogues as a Trend with movies

2.4. Gender Prediction

To benchmark our approach, we leveraged a pre-trained gender classification model. Specifically, we utilized the 'padmajabfrl/Gender-Classification' model from Hugging Face's repository. This model, developed for gender classification tasks, served as a baseline for evaluating our subsequent fine-tuning and customization efforts.

3. Dataset

To analyse character dynamics in the Marvel Cinematic Universe (MCU), we started by extracting character-dialogue mapping from publicly available sources [1, 3] for 18 movies of the MCU series. We then manually annotated the gender of each character for the corresponding film in our dataset.

Our dataset comprises 15 MCU movies with over 13,800 dialogues and 470 characters. This dataset encompasses characters from various storylines and arcs within the Marvel franchise. The breakdown of the dataset is as follows: 328 male characters, 127 female characters, and 15 neutral characters. The neutral characters consist of background characters whose gender cannot be extracted, for example, crowds, a narrator, a chorus, and a group of children.

By performing exploratory data analysis on the MCU dataset, we realised a significant disparity in the representation of males and females in the dataset distribution. By plotting the percentage of dialogues of males and females in each movie, we observe a trend that in the majority of the movies, 80% of the dialogues are delivered by male characters, while roughly 20% are spoken by females.

Moreover, we tried to analyse the frequency of conversations between genders to see the trends of interactions between genders. Most of the discussions stick to the group

Table 1. Number of Male and Female Characters in MCU Movies

Movie	Male Characters	Female Characters
Age of Ultron	26	12
Ant-Man	28	6
Avengers	39	13
Avengers: Endgame	39	13
Black Panther	27	16
Captain America	49	18
Civil War	33	9
Infinity War	38	9
Iron Man	29	8
Iron Man 2	26	10
Iron Man 3	28	18
Thor: Ragnarok	21	9
Thor	29	5
Thor: The Dark World	28	10
Winter Soldier	44	18

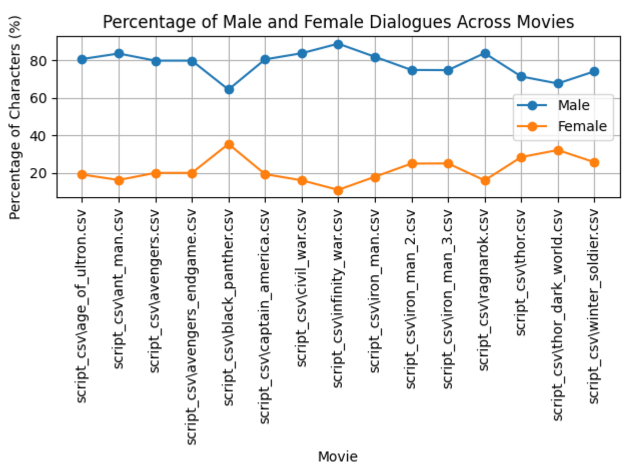


Figure 2. Gender Pair HeatMap which shows the percentage of conversations between two genders

(a) Male word cloud (b) Female word cloud

Figure 3. WORD CLOUDS

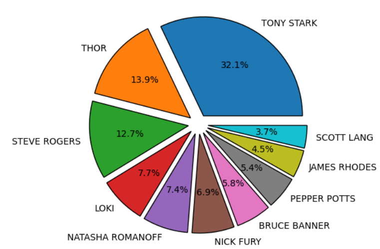


Figure 4. Top 10 Characters with the Most Dialogue across all movies

of the same genders, with 66% of all dialogue conversations between the same genders. Roughly 34% of the talks are between members of opposite sexes. We observe an interesting case here: the number of dialogues directed from males to females and those from females to males are almost the same, roughly 17% each. This shows that in a conversation, there is equal participation by both genders and no imbalance in the discussion between males and females.

To further explore character dynamics in the Marvel Cinematic Universe (MCU), we conducted extensive exploratory data analysis (EDA) using various visualization techniques. Specifically, we created word clouds(3) and pie charts(4 to understand better the distribution of dialogue and character representation across our dataset.

4. Methodology

4.1. Sentiment Analysis

Sentiment analysis determines the polarity of the text, which is positive, negative, or neutral. One of the famous tools for sentiment analysis is VADER.

It is an NLTK module that provides sentiment scores based on the words used. It is a rule-based sentiment analyzer in which the terms are generally labeled as per their semantic orientation as either positive or negative.

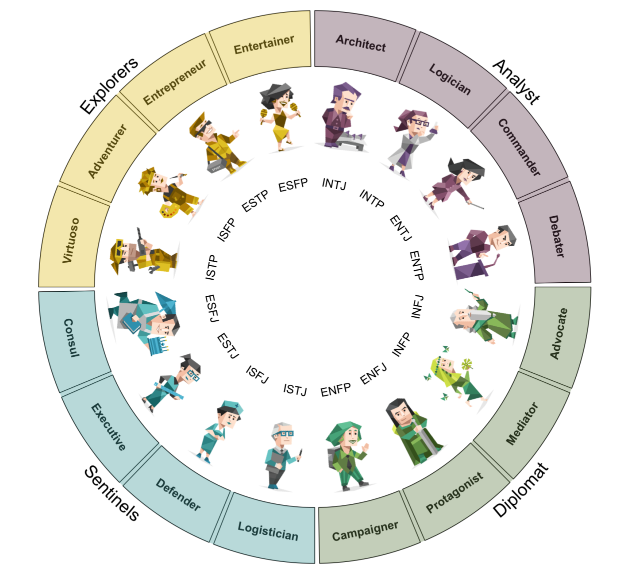


Figure 5. The MBTI demonstration of personality types vs true nature

To analyze sentiment trends across the Marvel Cinematic Universe (MCU), we employed a zero-shot classification technique on every piece of dialogue from each MCU movie. This approach allowed us to assess the sentiment of each character's dialogue without requiring pre-labelled training data, offering flexibility and accuracy in sentiment analysis.

After applying zero-shot classification to the entire dataset, we grouped the results by character, creating a comprehensive view of sentiment for each individual. This enabled us to track changes in character sentiment across the MCU timeline, providing insights into how characters evolve emotionally for the cinematic universe.

By examining sentiment shifts and trends for specific characters, we better understood narrative development and character arcs within the MCU. This analysis helped reveal how characters' emotions and sentiments change over time, contributing to the broader storytelling framework in the Marvel franchise.

4.2. MBTI Analysis

Myers-Briggs Type Indicator is a very famous test that claims to indicate differing "Psychological Types". The test assigns a binary value to each category: introversion or extroversion, sensing or intuition, thinking or feeling, and judging or perceiving.

Figures 5 and 6 give a high-level overview of the personalities. Figure 5 shows how MBTI relates to the proposed personalities. Figure 6 shows what each label means to one's personality.

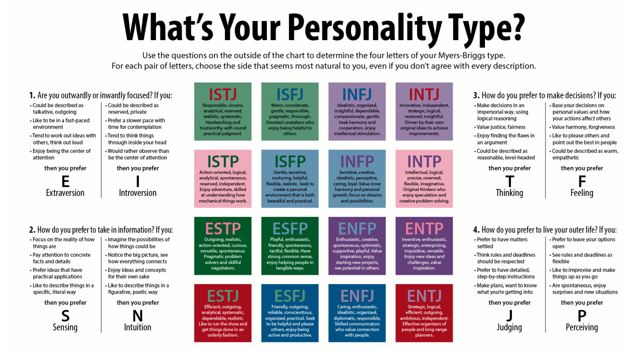


Figure 6. MBTI personality description of each binary label

Since this task is relatively unexplored, there were very few transformer-based models in the existing literature. Most attempts at this task were classical ML methods such as Naive Bayes and SVMs. In our work, we attempted to fine-tune the BERT and RoBERTa models to classify the personality.

An important point to note is that the dataset [2], is a highly imbalanced dataset, with I/E and N/S is exceptionally high(about 4:1 for both). This imbalance, along with the fact that the available dataset is of conversations on an E-forum about personalities, as opposed to our dataset of dialogues, due to which our model has very monotonic predictions for these 2 labels.

This is also very far from the true speculated representation in the true population, which would imply that the model would generalize very poorly.

Initially, we attempted to perform classification into precisely one of sixteen labels of personalities. However, our results on the validation split were very poor.

We also observe that since misclassifying only 1 of the 4 traits is not considered a completely wrong result, we remodelled our architecture to have 4 normalized logits, one for each trait.

Our observed results are shown in table 2

4.3. Emotion Analysis

For the task of emotion recognition and analysis on the dataset, we have chosen the Go Emotions [4] paper by Google Research, who developed a model for predicting emotions for each text from a set of 27 unique emotions and a neutral label. They used the Go Emotions dataset, a set of more than 58,000 Reddit comments labelled with their respective emotions. The emotions are each predicted using the sigmoid activation function; hence, this is a multi-class emotion classification task.

We utilized the pre-trained BERT-based model to perform zero-shot emotion classification on the MCU dataset. This way, we predicted emotions for each dialogue with a threshold of 0.3 over the sigmoid prediction values. While

Table 2. Results of the 2 models in personality analysis

Model	F1 for I/E	F1 for N/S	F1 for T/F	F1 for J/P
BERT	0.731	0.700	0.760	0.669
RoBERTa	0.664	0.642	0.704	0.632

analysing the emotions predicted on the MCU dataset, we observed that we received sparse predictions due to the high number of classified emotions. Hence, we tried to reduce the threshold to 0.0 and analyse the obtained results by aggregating the emotions. We observed better interpretability over the inferences with the threshold of 0.0, although the lower threshold added some noise over the emotions occurring with lower probabilities.

Henceforth, we tried to reduce the number of emotions to 6 with neutral and merged the emotions predicted. However, we observed that almost all characters, movies and gender combinations resulted in similar results, pointing to the possibility of averaging over the various emotions predicted. Due to poorer interpretability results, we went ahead with inferencing over 28 emotions instead, with a threshold of 0.0.

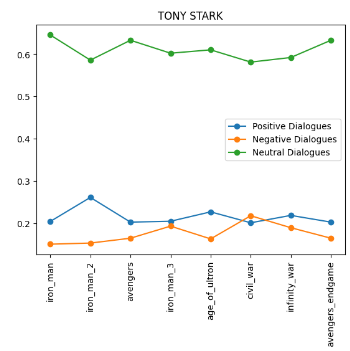
4.4. Gender Prediction

We worked on a gender classification task based on a given dialogue using Twitter data and then validated the results with a dataset of movie dialogues. Initially, we employed a pre-trained model, precisely the "padmajabfrl/Gender-Classification" from Hugging Face's repository, to benchmark our approach. F1 weighted was 0.39 and Accuracy was 0.45.

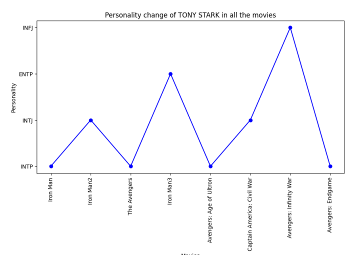
We fine-tuned a RoBERTa transformer by adding a single classification layer to its top. This was done with a dataset comprising 6,700 tweets from females and 6,194 tweets from males, aiming to enhance the model's gender classification capabilities.

Next, we froze the transformer model, maintaining its learned weights, and focused only on training the linear classification head with a movie-related dataset. We used tweets about the movie 'Avengers: Endgame' as our testbed to evaluate the model's performance. Given the potential class imbalance in this dataset, we measured our model's accuracy and weighted F1 score to ensure a robust evaluation.

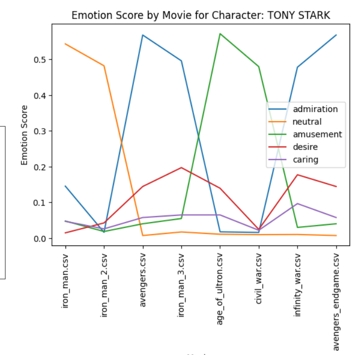
The F1 weighted score was particularly crucial in this context, as it accounts for class imbalance, providing a more reliable measure of the model's performance across different groups. This approach allowed us to assess the model's versatility and effectiveness in classifying gender within varying contexts.



(a) Sentiment Trend



(b) Personality Trend



(c) Emotion Trend

Figure 7. Graphs for Tony Stark

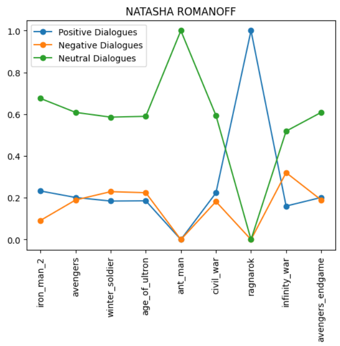
5. Results and analysis

We try to draw an analysis of 3 selected characters - 'Iron Man', 'Natasha Romanoff', and 'Thor Odinson'. We observe the trends across the three tasks and try to analyse the progression of the selected characters in the movies in chronological order.

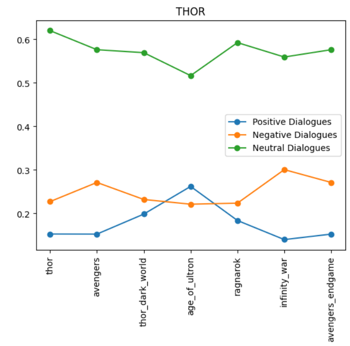
5.1. Tony Stark

Observations from the graphs(7):

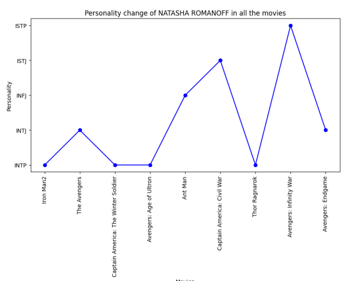
- For the movie 'Captain America: Civil War', we observe a hike in negative sentiment, personality becoming more judgmental and an evident dip in 'caring' and 'desire' emotion. During the civil war, Tony got to know that one of his friends was behind his parents' death; he was angry throughout the movie, which explains the hike in negative sentiment. Due to him being extremely furious, his caring attitude took a backseat.



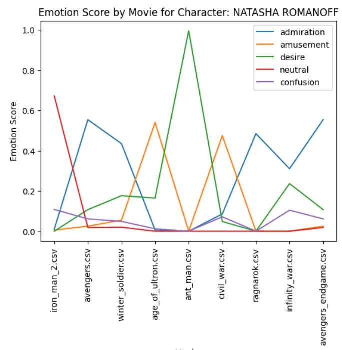
(a) Sentiment Trend



(a) Sentiment Trend

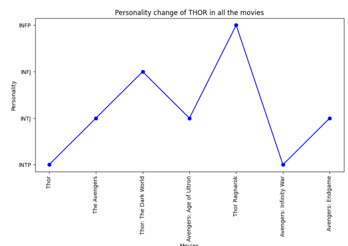


(b) Personality Trend

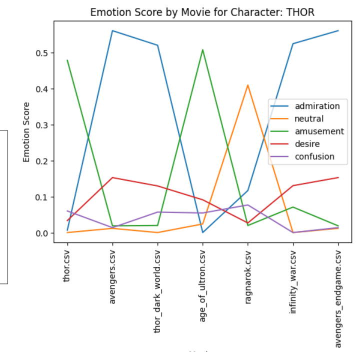


(c) Emotion Trend

Figure 8. Graphs for Natasha Romanoff



(b) Personality Trend



(c) Emotion Trend

Figure 9. Graphs for Thor Odinson

- For the movie ‘Iron Man 3’, we notice a little bit of a hike in negative sentiment, personality becoming extrovert and a significant hike in ‘desire’ emotion. In this movie, Tony’s enemy from the past is back to take revenge. In the fight, Tony’s home was destroyed, and he ended up abandoned alone, almost 400 miles away from his residence. There, he had to talk to many other people for help, which explains why his personality became an extrovert, and the restlessness of getting rid of the enemy and saving his people was the reason for the hike in ‘desire’ emotion.

5.2. Natasha Romanoff

Observations from the graphs(8):

- For the movie ‘Avengers Endgame’, we observe a dip in negative sentiment, personality becoming more judging and a significant hike in ‘admiration’ emotion. During the endgame, Natasha sacrifices herself for the soul stone to save the world. As she is exceedingly selfless throughout the movie, the negative sentiment drops and ‘admiration’ emotions become very strong.

- As compared to Tony Stark, Natasha’s character has been consistent throughout the universe. We can do some hikes in ‘Ant-Man’ and ‘Thor: Ragnarok’, but they were because she had a tiny number of dialogues in those movies.

5.3. Thor Odinson

Observations from the graphs(9):

- For the movie ‘Avengers’, his negative sentiment has hiked, his personality has changed to ‘judging’, and his admiration and desire have peaked. In this movie, all the Avengers were fighting against Thor’s brother Loki, which is a clear reason for the hike in negative sentiment. As Loki has always been a negative character, and Thor has always been trying to convert that, we can see a hike in desire.

- For the movie ‘Avengers: Age of Ultron’, his personality trait changed to ‘thinking’, and his amusement hiked, alongside a dip in admiration. In this movie, Thor’s character is portrayed as sarcastic and funny (perhaps because of his high strength), which may be why he hikes in amusement. ‘Thinking’ seems like a misclassified label as he is a very soft and very ‘feeling’ character towards his fellow Avengers.

5.4. Gender Prediction

The results(10) are all derived from a single movie, ‘Avengers: Endgame’. Observing the progression, we can

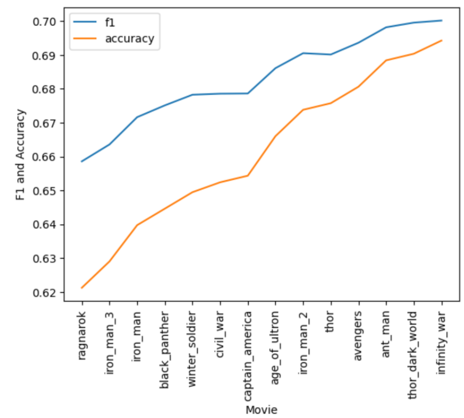


Figure 10. F1 scores and accuracy trend as we fine-tune on movies.

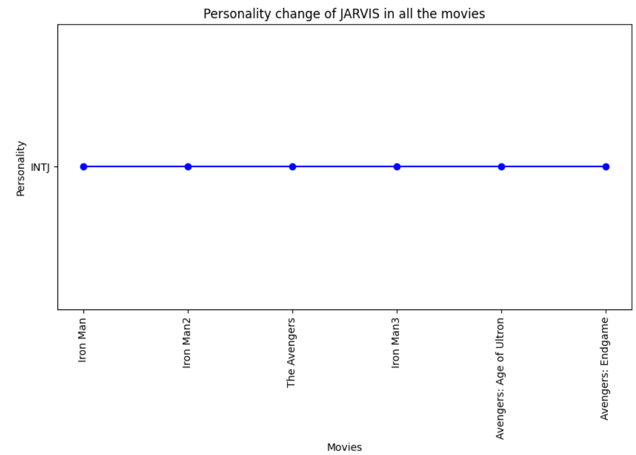


Figure 11. Personality of Jarvis over the Marvel Cinematic Universe

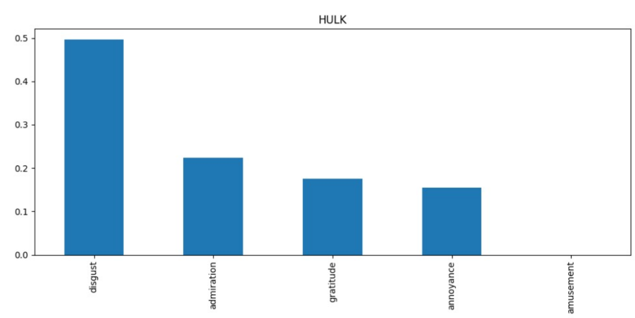


Figure 12. Emotional Analysis of Hulk over the Marvel Cinematic Universe

see that the performance metrics, such as F1 scores and accuracy, tend to improve as the model is fine-tuned with data from an increasing number of superhero movies. This improvement can be attributed to the unique linguistic patterns often associated with superhero characters.

In superhero films, characters tend to have distinctive speech patterns and mannerisms. As the model is exposed to more of these superhero-themed dialogues and scenarios, it better understands the nuances in character speech. This greater exposure helps the model recognize patterns and contextual cues, enhancing performance when applied to 'Avengers: Endgame' data.

Essentially, the improvement in results underscores the benefit of fine-tuning with domain-specific data. As the model encounters a broader spectrum of superhero narratives, it becomes better equipped to make accurate predictions and discern character traits through language use. The refinement process helps the model adapt and improve, demonstrating the importance of contextually relevant data in training effective machine learning models.

6. Interesting Results

6.1. Personality Analysis

JARVIS has a constant personality, which is consistent with the movies since JARVIS is an AI system and is the same throughout the movies.
Even though the reported F1s of our models were only at around 70% for each trait, the model seems to have classified the case of Jarvis perfectly.

Figure 11 shows how Jarvis has zero personality change over the timeline of Marvel movies.

6.2. Emotional Analysis

Hulk is a very aggressive character in the Marvel Universe. This emotional analysis also shows that "Disgust" is his dominant emotion.

Figure 12 shows that 'Disgust' is Hulk's dominant emotion.

6.3. Sentiment Analysis

Loki is known as the "God of mischief" and is famous for being notorious. In this sentiment analysis, we can see a spike in his positive dialogue in the movie 'Avengers: Infinity War' because that was the first time he cared for others.

Figure 13 shows the increase in positive dialogues.

7. Conclusion and Future Work

Through this work, we can conclude that there is a considerable difference in the representation of males and females in movies, as confirmed by the MCU series. We have observed disparity in representation, significance and characteristics of males and females. Males such as Tony Stark and Thor Odinson have been observed to have erratic development of characters through the series, whereas female characters such as Natasha have been seen to have more con-

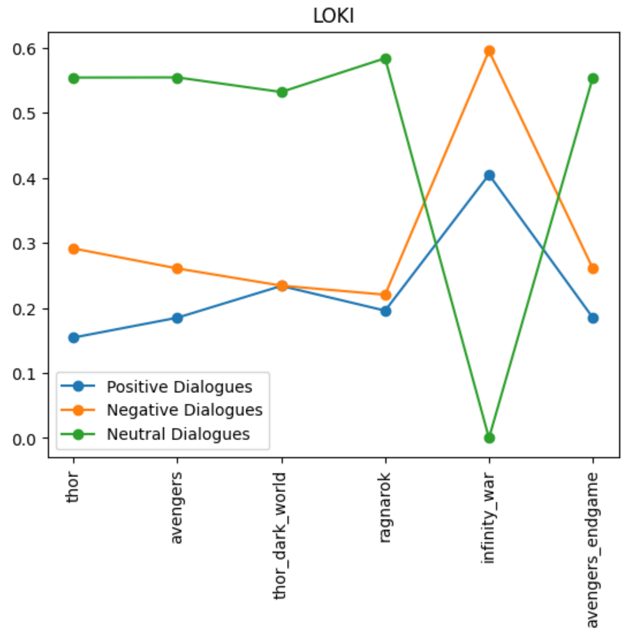


Figure 13. Sentiment Analysis of Loki over the Marvel Cinematic Universe

stant behaviour. The irregular behaviour points to an attempt at character development, which seems to have not been the case for Natasha (to some extent). Overall, we notice interesting differences between the representation of males and females in the movie series. Moving forward, we believe that movies should strive to have a more balanced representation of the characters and genders to avoid prejudices and biases being reinforced in content. Moreover, we look forward to further analysis of the MCU dataset and plan to conduct Bechdel tests on the movies to better learn about its characteristics.

References

[1] Marvel cinematic universe (mcu) movie scripts csv. Clean scripts of various movies. 2

[2] Mbti dataset csv. MBTI datasets of a Personality forum. 4

[3] Various Authors. Marvel cinematic universe (mcu) movie scripts. Movie script data used to create the dataset for analysis of character dynamics in the MCU. 2

[4] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*, 2020. 2, 4

[5] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014. 2

[6] Juan José Igartua and Isabel Matilde Barrios Vicente. Changing real-world beliefs with controversial movies: Processes and mechanisms of narrative persuasion. *Journal of Communication*, 63:514–531, 06 2012. 1

[7] Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79, 2017. 2

[8] Yisi Sang, Xiangyang Mou, Mo Yu, Dakuo Wang, Jing Li, and Jeffrey Stanton. Mbti personality prediction for fictional characters using movie scripts. *arXiv preprint arXiv:2210.10994*, 2022. 2