**Individual Report**

**Member name:** [Rahul Nayak](mailto:rrnayak@asu.edu)

**Evaluated by:** [Anuranjan Dubey](mailto:adubey37@asu.edu)[Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Date: 10/08/2023**

**Tasks Assigned:**

* Preparing individual progress report
* Evaluating [Avani Mundra](mailto:amudra@asu.edu)’s individual progress and in-depth reports
* Finding one more research paper related to my research domain

**Summary:**

* This week my tasks began by me searching research papers that were pertaining to my domain. I found one paper towards the end of the week, thus I will be completing an In depth study of this next week. In the meantime I studied 2 Not so important papers to expand my knowledge around the topic of Suspicious activity happening on social media.
* After studying [1] I did comprehend from the paper that Social media is an essential part of modern life, enabling individuals to freely share their thoughts and opinions. However, it also harbors harmful content such as fake news, hate speech, and cyberbullying, causing significant societal concerns and affecting mental health.Detecting and moderating such content is vital.
* Social media platforms like Facebook and Twitter have invested in policies and systems for content moderation, but the sheer volume surpasses manual efforts. Artificial Intelligence (AI), including Natural Language Processing (NLP) and Machine Learning (ML), offers automated solutions. While content detection has received substantial attention, content moderation has been relatively neglected.
* This research study comprises three parts: detecting detrimental content using NLP, moderating such content, and identifying research gaps. AI-driven moderation is crucial to handle the rising volume of harmful content on social media platforms. Further research is needed to improve moderation methods and address emerging challenges in this critical domain.
* The next paper [2] examines how extremist groups exploit social media for recruitment and radical propaganda. It reviews recent research on online extremism in text, highlighting gaps and opportunities in a systematic literature review of 45 studies from 2015 to 2020.
* The study emphasizes the need for a better understanding of online extremism trends and directions, offering a critical analysis of this emerging research field.

**Outcome:**

Studying [1] highlighted the importance of AI-driven content moderation on social media due to the prevalence of harmful content. [2] revealed the growing use of social media by extremist groups for recruitment and radical propaganda, emphasizing the need for a deeper understanding of online extremism trends. Both studies underscored the significance of addressing these issues in the digital landscape.

**References** *(with citation)*

[1] V. U. Gongane, M. V. Munot, and A. D. Anuse, “Detection and moderation of detrimental content on social media platforms: current status and future directions,” Social Network Analysis and Mining, vol. 12, no. 1, p. 129, 2022.

[2] S. Aldera, A. Emam, M. Al-Qurishi, M. Alrubaian and A. Alothaim, "Online Extremism Detection in Textual Content: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 42384-42396, 2021, doi: 10.1109/ACCESS.2021.3064178.

**Evaluation of Report  
  
Evaluation by:** [Anuranjan Dubey](mailto:adubey37@asu.edu) **Date: 8th October 2023**

**Is the weekly member report complete with all the major result(s) of the paper(s)? If not, provide as many examples of the major result(s) missing in the written report as possible. (in bullet form). [within 100 words]**

* Two papers are studied thoroughly, Specific findings or insights related to the detection of detrimental content using NLP.
* Insights into how extremist groups exploit social media for recruitment and radical propaganda. Details on the methods or techniques proposed in the paper for content moderation.
* Report provides an overview of the topics covered in the papers.

**Is each section of the guidelines sufficiently completed? If not, point out what is missing. [Normally within 40 words].**

Yes, each section of the guideline is sufficiently completed.

**Is the quality of this version of the written report satisfactory? If not, then why not? [Normally within 40 words]**

Yes.

**Approved by:** [Gautham Vijayaraj](mailto:gvijaya6@asu.edu) **Date: 10/9/2023**