**Individual Report**

**Member name:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)

**Evaluated by:** [Justin Young](mailto:jtyoun15@asu.edu) [Gautham Vijayaraj](mailto:gvijaya6@asu.edu)

**Date:** 10/12/23

**Tasks Assigned:**

* Preparation and Evaluation of Individual Progress Reports
* Read not-so important papers

**Summary:**

* Read the research papers classified as not-so important, which are to be read by all the team members.
* Evaluation of individual progress reports of one team member was completed successfully.
* The research paper read this week was ‘Detection and moderation of detrimental content on social media platforms: current status and future directions’.
* The paper focuses on the large amount of detrimental content on social media causing negative effects on the people consuming them and the ways to moderate and reduce such content.
* Here, the integrity of an honest user is not compromised as only those users who post detrimental content are taken into consideration.
* A major method for moderation of detrimental content covered in the research paper are legal provisions which have been made by certain governments on social media platforms.
* The methodology for social media content moderation proposed in the research paper is as follows:
  + Research question definition
  + Relevant topic collection from scientific literature and articles
  + Information mapping between literature and research questions
* Dataset creation is an important step for analysis of detrimental content on social media. The major concepts covered in this dataset could range from name of dataset, features, label categorization and skewness.
* Different Machine Learning algorithms are used on the dataset. These algorithms play an important role in detection of hate speech and other kinds of detrimental content on social media.
* The paper also discusses the techniques which could be used for moderation of detrimental content on social media. The techniques are:
  + Manual approach - Completely manual checking of detrimental content
  + Semi-automated technique - Partially manual, which uses AI to flag certain detrimental content
  + Automated technique - Completely automatic which uses automated tools to detect and moderate extremist and detrimental content.
* Overall, it is found from the paper that machine learning and AI methods are a boon to mankind to detect and moderate detrimental content on social media.

**Outcome:**

All the tasks assigned this week - preparation and evaluation of individual progress reports and reading of not-so important research papers was completed successfully.

**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.

**Evaluation of Report  
  
Evaluation by:** [Justin Young](mailto:jtyoun15@asu.edu) **Date:** 10/13/23

**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]**

* Yes, this report covers all major results of the tasks for this week

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

Yes

**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: 10/15/2023**