Project Title: Al Incident Database / Digital Safety Research Institute

PDS Version / Date: Feb 2nd, 2024

Project

Background:

The project focuses on analyzing media reports related to incidents caused by the deployment or development of artificial intelligence (AI) systems. The AI Incident Database (AIID), managed by the Digital Safety Research Institute (DSRI), aims to collect and analyze real-world harms caused by AI to educate stakeholders and promote responsible AI development

Desired Business Outcomes:

- 1. Gain insights into changing trends and rhetoric in Al journalism.
- 2.Enhance the AIID's ability to objectively evaluate media reports and identify emerging Al-related harms.
- 3. Provide recommendations for improving the indexing and treatment of media reporting within the AIID

Three Finger Summary					
Foreground	Trigger	Central Issue			
What are the key, relevant facts that define the current context?	What is the need for change?	What is the defining issue that needs to be addressed by the project?			
Foreground Answer	Trigger Answer	Central Issue Answer			
Analyzing media reports on Al-related incidents for the Al Incident Database to objectively evaluate and report on Al-related harms.	The need to understand evolving trends and rhetoric in AI journalism to enhance the AIID's effectiveness.	Identifying and analyzing changing patterns in AI journalism to improve incident evaluation and reporting in the AIID.			
Key Questions					
What are the prevailing trends and rhetoric in Al journalism, and how have they evolved over time?	Are some publishers consistently more detailed or nuanced in how they describe AI systems specifically?	How can computational techniques such as natural language processing (NLP) be leveraged to analyze large volumes of media reports effectively?			

Client expectations				
Client name	Kevin Paeth	Sean McGregor		
Position	Sponsor: Digital Safety Research Institute (DSRI)	Technical Specialist: Kevin Paeth	Manager: Sean McGregor	
Criteria for Quality and Action (how do you define success)	Constantly check and meet with clients for achieving requirements, weekly reflection with clients and group members. Complete the project within the designated time frame to ensure timely delivery of results and recommendations. Ensure that the analysis methodologies are well-documented and reproducible, allowing for validation and further research in the future. Ensure that the analysis and reporting adhere to ethical standards and respect the privacy and rights of individuals mentioned in the media reports. Identify opportunities for improvement in the analysis process and methodologies			