Internship Project 1

Title: Explore the usage of Chat-GPT to produce polarization annotated dataset

Description: The intern will be responsible for exploring the possibility of using Chat-GPT or other open-source and freely available Large Language Models (LLMs) (like Vicuna, WizardLLM) to generate a polarization dataset. The intern will work closely with our research team and will be responsible for constructing proper prompts to generate the content in the form of news articles or social media posts (like in Twitter) that contain polarizing content. The interns will also be expected to investigate potential limitations and restrictions posed by Chat-GPT or other LLMs, like ethical response guidelines, and develop methods to bypass them. Additionally, the intern will be required to process the generated content and obtain pre-defined ground-truth results to create a polarization dataset. The dataset will be used to train and evaluate machine learning models for detecting polarizing content.

Internship Project 2

Title: Discovery of Positively Affiliated Twitter Accounts

Description: The intern will work on a project that involves the use of the Twitter API to collect tweets surrounding a certain discussion, and process the interaction of users to determine a group of Twitter accounts that are strongly and positively affiliated with one another. The intern will work closely with our research team and will be responsible for collecting Twitter data using the Twitter API, processing and cleaning the data, and applying network analysis methods to identify groups of strongly and positively affiliated Twitter accounts. The intern will also be responsible for constructing a small subset of known positively affiliated accounts to evaluate the results to a certain extent. In terms of qualifications, the intern should have experience in programming with Python, and some familiarity with social network analysis concepts like nodes, edges, communities etc.

Internship Project 3

Title: Process, analyze and visualize video data from pitch competitions, political debates and court proceedings.

Description: As an intern, you will have the opportunity to work with a cutting-edge tool that processes video data, enabling you to extract valuable insights from a wide range of sources, including pitch competitions, political debates, and court proceedings. Furthermore, this internship will provide you with hands-on experience in utilizing advanced data processing techniques to handle video data from diverse contexts. You will have the chance to work with real-world datasets from pitch competitions, political debates, and court proceedings, honing your skills in data analysis and visualization to unlock key patterns and trends.

Internship Project 4

Title: Evaluate, clean, and optimize real-world big data projects codebases

Description: The intern will have the opportunity to work on real-world big data projects, gaining practical experience and will analyze existing code, identify areas for improvement, and implement effective strategies to optimize and refine the codebases. By refactoring the code for enhanced readability and efficiency, the candidate will contribute to the development of high-performing data analysis pipelines. Throughout the internship, the intern will have the privilege of working on modern big data analytics stack developing a deeper understanding of efficient coding techniques and optimization strategies. Moreover, this internship will allow the intern to work extensively with Apache Spark and its python library (PySpark). Through hands-on practice, she/he will gain skills and experience enhancing the candidate's technical abilities and will create a solid knowledge in the field of big data analytics.