Project Report: Suicide Prevention by Detection through Large Language Models (LLMs)

Team Members

1. Rohan Sawant
2. Dhruv Toshniwal
3. Abhaya Shukla

Introduction

This project aims to use Large Language Models (LLMs), including GPT-3.5 and LLaMA-2, for the early detection of suicidal tendencies and depression in individuals through their social media posts. The goal is to contribute to suicide prevention efforts, leveraging the latest advancements in Natural Language Processing (NLP).

Problem Statement

Suicide remains a critical public health issue globally, with early detection of suicidal ideation being vital for timely intervention. Traditional methods often fall short in identifying subtle signs of depression or suicidal thoughts, particularly in social media content. This project seeks to address this gap using the capabilities of LLMs.

Importance

- Mental Health Awareness: Enhancing efforts in mental health awareness and suicide prevention.
- Technological Advancement: Advancing the application of NLP in understanding human emotions and mental states.
- Early Intervention: Facilitating timely interventions by detecting early signs of mental distress.
- Research Contribution: Adding valuable insights to the fields of NLP and mental health.

Relation to Classwork and Interests

The project aligns with the principles and skills taught in our NLP class, demonstrating the practical application of NLP techniques in a real-world scenario. It also reflects our team's interest in the intersection of AI and mental health.

Resources and Dataset

1. Medium Article: "Suicide Prevention by Detection: Leveraging AI on Social Media" on HackerNoon.
2. Dataset: Kaggle's "Suicide Watch" dataset with 232,074 unique Reddit posts up to 2021 (166MB).
3. Research Paper: "Towards Suicide Prevention: Early Detection of Depression on Social Media" from Springer.

Methodology

- In-Context Learning: Leveraging LLMs for contextual understanding of social media content.
- Classifier Training: Developing a classifier to identify language patterns associated with depressive or suicidal ideation.
- Model Comparison and Evaluation: Assessing different LLMs for effectiveness in detection.

Execution Plan

1. Data Preparation: Processing the Kaggle dataset for model training.
2. Model Training and Tuning: Fine-tuning LLMs using the dataset.
3. Testing and Validation: Ensuring the accuracy and reliability of the models.

Evaluation Strategy

- Accuracy: Precision in identifying relevant posts.
- Sensitivity: Ability to correctly identify genuine cases of distress.
- Utility: Practical applicability in real-world scenarios.

Conclusion

This project aims to harness the potential of LLMs for a life-saving application in mental health. Through thorough research, model development, and evaluation, we aspire to make a meaningful contribution to suicide prevention and mental health awareness.