

Paper Title: ChatGPT for Suicide Risk Assessment on Social Media: Quantitative Evaluation of Model Performance, Potentials, and Limitations

Paper link:

https://aclanthology.org/2023.wassa-1.16/?fbclid=IwAR15m3HC_ZZ1U1GVsaPhYhFpN8oPF4-i7bz-XtAQA-IMNu_Le5rFw_eRSq4

1 Summary:

1.1 Motivation:

This study aims to make sure that chatGPT could be a reliable tool to detect suicide risks and empathize with people who are suffering from suicidal thoughts or depression.

1.2 Contribution:

The paper evaluates chatGPT using Zero-shot and few-shot learning techniques and two transformer-based models, ALBERT and DistilBERT. It also analyzes how changing temperature parameters affect chatGPT's ability to achieve suicide risk assessment.

1.3 Methodology:

The paper uses a dataset from the University of Maryland Reddit Suicidality Dataset(UMD), collected from the Reddit platform. The paper evaluates the ability of chatGPT to predict the level of suicide ideation on the dataset and compares the results with two fine-tuning classifiers. ALBERT and DistilBERT models were used to fine-tune the dataset and train them. It also evaluates the ability of chatGPT for zero-shot and Few-shot learning to predict the severity of suicidality of text. Uses different parameters like “message” parameter for input and “Temperature” parameter to control responses. While ChatGPT's performance is comparable to a fine-tuned DistilBERT, it falls considerably short (by 13% for F1-score) compared to a fine-tuned ALBERT model.

1.4 Conclusion:

These findings demonstrate the potential of ChatGPT as a tool for data annotation, particularly when utilizing a simple prompt design.

2 Limitations:

2.1 First Limitations:

Although this paper evaluates chatGPT's performance, it has not measured the other safety criteria like biases or privacy issues in using this model. This means, There is no surety of whether the user's identity or any information is secured.

2.2 Second Limitations:

Only focused on the initial steps of suicide risk assessment but did not explore the use of chatGPT in ongoing monitoring or intervention of the suicidal person.

3 Synthesis:

This study has a good scope in future applications to detect suicidality by analyzing social media posts and comments. Although this paper didn't work the afterward works like "what should be the next step after it detects a suicidal post", detecting the suicidal person can also be helpful to aware people. This paper has the potential to improve the study and it might also contribute to decreasing the suicide rate in society.