

**Application to use the UMD Reddit Suicidality Dataset, Version 2
& CLPsych 2024 Dataset in the CLPsych 2024 Shared Task**

Background

The CLPsych 2024 shared task will be making use of the UMD Reddit Suicidality Dataset, which was created by researchers at the University of Maryland and is described at http://umiacs.umd.edu/~resnik/umd_reddit_suicidality_dataset.html.

Although Reddit is a site for anonymous discussion and additional steps have been taken to de-identify the data, Reddit data (and particularly data related to mental health) can still be thought of as sensitive. As described on the UMD Reddit Suicidality Dataset web page, therefore, an application process has been put in place in collaboration with the American Association of Suicidology for sharing of the dataset. For the CLPsych 2024 shared task we are using a modified version of that application.

Note that this application is only for use of the data as part of the CLPsych 2024 shared task. If you wish to continue using the data after the CLPsych 2024 workshop, or to use it for other purposes, you will need to submit an application that also includes getting IRB or equivalent ethical review (see details [here](#)). However, for the CLPsych shared task, you can use the streamlined application below without needing to get your own IRB/ethical review approval.

Application for Use of the Dataset in the CLPsych 2024 Shared Task

I am applying for access to the UMD Reddit Suicidality Dataset and CLPsych 2024 data in order to participate in the CLPsych 2024 shared task. Other members of my team are as follows (or say N/A, for “not applicable”, if you are participating solo):

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____

Please list team members here one per line with their contact email addresses

My team and I confirm the following:

1. We will use this dataset ONLY for the CLPsych shared task and for the duration of the shared task. We will delete all dataset files and models that were trained on the dataset immediately after CLPsych 2024 workshop date (March 22nd, 2024)
2. During the CLPsych 2024 dataset we can use the data only for the sole purpose of participating in the shared task. (If you would like to use the dataset for your ongoing research, please submit the official application for the dataset access using the instructions posted here: https://users.umiacs.umd.edu/~resnik/umd_reddit_suicidality_dataset.html)

3. We confirm that we have read Benton, A., Coppersmith, G. and Dredze, M. (2017), “Ethical research protocols for social media health research”, in Proceedings of the First ACL Workshop on Ethics in Natural Language Processing (pp. 94-102), <http://www.aclweb.org/anthology/W17-1612>, and that in our use of this dataset we are committed to maintaining that paper’s broad ethical principles.

4. We commit to citing the "University of Maryland Reddit Suicidality Dataset", in any publications using or discussing this dataset using including appropriate references (you can find BibTeX entries on the dataset web page):

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik, "Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Postings", Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pages 25–36, New Orleans, Louisiana, June 5, 2018.

Ayah Zirikly, Philip Resnik, Özlem Uzuner, and Kristy Hollingshead. 2019. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology (CLPsych'19), 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Minneapolis, June 6, 2019.

5. We commit to acknowledging the assistance of the American Association of Suicidology in making the dataset available, in any publications.

6. The data and any derivatives will be stored only on password-protected servers at _____

[MY ORGANIZATION OR LIST MULTIPLE ORGANIZATIONS] where access will be restricted to me and my team using Unix group permissions or a reasonable equivalent;

7. Any copies of the data or derivatives of it that we create will be accompanied by a clear README.txt file identifying me as the contact person and stating that further re-distribution is not to take place without contacting me first.

8. We will refer any requests for redistribution of the data (outside my own team) to the providers of the dataset, rather than providing the data ourselves.

9. We commit to NOT submitting all or part of the data to any platform that would potentially store the data or use it as training data (e.g. OpenAI’s ChatGPT).

Signature by contact person/team lead:

Affirmed: _____ (please sign)

_____ (please print name)
_____ (please print email address)