

Data Enclave Use Agreements

Permitting Researchers Access to CLPsych 2022 Shared Task Data

1.0 Terms and Conditions

This Data Enclave Use Agreement is entered by and between National Opinion Research Center (dba NORC) and undersigned researcher who is authorized by Maria Liakata of Queen Mary University of London to access the data associated with the CLPsych 2022 Shared Task that include the UMD Reddit Suicidality Dataset annotated for moments of Change (MoC) (“UMDRSDv3”), a new Queen Mary University of London (QMUL) Reddit Dataset annotated for both Suicidality and MoC, and a subset of the eRisk Dataset annotated for MoC (the “Data”). Pursuant to the terms and provisions of this Agreement. The undersigned (referred to herein as the “Requestor”) affirm that they understand and agree with all conditions and responsibilities set forth herein. All conditions defined in this Agreement will remain in effect throughout the 4-month duration defined in this Agreement (see section 9.0: Duration of Agreement) or until a subsequent written agreement supersedes this Agreement.

2.0 Purpose

The purpose of this Agreement is to ensure the integrity and confidentiality of information contained in Data as used by the Requestor. The Agreement is customized to each Requestor’s needs and defines requirements and restrictions unique to each of the access parameters defined herein.

3.0 Data Description

After receiving approval from NORC, Requestor may access the Data and other information made available to them in the NORC Data Enclave®.

4.0 Conditions of Data Access

Requestor agrees to the following conditions for accessing Data maintained in the NORC Data Enclave®:

4.1 Requestor shall use Data solely for the Requestor's statistical research purposes associated with the CLPsych 2022 shared task. Data may not be used for any other purposes whatsoever, including administrative, regulatory, marketing, law enforcement, judicial, or other purposes.

4.2 Requestor agrees that he/she shall not attempt to capture, store or share any images, files or information accessed within the Data Enclave using any form of storage, screen capture software or devices (including any type of image recording device), screen sharing software or devices, or by allowing unauthorized users to view the Data Enclave.

- 4.3 Requestor shall not attempt to re-identify respondents, including without limitation persons, families, households, schools, establishments, firms, economic units or any other entities ("Respondents"); nor will any list of identities or raw data elements be published or otherwise distributed.
- 4.4 Within 24 hours of the time when a Requestor becomes aware that the identity or identifying information of any Respondent, Household, or Firm may have been disclosed, Requestor shall advise NORC of the disclosure in writing, inadvertent or otherwise, by sending an email at privacy@norc.org. Requestor shall use his/her reasonable efforts to destroy, retract, or otherwise safeguard from further dissemination of the materials containing the Respondent, Household, or Firm's identity or containing information from which the Respondent, Household, or Firm's identity may be determined.
- 4.5 All research output, such as statistical results and reports derived using statistical analysis applications in the Enclave must be reviewed by NORC and the CLPsych 2022 shared task organizers, solely for the purpose of ensuring that appropriate statistical disclosure protection standards have been met before being released to the Requestor.
- 4.6 Requestor shall not publish any work containing information or results derived from Data that identifies a Respondent, Household, Firm, or any other underlying entity.
- 4.7 Requestor shall respond to requests from NORC and the CLPsych 2022 shared task organizers for information on research progress and preliminary findings.
- 4.8 Requestor agrees and acknowledges that Requestor shall not export any Data from the Enclave in whole or in part, or any Trained Models (digital information derived from the Data that could be used to analyze other datasets, including but not limited to word or term lists, probability tables, or parameters of neural networks; except for small fragments included for illustrative purposes in Publications) from the Enclave unless permitted by NORC or the CLPsych 2022 shared task organizers in writing. Requestor is permitted to copy code, experimental results, and small fragments for illustrative purposes in Publications from the Enclave via the export request and review process described in the data enclave training guide (see Paragraph 7, below).
- 4.9 Requestor agrees to provide copies of any publication containing results derived from Data to NORC and the CLPsych 2022 shared task organizers (30) days prior to submission of such work, with the exception of the shared task paper, which is to be submitted on May 20th. Requestor agrees to provide copies of any work product containing results derived from Data for review by NORC and the CLPsych 2022 shared task organizers thirty (30) days prior to the submission of such work for publication for the purpose of ensuring that appropriate statistical disclosure protection standards have been met before being released to the public.

4.10 Requestor agrees to acknowledge the NORC Data Enclave® for any work containing information or results derived from the Data (e.g., *“The author(s) acknowledge the NORC Data Enclave® for providing researcher support and access to the data used in this research.”*). However, this shall in no way be construed as an endorsement of the Requestor’s work by NORC.

4.11 Requestor agrees to acknowledge the Data and cite the following papers per dataset.

- a. UMDRSDv3 (UMDRSDv2 with QMUL annotations), as described at http://users.umiacs.umd.edu/~resnik/umd_reddit_suicidality_dataset.html
- i. [1] Shing, Han-Chin, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. "Expert, crowdsourced, and machine assessment of suicide risk via online postings." In Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic, pp. 25-36. 2018.
 - ii. [2] Zirikly, Ayah, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. "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, pp. 24-33. 2019.

Respondent also agrees to acknowledging the American Association of Suicidology in any publications using this dataset.

- b. Queen Mary University of London (QMUL) Reddit Dataset annotated for both Suicidality and MoC
- i. [3] Adam Tsakalidis, Federico Nanni, Anthony Hills, Jenny Chim, Jiayu Song, and Maria Liakata. "Identifying Moments of Change from Longitudinal User Text". In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (to appear).
 - ii. [4] Overview of the CLPsych 2022 Shared Task (authors/exact title TBA)
- c. eRisk Dataset annotated for MoC as described at <https://tec.citius.usc.es/ir/code/eRisk2020.html>
- i. [5] David Losada, Fabio Crestani. "A Test Collection for Research on Depression and Language use". In Experimental IR Meets Multilinguality, Multimodality, and Interaction 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016.
 - ii. [6] David E. Losada, Fabio Crestani, Javier Parapar. Overview of eRisk 2020: Early Risk Prediction on the Internet. CLEF 2020, Lecture Notes in Computer Science 12260, 2020.
 - iii.

Respondent affirms having read Benton, A., Coppersmith, G., & Dredze, M. (2017, April. Ethical research protocols for social media health research. In Proceedings of the first ACL workshop on ethics in natural language processing, pp. 94-102) and commits to its broad ethical principles.

5.0 Violations of this Agreement

If NORC deem any aspect of this Agreement to be violated, they reserve the right to:

- a. Deny Requestor access to the Data and the NORC Data Enclave®.
- b. Withhold undelivered data output from Requestor.
- c. Report the violation to the appropriate authorities, including Requestor's Institution (and other applicable authorities) and recommend that sanctions be imposed.
- d. Invoke other remedies that may be available to NORC under law or equity, including injunctive relief to stop Requestor's use of any data or descriptions of data derived from the Data.
- e. Terminate this Agreement without any further cause and without notice.

6.0 Monitoring Usage

NORC reserves the right to take measures to monitor Requestor's use of the Data to ensure that the Requestor complies with all terms of this Agreement. These measures include, without limitation:

- a. NORC may maintain a full record of all computer-based interactions with the Data.
- b. NORC may review all Requestor's data output derived from Data.

7.0 Training

Prior to accessing the Data, Requestor agrees to review and follow the Data Enclave training guide, including the process by which approved researchers may access the Enclave session and submit data export requests.

8.0 Ownership of the Intellectual Property

NORC will actively and aggressively enforce their intellectual property rights to the fullest extent of the law.

"Intellectual Property Rights" means any rights existing now or in the future under patent law, copyright law, trademark law, database protection law, trade secret law, and any and all similar proprietary rights.

- a. Requestor acknowledges and agrees that all rights, including Intellectual Property Rights, to the architecture, computing systems, and computing environment of the Data Enclave, including without limitation Data and all other data unless provided by a third party under an independent arrangement, information, documents, programs, trade secrets, proprietary rights, confidential information and facilities that comprise or are contained in the Data Enclave, and any improvement made thereto during the term of this Agreement if the result constitutes a derivative work, belong to NORC.

- b. Requestor acknowledges and agrees that all rights, including Intellectual Property Rights, to inventions, discoveries, improvements, concepts, work product and programs conceived or made by NORC, its officers, directors, agents, employees, licensees, contractors, related companies, and all other persons or entities retained by NORC during the term of this Agreement in connection with this Agreement, including any extensions thereof, belong to NORC.
- c. Except as provided in paragraph 8(a), NORC acknowledges and agrees that all rights, including Intellectual Property Rights, to inventions, discoveries, improvements, concepts, works and manuscripts conceived or made by the Requestor, during the term of this Agreement, including any extensions thereof, in connection with Requestor's research, belong to Requestor.

9.0 Duration of this Agreement

This Agreement is effective upon the date that it is approved by NORC ("Effective Date"). Parties signing this Agreement agree to amend this Agreement as necessary for NORC to comply with all applicable federal and state law requirements regarding privacy and confidentiality of restricted data. Any ambiguity in this Agreement shall be resolved to permit NORC to comply with all applicable federal and state law requirements regarding privacy and confidentiality of restricted data. The Requestor and NORC have the right to terminate this Agreement without cause at any time.

10.0 Liability

10.1 Requestor agrees that NORC is not liable for any damage to computer systems or loss of data while accessing data or other materials through the Data Enclave. Data contained in the Data Enclave may contain errors due to sampling equipment, sampling method, data storage media, or data transfer method. Requestor agrees to assume all responsibility for interpreting the data correctly. Under no circumstances shall NORC be liable for any direct, indirect, incidental, punitive, special, or consequential damages resulting from: Requestor's use or inability to use the Data Enclave, training materials or the on-site facility.

10.2 Any Data Set disclosed pursuant to this Agreement is understood to be provided "AS IS." NORC MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED. THERE ARE NO EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE DATA SET WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS. Notwithstanding, NORC, to the best of its knowledge and belief, has the right and authority to disclose the Data Set to Requestor.

11.0 Miscellaneous

11.1 Merger. This Agreement constitutes the entire Agreement between the parties with respect to the subject matter contained herein, and all prior agreements, understandings and negotiations are merged into this Agreement. This Agreement may not be modified except in writing signed by the authorized representatives of the parties.

11.2 Waiver. Waiver of any breach of this Agreement shall be ineffective unless in writing signed by the party waiving compliance, and shall not be considered a waiver of any other breach.

11.3 Related Entities Bound. This Agreement shall be binding on the parties and their successors, assigns, affiliates, subsidiaries, officers, servants, employees, agents and representatives and all parties in active concert or participation with any of them.

11.4 Authority and Acknowledgement. The parties represent and warrant that they have full authority to enter into this Agreement. Requestor acknowledges and agrees that the Requestor have read and understood the terms and provisions of this Agreement and agrees to be bound by those terms and provisions in their individual capacity.

11.5 Severability. The determination that any provision of this Agreement is invalid or unenforceable shall not invalidate this Agreement or render other provisions unenforceable, and this Agreement shall be construed and performed in all respects as if the invalid or unenforceable provisions were omitted, insofar as the primary purposes of this Agreement are not impeded.

11.6 Governing Law. This Agreement shall be governed by and construed in accordance with the laws of Illinois, without regard to its conflict of laws rules, and shall be applicable worldwide.

11.7 Agreement Not Construed Against Drafter. The rule of construction that interprets contracts against the drafter shall not apply to this Agreement.

In signing this form, I affirm that I have read and agree to the above.

Requestor

Print Name: _____ Date: _____

Email: _____ Phone: _____

Signature: _____

For Internal Use Only

Date Received _____ Receipt Control No. _____

Time Received _____ NORC Staff

Initials _____