**Supplementary Information**

**Using Machine Learning for Suicide Risk Detection from Social Media**

\*Yaakov Ophir,1,2 Refael Tikochinski,1 Christa Asterhan,1 Itay Sisso,1 Roi Reichart2

1The Hebrew University of Jerusalem, 2Technion – Israel Institute of Technology

**Psycho-diagnostic Tools**

**Suicide risk.** Suicide risk was measured using the well-established Columbia Suicide Severity Rating Scale (CSSRS) (Posner et al., 2011). The CSSRS was originally developed to help clinicians structure their clinical interviews and assess the existence and severity of suicide risk with high levels of accuracy. The scale demonstrated high sensitivity and specificity scores in suicide prediction and it is considered a “diagnostic tool of choice,” both in clinical settings and in empirical research (Drapeau, Nadorff, McCall, Titus, Barclay, & Payne., 2019; Weber, Michail, Thompson, & Fiedorowicz, 2017). Upon consultation with the principal developer of the CSSRS (Posner, personal written communication), we chose to administer the electronic self-report version of the scale, in light of the fact that the current research examined participants from a crowdsourcing platform. The electronic version of the CSSRS has been demonstrated to have psychometric validity and prediction accuracies that are comparable to the original clinician version of the scale (Mundt, Greist, Gelenberg, Katzelnick, Jefferson, & Modell, 2010; Viguera et al., 2015).

The scale consists of six binary (yes/no) items that are presented to the participants in two parts. In the first part, participants were asked to complete Item 1 that addressed a “wish to be dead” *(“Have you wished you were dead or wished you could go to sleep and not wake up?”*) and Item 2 that addressed “suicidal thoughts” (*“Have you actually had any thoughts of killing yourself?”*). Only if participants answered “yes” to item #2 on suicidal thoughts, they were then exposed to the second part of the scale that examined the severity of the risk. Item 3 addressed suicidal thoughts with method (*“Have you been thinking about how you might kill yourself?”*). Item 4 addressed suicidal intent (*“Have you had these thoughts and had some intention of acting on them?”*). Item 5 addressed suicide intent with specific plan (*“Have you started to work out or worked out the details of how to kill yourself? Do you intend to carry out this plan?”*), and Item 6 addressed actual suicide behaviors (*“Have you ever done anything, started to do anything, or prepared to do anything to end your life?”*). Participants who answered “yes” to this last item were then asked to indicate when they engaged in such behavior (over a year ago, between three months and a year ago, or within the last three months).

The modular structure of the CSSRS enables the extraction of two binary (yes/no) variables: a *general risk of suicide* (participants who met the criterion of the first part of the scale, that is answering “yes” to item 2) and a *high risk of suicide* (a sub-group of participants at suicide risk who also responded “yes” to at least one of the items in the second part of the scale). The total sum score of the “yes” answers to all six items serves as another indication for the severity of the suicide risk.In this study, the total score of the CSSRS was positively correlated with all the examined risk factors (Table A) and especially with depression (*r* = 0.46), thus indicating a high convergent validity of the scale (for further details on the convergent validity of the various scales, see the *Supplementary Information*).

**Major Depressive Disorder (MDD).** Major depression was measured using the Patient Health Questionnaire-9 (PHQ-9) (Kroenke, Spitzer, & Williams, 2001), a nine-item scale that targets the nine symptoms of depression described in the DSM. Each item (symptom) is scored from 0 to 3 (not at all, several days, more than half the days, and nearly every day). The scale can be used both as a continuous measure (range = 0 - 27) that measures the severity of the depression and as a dichotomous measure to estimate the presence of depression (yes/no). The dichotomous cut-off point for the presence of depression corresponds with the DSM criteria and can be calculated when five or more symptoms receive a score of at least “more than half the days” and when these symptoms include one of the two key symptoms of depression: low interest and depressed mood (Spitzer, Kroenke, & Williams, 1999). Given its well-established validity and high sensitivity and specificity, the PHQ-9 is preferred over all other screening tools for depression (El-Den, Chen, Gan, Wong, & O’Reilly, 2018). The internal consistency of the scale in the current sample was high (α = .90) and the correlation with suicide total scores was high (*r* = 0.46).

**Generalized Anxiety Disorder (GAD).** GAD was measured using a well-established, seven-item scale named GAD-7. Each item, scored from 0 to 3 (not at all, several days, more than half the days, and nearly every day), targets one of the seven symptoms of the disorder. The total score of the scale (range 0 – 21) serves as an indication for both the existence and the severity of the disorder. According to the developers, the cutoff point for GAD is set to 10 points or higher (Spitzer, Kroenke, Williams, & Löwe, 2006). The internal consistency of the scale in the current sample was high (α = .92). The evidenced comorbidity between GAD and major depression as indicated in a bivariate Pearson, was very high (*r* = 0.76).

**Depressive rumination (brooding).** Depressive rumination as mentioned above is a maladaptive pattern of thinking in which people focus on their depressive feelings and enter a repetitive loop of negative thoughts (Mor, Hertel, Ngo, Shachar, & Redak, 2014; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008; Watkins, 2008). Specifically, the unconstructive component of this ruminative thinking, which has been shown to be strongly associated with depression was named “brooding” (Schoofs, Hermans, & Raes, 2010). Brooding was measured using five items rated from 1 (almost never) to 4 (almost always) from the frequently used Ruminative Responses Scale (RRS) (Nolen-Hoeksema & Morrow, 1991). Respondents read a general statement about depressive events (“*People think and do many different things when they feel depressed*”) and are asked to indicate to what extent they engage in a given response. An example for a brooding response is: “*Think about a recent situation, wishing it had gone better*.” The internal consistency of the 5 brooding items, in the current sample was good (α = .82) and the correlation with depression was high (*r* = 0.62).

**Excessive worrying.** A second pattern of negative thinking is excessive and subjectively uncontrollable worries about the future (Brown, Antony, & Barlow, 1992). To assess excessive worrying patterns, we used the Penn State Worry Questionnaire (PSWQ) (Meyer, Miller, Metzger, & Borkovec, 1990). The PSWQ is a well-established research tool (Fresco, Mennin, Heimberg, & Turk, 2003) that comprises 16 items, rated on a five-point scale (1 = not as all typical of me, 5 = very typical of me). The items address various aspects of pathological worry including its excessiveness (e.g., “*Many situations make me worry*”) and the subjective feeling of uncontrollability (e.g., “*Once I start worrying, I cannot stop*”). The internal consistency of the PSWQ in the current sample was high (α = .96) and the correlation with depression was high (*r* = 0.56).

**Loneliness.** Experiences of loneliness were measured using the 10-item version of the UCLA-Loneliness Scale (Russell, 1996). The items, rated from 1 (Never) to 4 (Always), encompass various aspects of loneliness experiences (e.g., “*How often do you feel that you lack companionship*”). This version of the scale demonstrated high levels of convergent validity and internal consistency (Elphinstone, 2018). The internal consistency of the scale in the current sample was high (α = .92) and the correlation with depression was high (*r* = 0.60).

**Low satisfaction with life.** The general sense of satisfaction with life was measured using the Satisfaction With Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985). This short scale comprises five items, rated from 1 (strongly disagree) to 7 (strongly agree). All items are formulated in a positive manner (e.g., “*The conditions of my life are excellent*”). Although we were interested in low satisfaction with life, we kept the original positive style of the scale to “break” the overall negative atmosphere of the research and to promote participants’ attentiveness along the research. The SWLS demonstrated good psychometric characteristics (Pavot & Diener, 2009) and moderate-strong negative relationships with depression (Blais, Vallerand, Pelletier, & Brière, 1989; Schimmack, Oishi, Furr, & Funder, 2004). The internal consistency of the scale in the current sample was high (α = .93) and the negative correlation of this positive scale with depression was high (*r* = -0.53).

**Personality traits.** Personality traits were assessed using the short version of the Big Five Inventory (BFI) (Rammstedt & John, 2007). The BFI-10 includes ten items that target the five clusters of personality traits, originally formulated in the standard 44-item BFI: Extraversion, Neuroticism, Openness to Experience, Agreeableness and Conscientiousness (John & Srivastava, 1999). Each trait in the BFI-10 is measured by only two items that are rated from 1 (disagree strongly) to 5 (agree strongly). The BFI-10 achieved high levels of reliability and validity (Rammstedt & John, 2007) and is currently widely used in research settings. Consistent with the literature on depression, the correlation between the personality trait of neuroticism and depression was high (*r* = 0.51).

The convergent validity of the psychosocial scales was high. As expected, the total score of the suicide scale was positively correlated with all the risk factors examined in the study and especially with depression (*r* = 0.46). Consistent with the literature on depression described above, the comorbidity between depression and anxiety was very high (*r* = 0.76) and the four psychosocial risk factors (i.e., brooding, excessive worry, loneliness, and low satisfaction with life) were strongly correlated with depression (Pearson’s *r* ranging from 0.53 to 0.62). The personality trait of neuroticism was also strongly correlated with depression (*r* = 0.51).

**Ethical Considerations**

Crowdsourcing-based suicide research involves an ethical challenge: how to safeguard the well-being of suicidal participants, without the possibility of face-to-face interactions? To address this ethical challenge, we adhered to an online suicide research protocol developed recently by an expert consortium (Ophir et al., under review). Prior to consenting to participate in the study, participants were informed that if their responses would indicate some form of suicidal risk, we would contact them through the data collection platform. Each participant who met the CSSRS criterion for general suicide risk (i.e., suicidal thoughts with or without a specific method or a concrete plan) then received a designated letter in which we encouraged them to seek help and provided them with a list of available “hotlines” and national mental health services. The complete description of the protocol and the ethical considerations made in the current research are available upon request.

**Data Quality**

In light of recent concerns regarding the quality of crowdsourcing-based data, we applied a newly developed rigid data quality assurance protocol (Ophir, Sisso, Asterhan, Tikochinski, & Reichart, 2019). The inclusion criteria were: having a Facebook account and having previous proven experience in MTurk-based studies. Proven experience was defined as past completions of at least 100 MTurk tasks, with a minimum of 95% success rate. To avoid bogus responses, we limited the participation to US residents and excluded users with suspicious Internet Protocol (IP) addresses (Prims, Sisso, & Bai, 2018). To ensure the quality of the unsupervised responses, we implemented a designated inattentiveness scale that comprised eight hidden attention checks. These checks included four types of data-quality measurements (i.e., “infrequency items,” “time measurements,” “person-total correlation,” and “long string analysis”), which were embedded in the various self-report scales of the study (Ophir et al., 2019).

**Architecture of the ANN-based Models**

As illustrated in Figure 1 of the main article, the Single-Task Model (STM) consisted of an input and an output layer, which are connected by a set of fully-connected layers. In contrast, the Multi-Task Model (MTM) contained three additional hierarchically organized auxiliary layers: Facebook content → personality traits → psychosocial risks → psychiatric disorders → suicide risk. As illustrated in Figure 2, each auxiliary layer is accompanied by a set of fully-connected layers, thus forming several “subnetworks.”

The subnetwork located at the bottom of the model (i.e., the Personality traits) is activated directly by the input layer (Facebook content), while the subnetworks at the middle (Psychiatric disorders and Psychosocial risks) are activated by the previous subnetwork’s output, which is concatenated with outputs from a shared set of fully-connected layers. The shared set of layers is activated directly from the input layer and allows the subnetworks to get direct information from the input layer (and not just from the previous subnetwork). This architecture introduces inductive bias to the suicide prediction model through the auxiliary tasks, while learning a shared set of parameters for the multiple tasks to reduce the risk for overfitting. Finally, the Suicide layer at the top of the model is activated by the output generated from the Psychiatric disorders layer and from the outputs of the shared set of hidden layers (Figure 2).

The loss function of the STM models is the *binary cross-entropy*:

Where *N* is the number of training examples, indicates whether participant *i* belongs to the suicide group () or not () according to the ground truth, and indicates the probability of as predicted by the model.

The loss function of the MTM is the sum of the output layer’s and the auxiliary layers’ loss functions:

Where is the binary cross-entropy loss function like before, and is the sum of all *mean squared errors* (MSEs) calculated for each of the auxiliary variables in the set *A*={Depression, Anxiety, Brooding, Worry, SWL, Lonely, Open, Conscientious, Extravert, Agreeable, Neurotic}:

where *N* is the number of training examples, is a continuous variable representing the ground truth score of the auxiliary-variable *a* for subject *i*, and is the predicted score for this variable according to the model.

The textual content of the Facebook postings was encoded using ELMo, a state-of-the-art ANN framework for “Embeddings from Language Models” (Peters et al., 2018). ELMo comprises a deep language model through multiple bi-directional Long-short-Term-Memory (LSTM) layers. ELMo has been shown to produce contextualized word embeddings that are more effective in many Natural Language Processing (NLP) tasks, compared to state-of-the-art non-contextualized embeddings such as Glove (Pennington, Socher, & Manning, 2014). Furthermore, ELMo is especially relevant to social media language. This is because ELMo is character-based (rather than word-based), thus allowing the system to make representations also to non-words (i.e. words that do not appear in formal dictionaries), as well as to expressions that did not appear in the learning phase. Using a pre-trained ELMo model (available at https://tfhub.dev/google/elmo/2), we extracted a 1024-dimensional embedding vector for each Facebook post in our data through mean-pooling over the contextualized word embeddings generated for the post. The overall textual-activity of the user was represented as the average of its post vectors. The resulting 1024-dimensional vector (per user) was then used as the input to the ANN models.

**ANN-based Models – Parameter Estimation (Learning)**

The optimization of the model was conducted with batch sub-gradient descent (batch-size of 32), using the back-propagation algorithm (Goodfellow, Bengio, & Courville, 2016) and the RMSProp optimizer (Tieleman & Hinton, 2012) with a momentum parameter of 0.9. The hyper-parameters of the models were tuned using a grid-search method. These hyper-parameters included the number of fully connected layers {1, 2, 3},[[1]](#footnote-1) the number of neurons in each layer {16, 32, 64, 128, 256, 512, 1024}, and the type of the activation function {*hyperbolic tangent*, *sigmoid*}. The hyper-parameters of the optimization algorithm were the learning rate {0.001, 0.005, 0.01, 0.05}, and the number of epochs {1000, 2500, 5000}.

The final hyper-parameters of the STM included: 3 fully connected layers, 32 neurons, an activation function of *hyperbolic tangent*, a learning rate of 0.01 with 2,500 epochs. The final hyper-parameters of the MTM included: 2 fully connected layers, 16 neurons, an activation function of *hyperbolic tangent*, a learning rate of 0.001 with 5,000 epochs.

**Four Possible Classes of Suicide Risk Predictions**

The ANN models produced binary (yes/no) predictions regarding the two (general/high) suicide risk variables for each Facebook user. These predictions were categorized into one of four possible classes: True Positive, in which a suicidal user is correctly detected (true) by the model as suicidal (positive); False Positive, in which a non-suicidal user is incorrectly detected (false) as suicidal (positive); True Negative in which a non-suicidal user is correctly determined (true) as not suicidal (negative); and False Negative in which a suicidal user is incorrectly determined by the model (false) as non-suicidal (negative).

**Figure A. Illustration of the hierarchical "pyramid" of risk factors for suicide**

Note: The bottom of the proposed pyramid consists of the big five personality traits (i.e., openness; conscientious; extraversion; agreeableness; and neuroticism). The middle layers consist of the psychosocial risk factors (i.e., depressive rumination, worries, loneliness, and low satisfaction with life) and the psychiatric disorders (i.e., depression and anxiety), and the top layer consists of the predicted output, which is the two types of binary suicide variables (i.e., general and high suicide risk).

**Table A.** Descriptive statistics and Correlations (*N* = 1,650).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Suicide | Depression | Anxiety | Brooding | Worry | SWL | Lonely | Open | Conscientious | Extravert | Agreeable | Neurotic |
| Means  (SD) | 0.8  (1.35) | 6.95  (5.93) | 13.62  (5.48) | 10.54  (3.52) | 49.42  (15.71) | 20.66  (8.14) | 23.42  (6.78) | 7.66  (1.98) | 7.64  (1.86) | 5.53  (2.38) | 6.94  (2.03) | 6.42  (2.44) |
| Depression | .459\*\* |  |  |  |  |  |  |  |  |  |  |  |
| Anxiety | .381\*\* | .760\*\* |  |  |  |  |  |  |  |  |  |  |
| Brooding | .390\*\* | .624\*\* | .648\*\* |  |  |  |  |  |  |  |  |  |
| Worry | .331\*\* | .566\*\* | .714\*\* | .645\*\* |  |  |  |  |  |  |  |  |
| SWL | -.360\*\* | -.534\*\* | -.449\*\* | -.458\*\* | -.423\*\* |  |  |  |  |  |  |  |
| Lonely | .384\*\* | .599\*\* | .508\*\* | .548\*\* | .490\*\* | -.607\*\* |  |  |  |  |  |  |
| Open | .072\*\* | -.005 | .020 | .009 | -.006 | -.012 | -.059\* |  |  |  |  |  |
| Conscientious | -.185\*\* | -.341\*\* | -.226\*\* | -.293\*\* | -.224\*\* | .269\*\* | -.302\*\* | .103\*\* |  |  |  |  |
| Extravert | -.179\*\* | -.259\*\* | -.236\*\* | -.209\*\* | -.287\*\* | .273\*\* | -.395\*\* | .143\*\* | .153\*\* |  |  |  |
| Agreeable | -.209\*\* | -.273\*\* | -.301\*\* | -.234\*\* | -.280\*\* | .262\*\* | -.351\*\* | .040 | .113\*\* | .199\*\* |  |  |
| Neurotic | .315\*\* | .506\*\* | .628\*\* | .561\*\* | .779\*\* | -.393\*\* | .468\*\* | -.061\* | -.289\*\* | -.323\*\* | -.304\*\* |  |

Note: Suicide = the total score of the CSSRS; SWL = Satisfaction With Life scale. Notice that the current research addressed low satisfaction with life whereas the SWL is formulated in a positive manner (i.e., high satisfaction with life). This positive formulation explains the negative correlation between SWL and depression.

**Table B.** Detection performance of the models using BERT

Table B provides a comparison of the results of the STM and MTM models, between the case where the text representation is made by ELMo (the text representation method that was employed in the main study) and the case where the text is represented by the recent attention-based BERT model (Bidirectional Encoder Representations from Transformers) (Devlin, Chang, Lee, & Toutanova, 2019). The similarities between the two cases included an equivalent range of AUC scores and improved predictions when the MTM, which included the theory-driven auxiliary factors, was applied (compared with the STM). The differences between the two cases included better BERT performance on the high risk group compared to the general risk group (an opposite phenomenon is observed with ELMo) and a weaker increase in BERT performance when moving from the entire sample to the Active users group, compared to the more stronger improvement observed with ELMo. The overall similar patterns indicate that our main conclusions, and particularly the one about the importance of theory-driven multi-task modeling for suicide risk prediction, are independent of the specific text representation method employed by the model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | General suicide risk | | High suicide risk | |
| Model | STM | MTM | STM | MTM |
| AUC for All users, using ELMo | .567 | .602 | .555 | .571 |
| AUC for All users, using BERT | .559 | .643 | .639 | .712 |
| AUC for Active users, using ELMo | .608 | .759 | .606 | .690 |
| AUC for Active users, using BERT | .584 | .695 | .668 | .724 |

Note: STM = Single Task Model; MTM = Multiple Tasks Model; AUC = Area Under the receiver operating characteristic Curve.

**Table C.** Term Frequency Inverse Document Frequency (TF-IDF)

The following table presents the hundred most frequent words that best distinguished between the four classes of prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **True Positive** | **False Positive** | **True Negative** | **False Negative** |
| 1 | anymore | president | lord | fight |
| 2 | lose | husband | shall | ufb |
| 3 | kinda | season | blessed | office |
| 4 | probably | games | king | daddy |
| 5 | mad | gun | gift | fire |
| 6 | poor | strong | james | system |
| 7 | positive | wonder | christ | email |
| 8 | room | cat | church | experience |
| 9 | pretty | continue | missed | nobody |
| 10 | sitting | update | drink | shows |
| 11 | top | news | loving | ufd |
| 12 | cheese | damn | anybody | just |
| 13 | air | paid | god | america |
| 14 | eating | writing | monday | war |
| 15 | sick | none | wedding | none |
| 16 | okay | words | faith | questions |
| 17 | pain | talking | christmas | king |
| 18 | cry | change | jesus | ya |
| 19 | actually | ass | coffee | etc |
| 20 | instead | hot | prayer | pray |
| 21 | seriously | cold | however | like |
| 22 | easy | test | prayers | day |
| 23 | bit | watched | version | bus |
| 24 | reason | within | history | prayers |
| 25 | game | children | kill | one |
| 26 | clothes | fb | wishes | can |
| 27 | mother | asking | answer | happy |
| 28 | daughter | knew | state | know |
| 29 | hurt | grow | comment | get |
| 30 | worst | cancer | law | whats |
| 31 | bad | blood | important | will |
| 32 | account | shit | heaven | time |
| 33 | felt | red | mental | five |
| 34 | theyre | yesterday | father | texas |
| 35 | wow | dad | dr | sister |
| 36 | enjoy | holiday | ready | soul |
| 37 | lady | country | son | running |
| 38 | cut | near | pass | street |
| 39 | sleep | look | spirit | coming |
| 40 | fuck | loved | ufc | green |
| 41 | supposed | weeks | email | yeah |
| 42 | bed | looked | peace | american |
| 43 | pizza | funny | line | fine |
| 44 | quite | text | thanksgiving | along |
| 45 | gets | around | perfect | attention |
| 46 | guess | support | ufb | close |
| 47 | drive | relationship | war | human |
| 48 | door | course | save | died |
| 49 | thinking | putting | lol | company |
| 50 | surgery | wanted | together | problems |
| 51 | gonna | seems | dear | aint |
| 52 | literally | found | thank | business |
| 53 | thats | couple | just | prayer |
| 54 | sleeping | world | great | share |
| 55 | bitch | daily | sunday | area |
| 56 | cream | john | happy | listen |
| 57 | heart | looks | working | bit |
| 58 | wonderful | ask | fall | water |
| 59 | big | several | day | truly |
| 60 | arent | posts | cause | missing |
| 61 | might | moving | st | go |
| 62 | fucking | half | given | character |
| 63 | hospital | age | today | page |
| 64 | told | seeing | help | people |
| 65 | sad | company | men | store |
| 66 | doesnt | longer | love | love |
| 67 | couldnt | kept | choose | local |
| 68 | wall | months | friends | group |
| 69 | favorite | tv | brother | learned |
| 70 | taking | florida | others | retweeted |
| 71 | cleaning | high | holy | send |
| 72 | stupid | others | time | busy |
| 73 | nap | away | like | good |
| 74 | ugh | given | city | gonna |
| 75 | start | kids | everyone | song |
| 76 | entire | women | giving | now |
| 77 | brain | place | roll | anybody |
| 78 | car | hand | please | im |
| 79 | wear | sit | florida | park |
| 80 | times | weekend | mother | sorry |
| 81 | dinner | hey | copy | see |
| 82 | play | fear | question | pass |
| 83 | story | run | wife | gas |
| 84 | Isnt | voice | know | birthday |
| 85 | calling | dark | friend | wishes |
| 86 | white | called | forget | asking |
| 87 | spent | spend | cleaning | end |
| 88 | mind | eyes | teacher | books |
| 89 | online | learn | get | lives |
| 90 | hopefully | light | group | miss |
| 91 | ice | history | pray | everyone |
| 92 | making | true | child | jesus |
| 93 | rest | saw | can | missed |
| 94 | feet | body | go | dark |
| 95 | order | heard | known | figure |
| 96 | cute | open | safe | wedding |
| 97 | understand | happened | one | entire |
| 98 | type | move | busy | today |
| 99 | sure | write | american | came |
| 100 | summer | due | lets | forget |

Note: This table presents the hundred most frequent words that best distinguished between the four classes of prediction (True Positive, True Negative, False Positive, and False Negative), using *Term Frequency Inverse Document Frequency (TF-IDF)*.

**References**

Blais, M. R., Vallerand, R. J., Pelletier, L. G., & Brière, N. M. (1989). L’échelle de satisfaction de vie: Validation canadienne-française du “Satisfaction with Life Scale.” *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *21*, 210.

Brown, T. A., Antony, M. M., & Barlow, D. H. (1992). Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample. *Behaviour Research and Therapy*, *30*, 33–37.

Devlin, J., Changm M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.

Drapeau, C. W., Nadorff, M. R., McCall, W. V., Titus, C. E., Barclay, N., & Payne, A. (2019). Screening for suicide risk in adult sleep patients. *Sleep Medicine Review,* *46*, 17–26.

El-Den, S., Chen, T. F., Gan, Y. L., Wong, E., & O’Reilly, C. L. (2018). The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. *Journal of Affective Disorders, 225*, 503-522.

Elphinstone, B. (2018). Identification of a suitable short‐form of the UCLA‐Loneliness Scale. *Australian Psychologist*, *53*, 107–115.

Fresco, D. M., Mennin, D. S., Heimberg, R. G., & Turk, C. L. (2003). Using the Penn State Worry Questionnaire to identify individuals with generalized anxiety disorder: A receiver operating characteristic analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *34*, 283–291.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In O. P. John & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (2nd ed.) (pp. 102–138)*.* New York: Guilford Press.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy*, *28*, 487–495.

Mundt, J. C., Greist, J. H., Gelenberg, A. J., Katzelnick, D. J., Jefferson, J. W., & Modell, J. G. (2010). Feasibility and validation of a computer-automated Columbia-Suicide Severity Rating Scale using interactive voice response technology. *Journal of Psychiatric Research*, *44*(16), 1224-1228.

Mor, N., Hertel, P., Ngo, T. A., Shachar, T., & Redak, S. (2014). Interpretation bias characterizes trait rumination. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 67–73.

Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality and Social Psychology*, *61*, 115–121.

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*, 400–424 .

Ophir, Y., Amichai Hamburger, Y., Brunstein Klomek, A., Levi-Belz, Y., Hadlaczky, G., Yom-Tov, E., Zalsman, G., &amp; Gould, M. (under review). The ethics of suicide research online: A consensual protocol for crowdsourcing-based studies on suicide.

Ophir, Y., Sisso, L., Asterhan, C. S. C., Tikochinski, R., & Reichart, R. (2019). The Turker blues: Hidden factors behind increased depression rates among Amazon’s Mechanical Turkers. *Clinical Psychological Science*, *8*(1), 65-83. https://doi.org/10.1177/2167702619865973

Pavot, W., & Diener, E. (2009). Review of the satisfaction with life scale. In E. Diener (Ed.) *Assessing well-being* (pp. 101–117). New York: Springer.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP) (pp. 1532–1543).

Peters, M. E., Neumann, M., Lyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, Z. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365.*

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., et al. (2011). The Columbia–Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry, 168*(12):1266–77.

Prims, J. P., Sisso, I., & Bai, H. (2018). Suspicious IP online flagging tool. Retrieved on October 28, 2019, from https://itaysisso.shinyapps.io/Bots.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203–212.

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, *66*, 20–40.

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin*, *30*, 1062–1075.

Schoofs, H., Hermans, D., & Raes, F. (2010). Brooding and reflection as subtypes of rumination: Evidence from confirmatory factor analysis in nonclinical samples using the Dutch Ruminative Response Scale. *Journal of Psychopathology and Behavioral Assessment*, *32*, 609–617.

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Patient Health Questionnaire primary care study. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*, *282*, 1737–1744.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*, 1092–1097.

Tieleman, T. & Hinton., G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, *4*(2), 26–31.

Viguera, A. C., Milano, N., Laurel, R., Thompson, N. R., Griffith, S. D., Baldessarini, R. J., & Katzan, I. L. (2015). Comparison of electronic screening for suicidal risk with the patient health questionnaire item 9 and the Columbia suicide severity rating scale in an outpatient psychiatric clinic. *Psychosomatics*, *56*(5), 460-469.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological Bulletin*, *134*, 163.

Weber, A. N., Michail, M., Thompson, A., & Fiedorowicz, J. G. (2017). Psychiatric emergencies: Assessing and managing suicidal ideation. *Medical Clinics of North America, 101*(3), 553–71.

1. Note that all the sub-networks of the MTM had the same number of fully connected layers. [↑](#footnote-ref-1)