Profiling Depression in Neutral Reddit Posts

Prediction-Insight Tradeoffs and Mental Health Technology Applications

Molly E. Ireland
Department of Psychological Sciences
Texas Tech University
Lubbock, TX, USA
molly.ireland@ttu.edu

Jonathan Schler & Gilad Gecht
Department of Computer Science
Holon Institute of Technology
Holon, Israel
schler@hit.ac.il, giladge@my.hit.ac.il

Kate G. Niederhoffer
Independent Researcher
Austin, TX, USA
k.niederhoffer@gmail.com

ABSTRACT

Attempts to predict mental health vulnerabilities in everyday social media messages have proliferated rapidly over the last decade. It remains unclear how mental health technology providers or clinicians can use predictive linguistic models from such research to serve their clients. To explore methods of profiling mental health status based on language used in a neutral setting (i.e., where people are not discussing mental health), we used an array of machine learning and regression models to classify posts from neutral Reddit forums as written by depressed users (self-identified) or random controls (N = 1,121). Predictive linguistic features included dictionary-based categories (e.g., LIWC variables), parts of speech, and character n-grams. The best-performing classifier was a convolutional neural network model, which is not interpretable in a traditional sense by practitioners or mental health technology users. We discuss the possibilities of reverse-engineering psychological insights from "black box" machine learning models and otherwise using these models to advance theory in behavioral science. Finally, we consider ethical and practical considerations of our findings for mental health technology applications.

CONCEPTS

• Applied computing • Human-centered computing • Computing methodologies

KEYWORDS

• Applied computing • Human-centered computing • Computing methodologies

1 Introduction

For practitioners and researchers in mental health technology, balancing prediction and insight is a perpetual struggle with far-reaching ethical and practical implications. Both prediction (of a mental health event or status) and insight (into why that event is taking place) are clearly of use to mental health professionals, and neither is sufficient on its own, in terms of advancing science or fulfilling ethical obligations to clients or users. Using a decade-long record of depression forum users and presumed non-depressed comparison users as a test bed, the current project explores how to navigate this trade-off using a combination of neural network approaches and more interpretable logistic regression models. We then speculate on the possibilities and pitfalls of reverse-engineering insights from opaque machine learning models. Last, the ethical responsibilities of mental health technology developers and practitioners are discussed in the context of prediction-insight tradeoffs in basic and applied research in this field.

Data-driven methods from computer science and computational linguistics are increasingly mainstream, with machine learning models being used across diverse fields to predict outcomes ranging from movie preferences [1] to healthcare system burdens [2]. As deep learning and other neural networks become more accessible to people without computer science training, including psychologists and mental health practitioners, prioritizing prediction over insight has gained traction. If the aim of behavioral science is to predict real-life behavior, then prediction—even in cases where the underlying model is a black box that cannot be opened or interpreted—should have primacy over theoretical insight (see [3]). The standalone value of prediction is particularly salient in cases where the behavioral outcome has relevance to life-or-death outcomes, such as mental health crises (e.g., suicide attempts or substance use relapse [4]), disease outbreaks [5], or crime [6]. That is, if an event can be predicted, it may be possible to prevent it with interventions, even if the predictive features in the model are not known or knowable. More broadly, mental health practitioners may be interested in predicting more quotidian clinically-relevant events that may be bellwethers of future crises, such as discontinuing use of some mental health treatment (e.g., therapy or a mental health technology app).

Aside from the practical case for valuing prediction over insight, there are historical and empirical reasons for believing that observational data provides a better foundation for theory than hypotheses. A tradition of focusing on theories over naturalistic data (or insight over exploration) in psychology may have contributed to the ongoing replication crisis in psychology. In much of the 20th and early 21st century, unexpected psychological findings were shoehorned into existing theories post hoc, with authors claiming to have predicted what they found based on theory—a practice that ironically implicated exploratory research, and not incorrect theories, as the culprits behind reproducibility problems [7]. As early as Francis Bacon's work in the 17th century, philosophers and scientists have observed that beginning empirical research with theory rather than inductive observation tends to produce dogmatic research and confirmation biases [3, 8]. Indeed, a key cause of the replication crisis in psychology may have been researchers' tendency to believe established theories over their own and others' data, leading the field to discredit null results that, had they been published, would have helped falsify incorrect theoretical models [9, 10]. Thus, the replication crisis and subsequent reproducibility movement in psychology have led some to argue for more precise theory-testing [11], and others to propose that we prioritize prediction—even in cases where theoretical explanations for model-based predictions are not immediately clear [3, 12].

Mental health technology practitioners tend to be more concerned with responsibilities to their users or clients than theory development. The aims of mental health technology applications vary across platforms, but include reduced psychological distress (e.g., fewer or less severe symptoms), increased well-being, and improved social support, often achieved by providing a space for affiliation with others who share mental health conditions or life experiences [13]. Theory-based research has the complication, in applied mental health interventions, of being tied to specific measures (e.g., validated scales) that are easier to use in laboratory studies than the field. Naturalistic user data from mental health technology websites and apps show that a low proportion of users are willing to complete standard psychometrically validated self-report assessments of long-term mental health outcomes over multiple time points [14]. Predictive models based on rich naturalistic data will typically forecast mental health events (e.g., crises or turning points, such as suicide attempts [15]) better than sparse or absent survey data, regardless of how sound the surveys' underlying theories might be.

Clinical psychology, on the other hand, traditionally is more oriented towards explanatory theories than predictive models based on data alone. Whereas the goal with mental health technology platforms (e.g., apps, websites) may be to reach some threshold of accuracy for predicting users' mental health crises [16], clinicians tend to operate face-to-face and one-on-one, basing diagnostic decisions on case histories and structured clinical interviews rather than individual out-of-context statements [17]. Practicing therapists in particular may find limited value in opaque predictive models that cannot offer insights into human behavior or help reveal the mechanisms underlying changes in mental health. Focusing on interpretation and applications does not entail that theory always trumps data in clinical psychology, however. Indeed, many clinical psychologists see limited use in the fine-grained diagnostic categories and checklists of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5 [17]), preferring instead to use a transdiagnostic approach that emphasizes common mechanisms (e.g., affect dysregulation) over specific disorders [18]. That approach is broadly consistent with the perspective, gaining traction in computational social science [19] and social psychology [20], that the presumed ground truth of surveys will soon be replaced by quantitative measures of everyday behaviors, such as posting on social media or online forums, like Reddit.

Social psychology—the study of social influence in non-clinical populations—and computational social science tend to occupy a middle ground, integrating theory-based and data-driven methods. As the ongoing COVID-19 pandemic normalizes social isolation and associated depression, anxiety, and suicide ideation symptoms [21], social psychology is increasingly relevant to mental health—a realm previously considered strictly the jurisdiction of clinical or abnormal psychology. And as the mental health crises related to both COVID-19 and the George Floyd and Black Lives Matter movements become more salient, there is an increasingly urgent need for practical, user-friendly strategies for gaining both causal insights and near-future predictions from readily available naturalistic data, such as social media messages.

The following project uses an interdisciplinary computer science and social psychological approach, which focuses on maximizing predictive accuracy while at the same time considering the relevance of this work to theory-building and mental health technology applications. Specifically, in a sample of Reddit users posting over the course of 10 years, we used a variety of logistic regression and neural network models to predict whether a post from a neutral forum (i.e., not about depression or mental health) was written by a self-identified depressed individual or a randomly selected control user who had never posted in a depression forum.

2 Method

Data collection and analysis followed the following steps:

Step 1: *Data collection*. Using a Python-based crawler, we gathered a total of 303,649 posts from 16,617 subreddits representing a total of 11,630 users (averaging 26.1 posts per user) over 10 years (February 2009-February 2019).

Step 2: *Classifier Creation.* In order to detect depression-related posts as compared to neutral posts, we harvested 7,115 depressed posts (using the original post only, excluding replies and comments) from r/Depression and 10,000 non-depressed posts from r/AskReddit, representing prototypical depressed and neutral forums, respectively. We then built a classifier to distinguish between those categories (see Table 1).

Notably, r/AskReddit, the prototypical neutral forum, is not entirely positive. Rather, it covers a range of topics that site users are concerned or curious about, which at times includes daily life stressors (e.g., family conflict, pet peeves), but does not specifically or frequently deal with topics such as diagnosed mental health conditions. The aim was to select a comparison forum that represented the range of positive and negative conversational topics that a person would encounter on Reddit outside of a dedicated mental health forum.

Table 1. Neutral and Depression-Related Forum Examples

Depressed	Neutral
r/depressed	r/nocontextpics
r/depression	r/hardwareswap
r/depression_help	r/androidapps
r/ForeverAloneDating	r/todayilearned
r/depressionregimens	r/tipofmytongue

We used bag-of-words features and applied logistic regression, random forests, and an SVM classifier which achieved the following results (Figure 1; we randomly split the training and test data according to size in the chart). Overall precision was 0.88 and recall 0.91 for the logistic regression classifier.

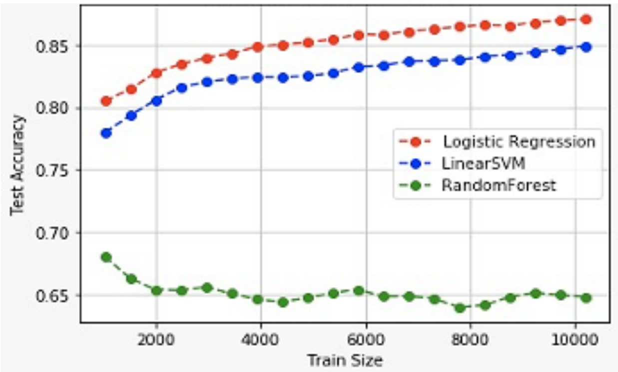

Figure 1. Model accuracy comparisons as a function of training sample size.

Step 3: *Post classification.* Based on the classifier built in step (2), we analyzed all subreddits (i.e., forums) with over 50 posts and classified the posts. Forums with at least 80% of posts classified as depressed were tagged as depressed forums, and those with fewer than 20% of posts classified as depressed were categorized as neutral forums (Table 2).

Note that many forums had between 20% to 80% posts classified as depressed and hence were not categorized either way, including mental health support forums (e.g., r/ADHD) and others that are not directly depression-related but are not neutral (e.g., in the latter group, r/me_irl commonly involves emotional self-disclosures that may be relevant to depression).

Table 2. Predicted and Actual Post Classifications

	Predicted Neutral	Predicted Depressed
Actual Neutral	8,912	1,088
Actual Depressed	1,130	5,984

Step 4: *Depressed user identification*. We identified users who posted in subreddits that were tagged as depressed in step (3) and identified themselves as depressed (i.e., stated "I was diagnosed as depressed" or "my depression," or mentioned explicit depression treatments), resulting in 1,593 users.

Step 5: *Collecting neutral posts by depressed individuals.* For the 1,593 users we found in step (4), we identified all their posts in other forums, including neutral forums (as detected in step 3). Of the users previously operationalized as depressed (step 4), 938 had also posted in a neutral subreddit. For these users, we collected 2,981 posts from neutral forums

Step 6: *Neutral user identification.* We gathered a random sample of 183 users who did not post in depression forums, had more than 12 posts on average, and also posted in the same neutral forums in which the depressed users had posted. We collected 2,822 posts from these users, resulting in the following dataset of users and posts from neutral forums (Table 3):

Table 3. Sample Sizes for Depressed and Control Users

	Neutral Users	Depressed Users
Individuals	183	938
Posts	2,822	2,981

Step 7: *Neutral post classification*. The aim was to build a classifier that can classify whether neutral posts (from a neutral forum) were written by a depressed person or not. To achieve this, we took the posts from step 6 (all written in neutral forums) and built a classifier that identifies author type (depressed or neutral). In order to prevent bias due to the difference in the number of users from each group, we balanced the groups to represent 140 random users from the depressed group (1,441 posts) and 134 from the neutral group (1,425 posts). Approaches included bag-of-word (top 10000 unigrams), Linguistic Inquiry and Word Count (LIWC) variables [22], character n-gram ($n = 3$, and character level sliding window $n = 150$ for the convolutional neural network, or CNN) all features were normalized to t-distribution such that $\mu = 0$ and $\sigma = 1$. We applied classifiers using logistic regression (BGFS solver and up to 1,000 iterations), random forest (100 estimators with gini function and tree pruning for depth and items per node), SVM (RBF kernel and $C = 1$), and CNN (3 convolutional layers, Adam optimizer and activation function). We ran a 10-fold cross-validation, where in each split, any given author (and that author's posts) were either in the training or test set. Results are shown in Table 4; other classifiers performed similarly to CNN.

Note that rather than expanding the neutral user sample, we reduced the depressed dataset to match the neutral sample. Future work will address that limitation by adding neutral user data.

Table 4. Precision/Recall for Logistic Regression and CNN

	Logistic Regression	CNN
Bag of words	0.51 / .51	n/a
BOW+LIWC	0.55 / 0.44	n/a
Char-n-gram + LIWC	0.41 / 0.61	0.62 / 0.88*

Note: CNN = convolutional neural network. *Without LIWC features but with character n-grams.

3 Discussion

Results showed that a CNN model using a character-level sliding window ($n = 150$) outperformed more standard logistic regression using traditional predictive language features (bag of words, LIWC, or character n-grams). Note that logistic regression and similar classifiers do not work well on large n-grams but rather use word-based features. Whereas theory-consistent linguistic markers of depression representing affective words (e.g., *sad*, *alone*) or cognitive processes (self-focus, e.g., *I*, *me*) would have helped explain why a given user's posts classified them as depressed or not [23], this more precise model, instead, leaves us wondering what led the model to perform as well as it did.

There are at least two main approaches for reverse-engineering theoretical insights from classifiers such as CNN, each with costs and benefits. First, it may be possible to correlate interpretable language features with the classifier to determine what cues it is focusing on (e.g., correlating all LIWC variables with a model’s predictions). Similarly, intuitive insights could be gained by asking clinicians to examine representative texts from the four cells in the classification model (Table 2). For experts trained in diagnosing mental health conditions, these texts may reveal patterns that are not apparent to either text analysis programs or researchers without clinical experience. Both of these quantitative and qualitative methods of reverse-engineering insights from black box models run the risk of identifying variables that are merely confounded with the true predictors, potentially leading researchers to base new models on psychometric red herrings.

In cases where it is not feasible or advisable to read the proverbial tea leaves of a machine learning-based classifier, a third option is to use such models to either falsify or qualify existing theoretical models. If a theory-based predictor of depression correlates with depression but fails to provide incremental validity beyond a machine learning-based classifier, then the underlying theory may be inadequate. Along these lines, recent research in psychology has shown that previously found links between self-referential language (specifically the pronouns “I” and “me”) and depression are better accounted for by a broader underlying trait—neuroticism, or negative emotionality—which may be a common factor in many emotional disorders [24, 25]. In the same way that those null results have advanced psychological theory, finding that opaque classifiers outperform intuitive models based on clinical observations and theory-driven research should drive psychologists to improve their theories [3]—even if it means returning to the first step of the scientific method (“observe”), embracing exploratory research [8, 12], and recognizing that mental health in a digital setting could manifest in different ways than in the traditional, offline clinical setting. Indeed, in the future, classifier models built on naturalistic data (e.g., Reddit posts, naturalistic recordings of conversations) may lead mental health experts to redefine mental health conditions, anchoring operationalizations more closely to lived experiences.

More broadly, the methods used in the present study—classifying neutral texts as being written by a person who is depressed or not—have the potential to improve psychological theories of depression by making their predictions more domain general. Basing models of depression primarily on behavior in theoretically relevant settings, such as therapy sessions or theory-based experiments, runs the risk of producing models that are limited to those specific contexts. Analyzing how people who identify as depressed behave in neutral conversations—for example, while inquiring about home improvement advice or commenting on video games—provides valuable naturalistic data that stands to improve the coverage of psychological theories of depression and other affective disorders. That is, independent of the practical applications of the deep learning models that performed best in the present study, we suggest that psychological theory would be improved by paying increased attention to how mental health conditions manifest in conversations that are not about mental health. Within-person comparisons also provide critical data on how an individual’s behavior changes across mental health relevant and neutral contexts, potentially revealing ways in which people attempt to (successfully and unsuccessfully) conceal or suppress mental health symptoms, such as anxiety [26].

The superiority of neural network models in the present data may relate to the nature of the classification task. Past studies of linguistic cues to depression have typically analyzed relatively long texts (e.g., essays, diaries) and measured depression symptoms on a continuous scale [27]. The task of classifying individual Reddit posts is complicated by both sparse data (short posts) and the binary outcome (depressed or not depressed). For example, one of the most face-valid predictors of depression in past work, negative emotional words (e.g., *depressed*, *crying*), tend to be used infrequently [28], especially outside of the context of close relationships [29, 30]. Thus, negative affect may have limited use in classifier models based on individual messages posted in public settings, like Reddit, even if it is a valid depression cue in other contexts.

The ability to generalize from the present data to clinical samples is likely limited by our focus on people who self-identified as depressed in online forums and thus may not meet standard clinical criteria for depression [17]. However, by analyzing the naturalistic, everyday behavior of people outside of a clinical population, these data may better represent the general population than samples in randomized clinical trials. Predictive models from this project may be of particular use to mental health technology providers who aim to prevent major depressive disorder before a first depressive episode or treat subclinical depressive symptoms outside of traditional psychotherapy settings. Supporting the generalizability of this approach, evidence from other fields suggests that effect sizes and replicability from observational studies are similar to those of randomized clinical trials [31], especially when cross-validation is used to estimate reproducibility [3].

Unobtrusive depression detection in a digital community has promise as a means to direct users to appropriate supports. This is a particularly favorable prospect when it involves consenting individuals in a dedicated context (e.g. depression subreddits, anxiety apps), with some transparency as to the means of identification or how a given person “scored” in a depressive range. The implications of a non-consenting individual in a neutral forum being identified as potentially depressed by an incomprehensible neural network model are problematic, even with the best of intentions. Comparable efforts by Facebook to identify at-risk users based on the site’s usage data have been ethically fraught [32]. Beyond issues related to consent, being classified by an unknown algorithm as part of a stigmatized group (e.g., suicidal, depressed) may harm users through self-fulfilling prophecies or self-stigma [33]. One promising implication of this work relates to lightweight micro-therapeutic interventions in the form of assists from digital agents [34]. In the same way Google’s Smart Reply offers potential response options in email, if a model were to unobtrusively identify a high probability of depression, digital agents could be empowered to provide users with messages regarding self-efficacy and skills-based resources, or otherwise guide the user to disclose emotions and identify support needs.

Carrying the potential of AI further, even an optimal deep learning-guided mental health application that individuals knowingly consent to use—for example, a digital phenotyping or "personal sensing" phone application that, based on individuals' language use and application activity, provides support itself or guides a person to exactly the support that they need, when then need it—may have ethical shortcomings [34]. People tend to thrive when they feel that they have control over their lives and are capable of helping themselves [35, 36]. Furthermore, people tend to build resilience and coping skills by overcoming daily life stressors [37]. Thus, even in a hypothetical utopian future, where applications of AI in mental health technology are optimally accurate, this technology may decrease individuals' perceived self-efficacy and ability to cope with distress. A truly optimal AI application for mental health monitoring and interventions may therefore be one that takes these concerns into consideration by building in mechanisms to help individuals build coping skills, self-confidence, and self-control, rather than giving users the impression that they are being sheltered or controlled. These goals could likely only be met under the governance of common-sense informed consent that users are fully aware of at all times, communicated in ways that make the intent and practice clear for all individuals. To ensure that users remain aware of what they have consented to over time, in the case of long-term use of services based on changing technology, it may be necessary to adopt a dynamic consent model that maintains an ongoing dialogue between the people who develop and use mental health technology [38].

Finally, despite cases like the present study where AI models outperform and have the potential to improve human theories, it is important to remember that theory can provide guidance to and scaffold machine learning models as well. Machine learning models are of course only as good as the data on which they are based. Theoretical models from the behavioral and social sciences can help to narrow the search space for both predictive features and data sources, increasing the likelihood that machine learning models will be able to predict outcomes efficiently and reliably in the relatively small samples that mental health care providers and researchers deal with most frequently (e.g., clients in psychotherapy, participants in a research experiment, or users of a mental health technology app). Furthermore, theories from psychology and other behavioral sciences, such as communication and sociolinguistics, can help to make sense of contextual moderators of model-based predictions. For example, theoretical insights may be able to provide predictions regarding how the ongoing coronavirus pandemic or Black Lives Matter movement will affect models based on data collected before these cultural events.

Prior to the widespread implementation of highly precise yet incomprehensible models in digital mental health, policymakers, developers, and designers should be aware of their responsibilities in the same way that researchers adhere to ethical standards of respect, beneficence, and justice [39]. If not the precise signals, mental health technology designers should ensure a means to communicate the way a model works to a user with a naive theory of a model, akin to the Sorting Hat in *Harry Potter* explaining its classification process to first-year students [40]. Absent such explanations, the promise of computational linguistic machine learning models may remain largely siloed in computer science—and using such methods in mental health technology may inflame users' fear of data-driven algorithms rather than encouraging them to seek evidence-based mental health support.

ACKNOWLEDGEMENTS

We are grateful for the insightful and detailed advice of two anonymous reviewers, and we thank Tyler Davis for his assistance with both the paper and the pre-recorded workshop presentation.

REFERENCES

[1] Blake Hallinan and Ted Striphas. 2016. Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media Soc.*, *18*, 1, (January 2016), 117-137. DOI: 10.1177/1461444814538646

[2] J. Madison Hyer, Aslam Ejaz, Diamantis I. Tsilimigras, Anghela Z. Paredes, Rittal Mehta, and Timothy M. Pawlik. 2019. Novel machine learning approach to identify preoperative risk factors associated with super-utilization of Medicare expenditure following surgery. *JAMA Surgery, 154,* 11, (November 2019), 1014-1021. DOI: 10.1001/jamasurg.2019.2979.

[3] Tal Yarkoni and Jacob Westfall. 2017. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.*, *12*, 6, (November 2017), 1100-1122. DOI: 10.1177/1745691617693393

[4] Nicholas B. Allen, Benjamin W. Nelson, David Brent, and Randy P. Auerbach. 2019. Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science provide a breakthrough? *J. Affect. Disord.*, *250*, (May 2019), 163-169. DOI: 10.1016/j.jad.2019.03.044

[5] Cuilian Li, Li Jia Chen, Xueyu Chen, Mingzhi Zhang, Chi Pui Pang, and Haoyu Chen. 2020. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Euro. Surveill.*, *25*, (March 2020), 2000199. DOI: 10.2807/1560-7917.ES.2020.25.10.2000199

[6] Hyeon-Woo Kang and Hang-Bong Kang. 2017. Prediction of crime occurrence from multi-modal data using deep learning. *PloS one*, *12*, 4, (April 2017), e0176244. DOI: 10.1371/journal.pone.0176244

[7] Stanley B. Klein. 2014. What can recent replication failures tell us about the theoretical commitments of psychology? *Theory & Psychology, 24,* (June 2014), 326–338. DOI: https://doi.org/10.1177/0959354314529616

[8] Francis Bacon. 1620/2000. *Novum Organon.* Cambridge University Press.

[9] David C. Funder. 2014. The real source of the replication crisis (blog post). (July 2014). Retrieved May 16 from https://funderstorms.wordpress.com/ page/2/.

[10] David C. Funder and Daniel J. Ozer. 2019. Evaluating effect size in psychological research: Sense and nonsense. *Adv. Methods. Pract. Psychol. Sci.*, *2*, (May 2019), 156-168. DOI: 10.1177/2515245919847202

[11] Klaus Oberauer and Stephan Lewandowsky. 2019. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.*, *26*, (Oct 2019), 1596-1618. DOI: 10.3758/s13423-019-01645-2

[12] Anne M. Scheel, Leonid Tiokhin, Peder M. Isager, and Daniel Lakëns. 2020. An Agenda for Exploratory Research. Manuscript under review. Human-Technology Interaction Group, Eindhoven University of Technology, Eindhoven, Netherlands.

[13] Mark Erik Larsen, Kit Huckvale, Jennifer Nicholas, John Torous, Louise Birrell, Emily Li, and Bill Reda. 2019. Using science to sell apps: Evaluation of mental health app store quality claims. *Digit. Med., 2,* 18 (March 2019). https://doi.org/10.1038/s41746-019-0093-1

[14] Theresa Fleming, Lynda Bevin, Mathijs Lucassen, Karolina Stasiak, Sarah Hopkins, and Sally Merry. 2018. Beyond the trial: Systematic review of real-

world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. *J. Med. Internet Res., 20*, (June 2018), e199. DOI: 10.2196/jmir.9275

[15] Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomed. Inform. Insights, 10*, (August 2018), 1-11. DOI: 10.1177/1178222618792860

[16] Allura L. Ralston, Arthur R. Andrews III, and Debra A. Hope. 2019. Fulfilling the promise of mental health technology to reduce public health disparities: Review and research agenda. *Clin. Psychol., 26,* (January 2019), e12277. DOI: 10.1111/cpsp.12277

[17] American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub.

[18] Tim Dalgleish, Melissa Black, David Johnston, and Anna Bevan. 2020. Transdiagnostic approaches to mental health problems: Current status and future directions. *J. Consult. Clin. Psychol.,* 88, 3, (March 2020), 179. DOI: 10.1037/ccp0000482

[19] Wu Youyou, David Stillwell, H. Andrew Schwartz, and Michal Kosinski. 2017. Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends. *Psychol. Sci. 28*, 3 (March 2017), 276-284. DOI: https://doi.org/10.1177/0956797616678187

[20] Ryan L. Boyd and James W. Pennebaker. 2017. Language-based personality: A new approach to personality in a digital world. *Curr. Opin. Behav. Sci., 18,* (December 2017), 63-68. https://doi.org/10.1016/j.cobeha.2017.07.017

[21] Nicola Montemurro. 2020. The emotional impact of COVID-19: From medical staff to common people. 2020. *Brain Behav. Immun.,* 20, (March 2020), 30411-6. Advance online publication. DOI: https://doi.org/10.1016/j.bbi.2020.03.032

[22] James W. Pennebaker, Roger J. Booth, Ryan L. Boyd, and Martha E. Francis. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

[23] Stephanie S. Rude, Eva-Maria Gortner, and James W. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cogn. Emot. 18*, 8 (December 2004), 1121-1133. DOI: https://doi.org/10.1080/02699930441000030.

[24] Allison M. Tackman, David A. Sbarra, Angela L. Carey, M. Brent Donnellan, Andrea B. Horn, Nicholas S. Holtzman, To'Meisha S. Edwards, James W. Pennebaker, and Matthias R. Mehl. 2019. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *J. Pers. Soc. Psychol., 116*, 5 (May 2019): 817-834. DOI: 10.1037/pspp0000187

[25] Daniel J. Paulus, Salome Vanwoerden, Peter J. Norton, and Carla Sharp. 2016. From neuroticism to anxiety: Examining unique contributions of three transdiagnostic vulnerability factors. 2016. *Pers. Individ. Differ., 94*, (May 2016), 38-43. DOI: 10.1016/j.paid.2016.01.012

[26] Molly E. Ireland & Micah Iserman. 2018. Within and between-person differences in language used across anxiety support and neutral reddit communities. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, *NAACL HLT,* (June 2018), 182-193. https://www.aclweb.org/anthology/W18-06.pdf#page=194

[27] Aubrey J. Rodriguez, Shannon E. Holleran, and Matthias R. Mehl. 2010. Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *J. Pers.,* 78, 2, (April 2010), 575-598. DOI: 10.1111/j.1467-6494.2010.00627.x

[28] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. https://repositories.lib.utexas.edu/handle/2152/31333

[29] Jenna L. Baddeley, James W. Pennebaker, and Christopher G. Beevers. 2013. Everyday social behavior during a major depressive episode. *Soc. Psychol. Pers. Sci., 4*, (July 2013), 445-452. DOI: 10.1177/1948550612461654

[30] Taleen Nalabandian and Molly E. Ireland. 2019. Depressed Individuals Use Negative Self-Focused Language When Recalling Recent Interactions with Close Romantic Partners but Not Family or Friends. 2019. In *Proc. of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, *ACL '19*, (June 2019), 62-73. Retrieved from https://www.aclweb.org/anthology/W19-3008.pdf

[31] John Concato, Nirav Shah, and Ralph I. Horwitz. 2000. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N. Engl. J. Med.,* 342, 25 (June 2000): 1887-1892. DOI: 10.1056/NEJM200006223422507

[32] Ian Barnett and John Torous. 2019. Ethics, transparency, and public health at the intersection of innovation and Facebook's suicide prevention efforts. *Ann. Intern. Med.*, *170*, 8, (April 2019), 565-566. DOI: 10.7326/M19-0366

[33] Geoff Bathje and John Pryor. 2011. The relationships of public and self-stigma to seeking mental health services. *J. Ment. Health Couns.*, *33*, (April 2011), 161-176. DOI: 10.17744/mehc.33.2.g632039274160411

[34] David C. Mohr, Katie Shilton, and Matthew Hotopf. 2020. Digital phenotyping, behavioral sensing, or personal sensing: Names and transparency in the digital age. *npj Digit Med, 3*, (March 2020), 1-2. DOI: 10.1038/s41746-020-0251-5

[35] Ralf Schwarzer (Ed.) 2014. *Self-efficacy: Thought control of action.* Taylor & Francis.

[36] Nicola Groth, Nina Schnyder, Michael Kaess, Andjela Markovic, Liz Rietschel, Susann Moser, Chantal Michel, Frauke Schultze-Lutter, and Stefanie J. Schmidt. 2019. Coping as a mediator between locus of control, competence beliefs, and mental health: A systematic review and structural equation modelling meta-analysis. *Behav. Res. Ther., 121*, (October 2019), 1-16. DOI: 10.1016/j.brat.2019.103442

[37] Jennifer A. DiCorcia and E. D. Tronick. 2011. Quotidian resilience: Exploring mechanisms that drive resilience from a perspective of everyday stress and coping. *Neurosci. Biobehav. Rev., 35,* (June 2011), 1593-1602. DOI: 10.1016/j.neubiorev.2011.04.008

[38] Megan Prictor, Harriet J. A. Teare, and Jane Kaye. 2018. Equitable participation in biobanks: The risks and benefits of a "dynamic consent" approach. *Front Public Health,* 6, (September 2018), 103442. DOI: 10.3389/fpubh.2018.00253

[39] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. Bethesda, MD, USA. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html

[40] J. K. Rowling. 1997. Harry Potter and the philosopher's stone. Bloomsbury Pub.