FU-TU-DFKI@eRisk 2025: A Linguistically Informed but Overdiagnosing Approach to Early Depression Detection

Notebook for the eRisk Lab at CLEF 2025

Elif Kara[1,6,*,†], Rosa Esther Martín Peña[2,3,6,†] and Lisa Raithel[4,5,6,†]

[1]*Freie Universität Berlin, Department of Philosophy and Humanities, Institute for English Language and Literature, Habelschwerdter Allee 45, 14195 Berlin, Germany*

[2]*Gottfried Wilhelm Leibniz Universität Hannover, Centre for Ethics and Law in the Life Sciences, Otto-Brenner-Straße 1, 30159 Hannover, Germany*

[3]*CAIMed – Center for Artificial Intelligence in Medicine, Hannover Medical School (MHH), Carl-Neuberg-Straße 1, 30625 Hannover, Germany*

[4]*Technische Universität Berlin, Quality and Usability Lab, Marchstraße 23, 10587 Berlin, Germany*

[5]*BIFOLD – Berlin Institute for the Foundations of Learning and Data, Ernst-Reuter-Platz 7, 10587 Berlin, Germany*

[6]*Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), DFKI Labor Berlin, Salzufer 15/16, 10587 Berlin, Germany*

Abstract

This paper describes the participation of the FU-TU-DFKI team in the eRisk 2025 Task 2, *Contextualized Early Detection of Depression*. We propose a hybrid approach that combines transformer-based modelling with linguistic and meta feature analysis. While our model achieved high recall, it exhibited low precision, resulting in an overall F_1-score of 0.29 in the official evaluation. We interpret this cautious behaviour as a tendency toward overdiagnosis. Beyond the technical system, we investigated the linguistic characteristics of user messages via corpus-linguistic methods, including Collostructional Analysis – a method for identifying statistically significant associations between words and grammatical constructions. Additionally, we examine the ethical implications of automated depression detection, and highlight the reductionist interpretation of complex affective utterances in such systems. Our submission emphasizes the importance of interpretability and caution in high-stakes, health-related NLP tasks, particularly when system performance remains limited.

Keywords

mental health, depression detection, transformer models, collostructional analysis, corpus linguistics, ethical NLP

1. Introduction

Depressive disorder is a serious mental health concern, affecting around 280 million adults worldwide according to the World Health Organization (WHO).[1] The condition has an impact on all phases and aspects of life, such as relationships, school, or work, making it a major public health concern. Nevertheless, many cases remain undiagnosed, are self-diagnosed, or are diagnosed only after significant delays, often resulting in worse outcomes for those affected [1, 2]. Early detection of depressive symptoms can enable more timely support and intervention, potentially improving quality of life and reducing long-term suffering [3, 4]. However, structural and social barriers often prevent individuals from seeking help. Even in well-funded healthcare systems, access to therapy can be limited [5]. Furthermore, mental health stigma remains widespread, making open conversations about psychological distress difficult for many people [6, 7, 8].

As a consequence, many individuals turn to social media platforms such as Reddit to express their thoughts, connect with others facing similar struggles, or seek informal advice [9]. The anonymity offered by these platforms allows users to share personal experiences more openly than they might in

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

†These authors contributed equally.

elif.kara@fu-berlin.de (E. Kara); rosa-esther.martin-pena@cells.uni-hannover.de (R. E. M. Peña); raithel@tu-berlin.de (L. Raithel)

[1] https://www.who.int/news-room/fact-sheets/detail/depression

offline contexts [10]. This makes social media a valuable, albeit noisy, source of linguistic and emotional data. Since most interactions on these platforms are text-based, language plays a central role in the way emotions and psychological states are communicated. As more and more patients express themselves online, there is growing interest in using Natural Language Processing (NLP) techniques to detect patterns and markers associated with the condition in large-scale text data [11].

All of these challenges are at the core of this year's eRisk 2025 workshop [12, 13], specifically Task 2, *Contextualized Early Detection of Depression*. In this paper, we present our system for this task which includes the following contributions:

- two corpus-linguistic pilot studies, including collostructional methods, on lexical characteristics of the training data to identify patterns and markers associated with depressive language (Section 3);
- a hybrid pipeline that combines a transformer-based prediction model (MentalBERT), handcrafted linguistic features, and contextual meta-information (Section 4); and
- a brief reflection on the ethical implications of applying NLP to social media data for mental health research (Section 5).

2. Related Work

Over the past decade, NLP has emerged as a powerful tool for studying mental health through language. A growing body of work has focused on extracting linguistic signals related to psychological wellbeing [14, 15], stress [16], anxiety [17, 18], schizophrenia [19, 20] and depression [17, 21, 22]. Among these, depression detection remains one of the most extensively studied applications. Shared tasks such as CLPsych and eRisk have driven methodological advances by providing annotated datasets and realistic evaluation settings for early detection [23, 24]. Systems developed for these tasks have explored a wide range of techniques, including keyword-based lexica, topic modelling [25], psycholinguistic feature extraction [26], and (deep) machine learning approaches, such as XGBoost or CNNs [27, 28].

Depression has also received particular attention from a linguistic perspective. Studies investigating linguistic markers of psychological distress have consistently reported correlations between depression and specific features, such as the frequency of first-person singular pronouns (FPSPs) [29, 30, 31], negatively valenced words [32, 33, 34], absolutist language [35], and a preference for past-tense verbs [36]. Among these, FPSP frequency has emerged as a particularly robust marker of depression, as found by a meta-analysis [37], and frequently reported both across analytical approaches [38, 39, 31] and languages [39, 31, 40]. This observation aligns with psychological theories positing that depression is associated with maladaptive self-focused attention schemas [41]. These studies demonstrate that linguistically grounded features, when integrated into NLP models, offer a scalable and transparent means of extracting mental health signals from user-generated content. In addition, they make it possible to monitor language use longitudinally and to identify early indicators of depression – even in the absence of explicit self-disclosure. However, challenges remain, particularly in achieving high precision and interpretability, and in addressing the (ethical) complexities of real-world deployment.

3. Linguistic Analysis

Dataset Both pilot studies were conducted on the official eRisk 2025 training data, which combines data from previous eRisk challenges in 2017, 2018 and 2022. It consists of the full conversational history of individual Reddit users, divided into a target group (henceforth POS) comprising depressed users, and a control group (henceforth NEG) comprising non-depressed users. The users in POS were selected based on statements disclosing a depression diagnosis; all posts containing such statements were subsequently removed from the dataset. For further information, see Losada and Crestani [42].

Table 1 provides an overview of the basic statistics of POS and NEG. NEG is roughly ten times larger than POS. This class imbalance is appropriate for statistical linguistic analysis, as the larger control group allows for more stable frequency estimates, provided that the NEG data reflects diverse and representative language use. However, we acknowledge that Reddit does not reflect general population demographics or mental health prevalence, as discussed in Section 5.

Table 1
Descriptive Statistics of the Cohorts

	POS	NEG
Total Number of Users	312	2,795
Total Word-Form Tokens	5,572,340	53,724,447
Mean No. of Messages per User	40.4422	33.0902
Mean Sentence Length	3.2414	2.7809
MSTTR (Segment: 1000)	0.0145	0.0088

POS users tend to have a larger posting history than NEG users. Moreover, while sentences overall are rather short, POS sentences are both longer and more lexically diverse[2] than NEG.

3.1. Pilot Study 1: First-Person Singular Pronoun Use

This study is motivated by FPSPs use being postulated as a robust linguistic marker of depression. First, we explore the relative distribution of *I*, followed by verbal associations with *I* in the two cohorts.

Distribution of FPSPs The lemma form of *I* – comprising *I* and *me* but excluding *my* and other forms of self-reference – occurs with a relative frequency of 4.81% (N=267,868) in the POS group and 2.50% (N=1,344,950) in the NEG group, confirming findings from previous studies. Deviation of Proportions (DP) was applied as a measure of dispersion.[3] POS yields a DP of 0.48 and NEG 0.49, indicating uneven distribution in both groups, with a minor skew of POS towards greater evenness. Moreover, the FPSP *I* is absent from only one depressed user's (0.32%) posting history, compared to 216 users (7.73%) in the control group. This supports and strengthens the observation that FPSP usage in the depressed dataset is not only more frequent but also more evenly distributed, as reflected in a narrower range.

Verbal Associations with *I* in POS vs NEG To explore why FPSPs are more frequent in the depression data, we turn to a qualitative investigation of how users with and without depression predicate states and actions about themselves. We operationalize this as an analysis of verb associations with the FPSP. To do so, we extract all instances of the construction [*I* + VERB] – plus optional slots for one adverb and up to two auxiliary verbs – from both datasets. The lists of verb lemmas are submitted to two subtypes of CA, implemented via the `collostructions` R package [44].

Collostructional Analysis CA, developed by Stefanowitsch and Gries [45] (see also [46, 47]) is a quantitative approach that offers insights into co-occurrence phenomena at the form-function interface.[4] It has been applied extensively to uncover systematic patterns in how lexical items associate with grammatical constructions across different languages and registers.[5] Thus, this method offers insights into both structural properties of language and the cognitive mechanisms underlying its use, supporting research on mental health and language, as demonstrated in a recent study [50].

We apply Distinctive Collexeme Analysis (DCA) to measure the association of verbs with [*I* + VERB] in POS, and compare them against the association of verbs with the same construction in NEG. This allows us to determine verbal associations characteristic of each cohort, emphasizing their differences.

[2] We used Mean Segmental Type-Token Ratio (MSTTR) to measure lexical diversity as it is insensitive to varying text lengths.
[3] DP compares the expected distribution of a linguistic unit across corpus segments to the observed distribution, with 0 indicating perfectly even and 1 maximally uneven distribution [43].
[4] CA is grounded in the theoretical framework of Construction Grammar, which assumes that linguistic units on all levels (words, morphemes, phrases, and sentences) are form-meaning pairings in the Saussurean sense, called *constructions* [48, 49].
[5] E.g., the English ditransitive construction [VERB + NP + NP] strongly attracts verbs of transfer (e.g., *give*, *send*, *offer*) while the caused-motion construction [VERB + NP + PP/AdvP] attracts verbs of placement (e.g., *put*, *place*, *throw*) [45].

Results Table 3, in Appendix A, displays the verbs most strongly, positively associated with each dataset (all p<.0001). A glance at the highest-ranking verbs reveals that POS strongly attracts verbs encoding negative sentiment (e.g., *struggle*, *suffer*, *cry*, *lose*), mental health-related verbs (e.g., *diagnose*, *prescribe*, *hospitalize*, *misdiagnose*), as well as emotion verbs (e.g., *feel*). In contrast, NEG strongly attracts verbs with more neutral sentiment (e.g., *modify*, *see*, *think*, *mean*, *watch*), indicating a more casual, conversational tone. A post-hoc sentiment analysis of the [I + VERB] constructions via NLTK's VADER [51], using the three polarity labels *positive*, *negative*, and *neutral*, confirms that negative sentiment is more common among the highest-ranking verb associations in the depressed cohort (18% in POS vs 8% in NEG). Additionally, we cross-examined the results from the DCA with a Simple Collexeme Analysis (SCA) that compares verb associations with [I + VERB] against their overall corpus frequency.[6] We did this for each cohort in order to identify intra-cohort associations. Remarkably, nearly all of the top 100 collexemes are shared between the cohorts, and reflect neutral sentiment (see Table 6). Thus, although an independent analysis shows both groups are associated with neutral language, a contrastive approach reveals clear differences in affective characterization.

3.2. Pilot Study 2: Important Concepts

This study is motivated by two goals: first, to complement NLP methods for concept detection, such as topic modelling and TF-IDF, and second, to build on the preliminary finding from Pilot Study 1 that mental health-related verbs are characteristically predicated about the self in the depressed cohort. For this, we extract all word-form tokens from each cohort and compare the resulting word lists against a 440-million-word subset of the Corpus of Contemporary American English (COCA) [52].[7] In addition to this corpus-level analysis, we conducted KAs for each year in which users posted messages (2009 to 2021), in order to track lexical trends over time. While this does not inform early risk detection, it serves as a form of cross-validation to identify which concepts recur across years in the POS and NEG datasets.

Keyword Analysis Keyword Analysis (KA) is a corpus-linguistic approach to identifying tokens that are *key* in a given corpus. Like CA, KA is a transparent statistical method aimed at detecting words associated with a target corpus; in the case of KA, relative to a larger reference corpus [53]. When applied to highly specialized corpora – for example, comprising depression data – KA can reveal deviations from the lexical norms and conceptual patterns considered typical for the broader speech community represented by the control corpus.

Results Table 6, in Appendix C, displays the top keywords, indicating strong thematic differences between the datasets: POS shows an overrepresentation of first-person pronouns, along with affective (*feel*, *love*), mental health-related (*depression*) and clinical (*meds*) vocabulary. In contrast, NEG shows an overrepresentation of words related to financial and transactional discourse (*binance*, *wallet*, *account*, *cryptocurrency*). As expected, COCA keywords reflect more formal and informational language, comprising function words (*the*, *of*, *in*), including third-person pronouns (*his*, *he*), and proper nouns tied to political discourse (*president*, *national*, *united*). The association with function words reflects general patterns of natural language use, while the other salient categories likely stem from COCA's composition which is biased toward formal, written genres. All of these patterns are also reflected in the annual KAs: the most prominent keywords per cohort recur consistently across all 13 years. In addition, some overlap of genre-specific expressions[8] was observed in both POS and NEG, demonstrating how lexical choices are influenced by multiple factors, which studies on mental health and language must account for.

[6] The research designs of both a DCA and an SCA are schematically illustrated in Table 5, in Appendix B.

[7] The COCA covers eight genres (spoken, fiction, academic, newspapers, etc.) published between 1990 and 2019, and is widely considered to be a balanced corpus of present-day American English.

[8] Including second-person pronouns (*you*), platform- and medium-specific expressions (*reddit*, *lol*, *haha*, *fuck*), as well as conversational markers (*thanks*).

4. Pipeline

Our pipeline is based on the predictions of a transformer-based [54] encoder-only model [55], the linguistic analyses, and metadata, either extracted directly from the incoming messages or inferred from the broader conversation context. We opted for an encoder-only architecture over a Large Language Model (LLM) due to its efficiency, interpretability, and compatibility with additional linguistic and meta features. The pipeline is illustrated in Figure 1.

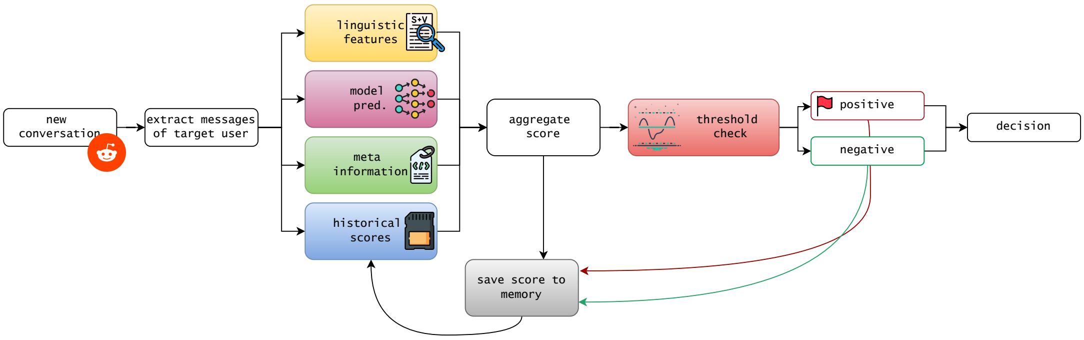

Figure 1: The final pipeline. User messages are processed using a transformer-based model fine-tuned for binary classification (user is *likely* vs *not likely* to be diagnosed with depression). We extract the probability assigned to the *likely* class, and combine it with linguistic and meta features to make the final decision for each message.

Model For model predictions, we use MentalBERT [56],[9] a BERT-based model [55] which was continuously pretrained on English Reddit posts related to mental health.[10] This domain-specific pretraining allows the model to better handle informal language and topic-specific expressions. We fine-tuned MentalBERT on a balanced subset of our training data for a binary classification task: given a message, the model predicts whether the user is likely (*positive*) or not likely (*negative*) to exhibit signs of depression. Following hyperparameter tuning, the best configuration achieved an F_1-score of 0.63 on the positive class on a held-out validation set, reflecting the difficulty of the task.

Features For each new conversation, we retrieve the target user's messages in chronological order and extract linguistically motivated features based on the analyses. Specifically, we scanned each message for (a) instances where *I* was followed by a verb[11] associated with the POS group within a five-token window (see Pilot Study 1), and (b) for keywords associated with the POS group (see Pilot Study 2). Additionally, we incorporated a small set of meta-level behavioural indicators: (a) the *night writer* feature, which captures how frequently a user posted messages between 11:00 pm and 06:00 am, motivated by the established link between sleep disturbances and depressive symptoms [57, 58]; and (b) a sentiment classification pipeline based on a pretrained model from the Huggingface Transformers library,[13] given prior findings of increased negative sentiment in depression [59, 60] and in our own linguistic analyses. Both the linguistic and meta features served to bias the final decision of the system in case the prediction model was not confident.

Decision Logic To ensure stability and avoid unreliable decisions based on limited data, the system waits until a user has posted a sufficient number of messages before making a prediction.[14] The final decision integrates weighted model probabilities, linguistic features, and metadata through a threshold-based logic, as outlined in Appendix D. Upon receiving a new batch of messages from a user, we

[9] https://huggingface.co/mental/mental-bert-base-uncased; MentalBERT was the best model in preliminary experiments.
[10] A subset of the eRisk18 T1 dataset [42] was included in the model's training set.
[11] We applied lemmatization using the spaCy library[12] to increase robustness by including inflected forms.
[13] Using the default model https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english with three labels, positive, negative, and neutral.
[14] We set this threshold to 5 messages and require that at least 2 rounds have already been processed.

use top-k (with $k = 3$) averaging to capture peak signals across their history while reducing noise, and combining these with historical scores to update the assessment, which is then saved to inform upcoming predictions.

4.1. Results and Preliminary Analysis of Error Sources

Table 2 presents the official evaluation results of our system, as provided by the task organizers. Due to hardware issues, our submission processed only 449 user threads out of the intended 1,280. The resulting

Table 2
The results as provided by the task organizers, including precision (P), recall (R), F_1-score (F_1), ERDE (5 or 50), latency of responses (latency$_{TP}$), speed, and latency-aware F_1-score ($F_{latency}$).

P	R	F_1	ERDE$_5$	ERDE$_{50}$	latency$_{TP}$	speed	F$_{latency}$
0.17	0.97	0.29	0.16	0.07	11.00	0.96	0.28

F_1 score is 0.29, which is, unfortunately, to be found within the lower end of the participating teams' scores. However, the model demonstrated a high recall of 0.97, indicating that it successfully identified most users with depression. This came at the cost of very low precision (0.17), meaning the system overdiagnosed users and produced a high number of false positives. The early risk detection error (ERDE) scores [42] further reflect the system's cautious behaviour: with a latency$_{TP}$ of 11 messages, the model generally waited for a substantial amount of user data before making a positive prediction. While this helped avoid premature decisions, it limited the model's ability to detect depression early. This is also reflected in the low latency-aware F_1-score (0.28), despite an overall speed score of 0.96.

5. Discussion and Conclusion

Our system for eRisk 2025 Task 2 sought to balance predictive performance with interpretability by combining a transformer-based model with linguistically motivated features. While the system achieved relatively high recall, low precision resulted in an overall F_1-score of 0.29. This outcome reflects our emphasis on avoiding false negatives in a high-risk domain, but also highlights the trade-off between sensitivity and specificity in early depression detection. Several design choices may have contributed to these outcomes: we relied on a single model with heuristically selected parameters and no uncertainty calibration. In future work, we aim to better integrate linguistic and contextual features to improve transparency and accuracy, as well as to explore ensemble approaches, LLMs, and more adaptive decision logic.

Independent of model performance, data limitations pose broader concerns in this domain: social media data lacks clinical validation, and demographic biases, such as the underrepresentation of older adults or individuals with limited digital access, reduce generalizability.

From an ethical perspective, we acknowledge that language-based models in mental health contexts are not neutral. For example, they encode assumptions about the relevance of emotional expressions, introducing an algorithmic biases whereby complex expressions of distress (e.g., *I cried*) may be pathologized or misinterpreted. While we incorporated such tools into our pipeline, we recognize their potential as well as their limited generalizability: apparent predictive accuracy may stem from superficial lexical cues, ultimately compromising model robustness. In our case, the sentiment classification pipeline, although treated with caution, proved particularly unreliable and may have introduced noise. Moreover, such biases raise ethical concerns about whose emotional registers are recognized and whose are overlooked [61, 62].

Building on these concerns, we propose a broader reframing of emotional language in mental health research: one that recognizes that linguistic expression is shaped by multiple contextual factors beyond mental health status, such as genre, interactional context, and communicative intent. Future research should incorporate these dimensions in both model design and evaluation.

Acknowledgments

We would like to thank the organizers of CLEF eRisk 2025 for this interesting task. We found it to be both technically and conceptually rewarding as it challenged us to think beyond raw performance.

This work was supported by the German Federal Ministry of Education and Research (BIFOLD25B), and conducted within the framework of the CAIMed – Center for Artificial Intelligence in Medicine, which supports interdisciplinary research at the intersection of AI, ethics, and medicine.

Declaration on Generative AI

The authors used Claude 3.7 Sonnet as a programming aid for the sentiment classification in the corpus-linguistic component. The analysis and interpretation of the results were carried out independently by the authors.

CrediT Authorship Contribution Statement

Elif Kara: Linguistic Analysis (Conceptualization, Methodology, Formal Analysis, Investigation, and Data Curation); Writing – Original Draft (Related Work, Linguistic Analysis); Writing – Review & Editing; Supervision **Rosa Esther Martín Peña**: Ethical Analysis (Conceptualization, Methodology, Formal Analysis and Investigation); Writing – Original Draft (Discussion and Conclusion) **Lisa Raithel**: Machine Learning Pipeline (Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, and Visualization); Writing – Original Draft (Introduction, Related Work, Pipeline, Discussion and Conclusion)

References

[1] A. Handy, R. Mangal, T. S. Stead, R. L. Coffee, L. Ganti, Prevalence and impact of diagnosed and undiagnosed depression in the United States, Cureus (2022). doi:10.7759/cureus.28011.

[2] R. M. Epstein, P. R. Duberstein, M. D. Feldman, A. B. Rochlen, R. A. Bell, R. L. Kravitz, C. Cipri, J. D. Becker, P. M. Bamonti, D. A. Paterniti, "I didn't know what was wrong:" How people with undiagnosed depression recognize, name and explain their distress, Journal of General Internal Medicine 25 (2010) 954–961. doi:10.1007/s11606-010-1367-0.

[3] C. Kraus, B. Kadriu, R. Lanzenberger, C. A. Zarate, S. Kasper, Prognosis and improved outcomes in Major Depression: A review, Translational Psychiatry 9 (2019). doi:10.1038/s41398-019-0460-3.

[4] C. Buntrock, M. Harrer, A. A. Sprenger, S. Illing, M. Sakata, T. A. Furukawa, D. D. Ebert, P. Cuijpers, Psychological interventions to prevent the onset of Major Depression in adults: A systematic review and individual participant data meta-analysis, The Lancet Psychiatry 11 (2024) 990–1001. doi:10.1016/s2215-0366(24)00316-x.

[5] A. M. Boerema, M. Ten Have, A. Kleiboer, R. de Graaf, J. Nuyen, P. Cuijpers, A. T. F. Beekman, Demographic and need factors of early, delayed and no mental health care use in Major Depression: A prospective study, Bmc Psychiatry 17 (2017) 367. doi:10.1186/s12888-017-1531-8.

[6] S. Tomczyk, S. Schmidt, H. Muehlan, S. Stolzenburg, G. Schomerus, A prospective study on structural and attitudinal barriers to professional help-seeking for currently untreated mental health problems in the community, The Journal of Behavioral Health Services I& Research 47 (2019) 54–69. doi:10.1007/s11414-019-09662-8.

[7] S. Clement, O. Schauman, T. Graham, F. Maggioni, S. Evans-Lacko, N. Bezborodovs, C. Morgan, N. Rüsch, J. S. L. Brown, G. Thornicroft, What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies, Psychological Medicine 45 (2014) 11–27. doi:10.1017/s0033291714000129.

[8] Z. Xu, F. Huang, M. Kösters, T. Staiger, T. Becker, G. Thornicroft, N. Rüsch, Effectiveness of interventions to promote help-seeking for mental health problems: Systematic review and meta-analysis, Psychological Medicine 48 (2018) 2658–2667. doi:10.1017/s0033291718001265.
[9] M. D. Choudhury, E. Kıcıman, Characterizing and predicting mental health disclosures on social media, in: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (icwsm), AAAI Press, 2014, pp. 71–80. doi:10.1609/icwsm.v8i1.14526.
[10] T. D. Afifi, E. D. Basinger, J. A. Kam, The extended theoretical model of communal coping: Understanding the properties and functionality of communal coping., Journal of Communication 70 (2020) 424–446. doi:10.1093/joc/jqaa006.
[11] T. Zhang, A. M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental illness detection: A narrative review, npj Digital Medicine 5 (2022). doi:10.1038/s41746-022-00589-7.
[12] J. Parapar, A. Perez, X. Wang, F. Crestani, Overview of eRisk 2025: Early Risk Prediction on the Internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 16th International Conference of the CLEF Association, CLEF 2025, Madrid, Spain, September 9-12, 2025, Proceedings, Part II, volume To be published of *Lecture Notes in Computer Science*, Springer, 2025.
[13] J. Parapar, A. Perez, X. Wang, F. Crestani, Overview of eRisk 2025: Early Risk Prediction on the Internet (Extended Overview), in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025), Madrid, Spain, 9-12 September, 2025, volume To be published of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025.
[14] N. Fujikawa, Q. T. Nguyen, K. Ito, S. Wakamiya, E. Aramaki, Loneliness episodes: A Japanese dataset for loneliness detection and analysis, in: O. De Clercq, V. Barriere, J. Barnes, R. Klinger, J. Sedoc, S. Tafreshi (Eds.), Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 280–293. doi:10.18653/v1/2024.wassa-1.23.
[15] T. Tseriotou, J. Chim, A. Klein, A. Shamir, G. Dvir, I. Ali, C. Kennedy, G. Singh Kohli, A. Hills, A. Zirikly, D. Atzil-Slonim, M. Liakata, Overview of the CLPsych 2025 shared task: Capturing mental health dynamics from social media timelines, in: A. Zirikly, A. Yates, B. Desmet, M. Ireland, S. Bedrick, S. MacAvaney, K. Bar, Y. Ophir (Eds.), Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2025), Association for Computational Linguistics, Albuquerque, New Mexico, 2025, pp. 193–217. URL: https://aclanthology.org/2025.clpsych-1.16/.
[16] M. Mendula, S. Gabrielli, F. Finazzi, C. Dompe, M. Delucis, Unveiling mental health insights: A novel NLP tool for stress detection through writing and speaking analysis to prevent burnout, AHFE International 122 (2024) 164–174.
[17] A. Fine, P. Crutchley, J. Blase, J. Carroll, G. Coppersmith, Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data, in: D. Bamman, D. Hovy, D. Jurgens, B. O'Connor, S. Volkova (Eds.), Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, Association for Computational Linguistics, 2020, pp. 50–54. doi:10.18653/v1/2020.nlpcss-1.6.
[18] D. Zarate, M. Ball, M. Prokofieva, V. Kostakos, V. Stavropoulos, Identifying self-disclosed anxiety on Twitter: A Natural Language Processing approach, Psychiatry Research 330 (2023) 115579. doi:10.1016/j.psychres.2023.115579.
[19] S. Just, E. Haegert, N. Kořánová, A.-L. Bröcker, I. Nenchev, J. Funcke, C. Montag, M. Stede, Coherence models in schizophrenia, in: K. Niederhoffer, K. Hollingshead, P. Resnik, R. Resnik, K. Loveys (Eds.), Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 126–136. doi:10.18653/v1/W19-3015.
[20] I. Nenchev, T. Scheffler, M. de la Fuente, H. Stuke, B. Wilck, S. A. Just, C. Montag, Linguistic markers of schizophrenia: a case study of Robert Walser, in: A. Yates, B. Desmet, E. Prud'hommeaux, A. Zirikly, S. Bedrick, S. MacAvaney, K. Bar, M. Ireland, Y. Ophir (Eds.), Proceedings of the 9th

Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024), Association for Computational Linguistics, St. Julians, Malta, 2024, pp. 41–60. URL: https://aclanthology.org/2024.clpsych-1.4.

[21] N. A. Abdelkadir, C. Zhang, N. Mayo, S. Chancellor, Diverse perspectives, divergent models: Cross-cultural evaluation of depression detection on Twitter, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (volume 2: Short Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 672–680. doi:10.18653/v1/2024.naacl-short.58.

[22] A.-M. Bucur, A. Moldovan, K. Parvatikar, M. Zampieri, A. Khudabukhsh, L. Dinu, Datasets for depression modeling in social media: An overview, in: A. Zirikly, A. Yates, B. Desmet, M. Ireland, S. Bedrick, S. MacAvaney, K. Bar, Y. Ophir (Eds.), Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2025), Association for Computational Linguistics, Albuquerque, New Mexico, 2025, pp. 116–126. URL: https://aclanthology.org/2025.clpsych-1.10/.

[23] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, M. Mitchell, CLPsych 2015 shared task: Depression and PTSD on Twitter, in: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 31–39. doi:10.3115/v1/W15-1204.

[24] D. E. Losada, F. Crestani, J. Parapar, Overview of eRisk: Early risk prediction on the internet, in: P. Bellot, C. Trabelsi, J. Mothe, F. Murtagh, J. Y. Nie, L. Soulier, E. SanJuan, L. Cappellato, N. Ferro (Eds.), Experimental Ir Meets Multilinguality, Multimodality, and Interaction, Springer International Publishing, 2018, pp. 343–361. doi:10.1007/978-3-319-98932-7_30.

[25] D. Maupomé, M. D. Armstrong, F. Rancourt, T. Soulas, M.-J. Meurs, Early detection of signs of pathological gambling, self-harm and depression through topic extraction and neural networks., in: CLEF (Working Notes), 2021, pp. 1031–1045.

[26] S. Zanwar, D. Wiechmann, Y. Qiao, E. Kerz, SMHD-GER: A large-scale benchmark dataset for automatic mental health detection from social media in German, in: A. Vlachos, I. Augenstein (Eds.), Findings of the Association for Computational Linguistics: EACL 2023, Association for Computational Linguistics, Dubrovnik, Croatia, 2023, pp. 1526–1541. doi:10.18653/v1/2023.findings-eacl.113.

[27] E. Campillo-Ageitos, J. Martinez-Romo, L. Araujo, UNED-MED at eRisk 2022: depression detection with TF-IDF, linguistic features and embeddings, in: Proceedings of the Working Notes of CLEF 2022, 2022, pp. 864–874.

[28] A. Husseini Orabi, P. Buddhitha, M. Husseini Orabi, D. Inkpen, Deep learning for depression detection of Twitter users, in: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, Association for Computational Linguistics, 2018, pp. 88–97. doi:10.18653/v1/w18-0609.

[29] M. de Choudhury, S. Counts, E. Horvitz, Social media as a measurement tool of depression in populations, in: Proceedings of the 5th Annual Acm Web Science Conference, WebSci '13, ACM, New York, NY, 2013, pp. 47–56. doi:10.1145/2464464.2464480.

[30] J. W. Pennebaker, The secret life of pronouns, New Scientist 211 (2011) 42–45. doi:10.1016/s0262-4079(11)62167-2.

[31] D. Smirnova, P. Cumming, E. Sloeva, N. Kuvshinova, D. Romanov, G. Nosachev, Language patterns discriminate mild depression from normal sadness and euthymic state, Frontiers in Psychiatry 9 (2018) 1–11. doi:10.3389/fpsyt.2018.00105.

[32] J. L. Baddeley, J. W. Pennebaker, C. G. Beevers, Everyday social behavior during a Major Depressive Episode, Social Psychological and Personality Science 4 (2013) 445–452. doi:10.1177/1948550612461654.

[33] G. Gkotsis, A. Oellrich, T. Hubbard, R. Dobson, M. Liakata, S. Velupillai, R. Dutta, The language of mental health problems in social media, in: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, 2016, pp. 63–73. doi:10.18653/v1/W16-0307.

[34] N. Ramirez-Esparza, C. K. Chung, E. Kacewicz, J. W. Pennebaker, The psychology of word use in depression forums in English and in Spanish: Texting two text analytic approaches, in: ICWSM 2008 - Proceedings of the 2nd International Conference on Weblogs and Social Media, 2008, pp. 102–108. doi:10.1609/icwsm.v2i1.18623.

[35] M. Al-Mosaiwi, T. Johnstone, In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation, Clinical Psychological Science 6 (2018) 529–542. doi:10.1177/2167702617747074, pMID: 30886766.

[36] A. Arntz, L. D. Hawke, L. Bamelis, P. Spinhoven, M. L. Molendijk, Changes in natural language use as an indicator of psychotherapeutic change in personality disorders, Behaviour Research and Therapy 50 (2012) 191–202. doi:10.1016/j.brat.2011.12.007.

[37] T. Edwards, N. S. Holtzman, A meta-analysis of correlations between depression and first person singular pronoun use, Journal of Research in Personality 68 (2017) 63–68. doi:10.1016/j.jrp.2017.02.005.

[38] D. Davis, T. C. Brock, Use of first person pronouns as a function of increased objective self-awareness and performance feedback, Journal of Experimental Social Psychology 11 (1975) 381–388.

[39] J. Zimmermann, T. Brockmeyer, M. Hunn, H. Schauenburg, M. Wolf, First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients, Clinical Psychology & Psychotherapy 24 (2016) 384–391. doi:10.1002/cpp.2006.

[40] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, F. Sanz, Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis, Journal of Medical Internet Research 21 (2019) e14199. doi:10.2196/14199.

[41] A. T. Beck, Depression: Clinical, Experimental, and Theoretical Aspects, Harper & Row, New York, NY; Evanston, IL; London, UK, 1967.

[42] D. E. Losada, F. Crestani, A test collection for research on depression and language use, in: N. Fuhr, P. Quaresma, T. Gonçalves, B. Larsen, K. Balog, C. Macdonald, L. Cappellato, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, volume 9822, Springer International Publishing, Cham, 2016, pp. 28–39. doi:10.1007/978-3-319-44564-9_3.

[43] S. T. Gries, Dispersions and adjusted frequencies in corpora, International Journal of Corpus Linguistics 13 (2008) 403–437. doi:10.1075/ijcl.13.4.02gri.

[44] S. Flach, Collostructions: An R implementation for the family of collostructional methods. R package version 0.2.0, 2021. URL: https://sfla.ch/collostructions/.

[45] A. Stefanowitsch, S. T. Gries, Collostructions: Investigating the interaction of words and constructions, International Journal of Corpus Linguistics 8 (2003) 209–243. doi:10.1075/ijcl.8.2.03ste.

[46] S. T. Gries, A. Stefanowitsch, Extending collostructional analysis: A corpus-based perspective on alternations, International Journal of Corpus Linguistics 9 (2004) 97–129. doi:10.1075/ijcl.9.1.06gri.

[47] A. Stefanowitsch, Collostructional analysis, in: T. Hoffmann, G. Trousdale (Eds.), The Oxford Handbook of Construction Grammar, Oxford University Press, Oxford, UK, 2013, pp. 290–306. doi:10.1093/oxfordhb/9780195396683.013.0016.

[48] C. J. Fillmore, Syntactic intrusions and the notion of grammatical construction, in: Annual Meeting of the Berkeley Linguistics Society, volume 11, Linguistic Society of America, Berkeley, 1985, pp. 73–86. doi:10.3765/bls.v11i0.1913.

[49] A. E. Goldberg, Constructions: A Construction Grammar Approach to Argument Structure, University of Chicago Press, Chicago, 1995.

[50] E. Kara, What depression feels like: A collostructional analysis of patient and caregiver perspectives, Zeitschrift für Anglistik und Amerikanistik 72 (2024) 249–282. doi:10.1515/zaa-2024-2027.

[51] C. J. Hutto, E. Gilbert, VADER: A parsimonious rule-based model for sentiment analysis of social media text, Proceedings of the International AAAI Conference on Web and Social Media (2014). doi:10.1609/icwsm.v8i1.14550.

[52] M. Davies, The Corpus of Contemporary American English: 450 million words, 1990–2012, 2008. URL: http://corpus.byu.edu/coca.

[53] M. Scott, PC analysis of key words — and key key words, System 25 (1997) 233–245. doi:10.1016/s0346-251x(97)00011-0.

[54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 31st Conference on Neural Information Processing Systems (NIPS 2017 (2017) 11.

[55] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2018. doi:10.48550/ARXIV.1810.04805.

[56] S. Ji, T. Zhang, L. Ansari, J. Fu, P. Tiwari, E. Cambria, MentalBERT: Publicly available pretrained language models for mental healthcare, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7184–7190. URL: https://aclanthology.org/2022.lrec-1.778/.

[57] D. Nutt, S. Wilson, L. Paterson, Sleep disorders as core symptoms of depression, Dialogues in Clinical Neuroscience 10 (2008) 329–336. doi:10.31887/DCNS.2008.10.3/dnutt.

[58] S. Yasugaki, H. Okamura, A. Kaneko, Y. Hayashi, Bidirectional relationship between sleep and depression, Neuroscience Research 211 (2025) 57–64. doi:10.1016/j.neures.2023.04.006.

[59] T. Zhang, K. Yang, S. Ji, S. Ananiadou, Emotion fusion for mental illness detection from social media: A survey, Information Fusion 92 (2023) 231–246. doi:10.1016/j.inffus.2022.11.031.

[60] N. V. Babu, E. G. M. Kanaga, Sentiment analysis in social media data for depression detection using artificial intelligence: A review, Sn Computer Science 3 (2022) 74. doi:10.1007/s42979-021-00958-1.

[61] A. Benton, M. Mitchell, D. Hovy, Multitask learning for mental health using social media text, in: Proceedings of EACL 2017, 2017, pp. 152–162.

[62] C. Burr, Digital psychiatry: Risks, opportunities and recommendations, Frontiers in Digital Health 4 (2022). doi:10.1109/TTS.2020.2977059.

[63] T. R. C. Read, N. A. C. Cressie, Goodness-Of-Fit Statistics for Discrete Multivariate Data, Springer, New York, 1988. doi:10.1007/978-1-4612-4578-0.

[64] S. Evert, B. Krenn, Methods for the qualitative evaluation of lexical association measures, in: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Toulouse, France, 2001, pp. 188–195. doi:10.3115/1073012.1073037.

[65] A. Stefanowitsch, S. Flach, *Too big to fail* but *big enough to pay for their mistakes*: A Collostructional Analysis of the Patterns [*too* Adj *to* V] and [Adj *enough to* V], John Benjamins Publishing Company, Amsterdam, 2020, pp. 247–272. doi:10.1075/ivitra.24.13ste.

A. Pilot Study 1: DCA Results

This appendix (Table 3) displays the distinctive verb associations with the FPSP in POS vs NEG, the observed and expected frequencies, as well as the strength of association. The association measure is the G statistic from the log-likelihood ratio test, which compares observed co-occurrence frequencies against expected frequencies under the assumption of independence. This test is well suited for analysing linguistic data with uneven frequency distributions, which applies to the present datasets [63, 64].

Table 3
Top 40 verbs associated with the linguistic construction [*I* VERB] in a DCA of POS and NEG, displaying observed (Obs) and expected (Exp) frequencies, as well as the G value of the log-likelihood ratio test.

POS				NEG			
Verb	Obs	Exp	G	Verb	Obs	Exp	G
feel	17,651	20,245	1,930.21	*modify*	3	464	972.18
diagnose	103	390	889.01	*see*	3,897	5,301	493.96
ovulate	6	90	332.26	*think*	9,897	11,678	366.05
relate	307	473	278.39	*mean*	1,213	1,875	312.68
start	7,450	8,058	275.87	*watch*	544	855	150.90
struggle	699	888	209.53	*hear*	1,324	1,760	139.07
suffer	218	338	203.65	*agree*	948	1,294	119.41
try	14,592	15,279	194.84	*read*	1,028	1,370	109.60
cry	554	715	187.03	*run*	319	517	101.79
take	7,565	8,055	182.03	*accept*	113	244	100.68
wish	5,281	5,657	151.58	*believe*	979	1,295	98.86
lose	2,186	2,437	149.65	*post*	401	594	82.36
date	156	238	135.92	*pull*	81	175	71.76
want	21,003	21,669	132.60	*expect*	270	417	68.56
experience	636	767	121.31	*wonder*	901	1,149	68.04
work	4,539	4,846	118.14	*like*	3,037	3,461	65.12
know	28,005	28,719	117.09	*bet*	213	342	64.85
stop	1,730	1,921	110.03	*check*	300	448	64.46
end	1,517	1,691	103.61	*link*	31	89	57.13
prescribe	36	77	96.74	*add*	243	367	55.37
need	10,291	10,684	91.46	*suspect*	92	168	46.99
eat	1,579	1,735	82.39	*remember*	985	1,195	46.22
tell	6,341	6,622	74.44	*make*	1,503	1,748	42.76
smoke	316	389	73.19	*sell*	77	143	42.40
sleep	586	676	67.67	*guess*	1,876	2,143	41.35
become	886	993	65.82	*ride*	19	59	41.12
regress	3	21	65.44	*step*	17	55	40.85
develop	178	232	65.22	*create*	65	125	40.30
break	594	682	64.31	*doubt*	273	382	40.02
gain	230	289	63.47	*buy*	692	856	39.60
live	4,536	4,755	62.29	*disagree*	114	188	39.37
drink	667	757	60.73	*stand*	113	187	38.87
spend	2,056	2,202	58.50	*nod*	11	43	38.55
hate	5,689	5,920	56.52	*grab*	47	98	38.06
deal	414	484	56.09	*remove*	46	95	35.60
handle	191	240	52.43	*own*	97	162	34.92
hospitalize	6	22	49.16	*vote*	49	98	33.97
misdiagnose	3	17	48.06	*assume*	483	611	33.69
abuse	20	41	46.78	*write*	355	464	32.63
lack	100	135	45.37	*point*	54	103	32.34
fail	407	468	44.51	*search*	58	105	28.66

Table 4 displays the collexemes overlapping in SCAs of the two cohorts, as a baseline to the contrastive

DCA.

Table 4
99 of the top 100 verb associations with [*I* VERB] overlap in SCAs of POS and NEG (all p<.0001)

Verb lemmas associated with both cohorts
think, *feel*, *know*, *love*, *want*, *like*, *hope*, *see*, *try*, *guess*, *find*, *use*, *wish*, *need*, *start*, *say*, *hate*, *hear*, *wonder*, *agree*, *take*, *remember*, *look*, *believe*, *mean*, *tell*, *understand*, *read*, *live*, *miss*, *appreciate*, *ask*, *work*, *realize*, *notice*, *play*, *recommend*, *lose*, *buy*, *assume*, *enjoy*, *suppose*, *decide*, *spend*, *learn*, *make*, *doubt*, *watch*, *prefer*, *struggle*, *meet*, *put*, *plan*, *figure*, *give*, *imagine*, *keep*, *diagnose*, *end*, *wake*, *stop*, *bet*, *suggest*, *eat*, *consider*, *talk*, *post*, *leave*, *come*, *forget*, *move*, *dunno*, *cry*, *write*, *manage*, *experience*, *tend*, *relate*, *swear*, *expect*, *check*, *call*, *forgot*, *grow*, *sit*, *run*, *finish*, *pick*, *wear*, *drink*, *promise*, *fall*, *order*, *walk*, *pay*, *sleep*, *choose*, *suffer*, *turn*

B. Pilot Study 1: DCA and SCA Research Designs

This appendix (Table 5) illustrates the research designs of DCA and SCA. DCA assesses the association strength between a lexical item l and one construction c_1 over another related construction c_2. In contrast, SCA assesses the association strength of a lexical item l with a construction c, relative to all other lexical items occurring in c and outside of it (*!c*).

Table 5
Schematic contingency table for DCA (left) and SCA (right) [65]

		L		
		l	$!l$	Total
C	c_1	O_{11}	O_{12}	R_1
	c_2	O_{21}	O_{22}	R_2
	Total	C_1	C_2	N

		L		
		l	$!l$	Total
C	c	O_{11}	O_{12}	R_1
	$!c$	O_{21}	O_{22}	R_2
	Total	C_1	C_2	N

C. Pilot Study 2: KA Results

This appendix (Table 6) provides the highest-ranking keywords associated with the datasets. As before, the association metric is the G value of the log-likelihood ratio test.

Table 6
Top keywords associated with the datasets (all p<.0001)

POS	*i*, *my*, *nt*, *me*, *you*, *it*, *m*, *just*, *do*, *like*, *lol*, *so*, *feel*, *ve*, *really*, *if*, *am*, *depression*, *get*, *shit*, *but*, *your*, *haha*, *re*, *have*, *etc*, *reddit*, *edit*, *myself*, *s*, *someone*, *thanks*, *fuck*, *pretty*, *meds*, *fucking*, *love*, *definitely*, *ca*, *op*, *honestly*, *because*, *awesome*, *anxiety*, *try*, *people*, *idk*, *self*, *can*, *want*
NEG	*reddit*, *binance*, *account*, *i*, *your*, *you*, *nt*, *cryptocurrency*, *exchanging*, *email*, *wallet*, *crypto*, *fuck*, *gt*, *my*, *discount*, *m*, *implies*, *it*, *referral*, *lol*, *register*, *amp*, *just*, *r*, *awesome*, *check*, *will*, *select*, *trade*, *if*, *approached*, *send*, *code*, *is*, *click*, *post*, *substantial*, *thanks*, *connection*
COCA	*the*, *of*, *in*, *his*, *he*, *says*, *by*, *said*, *president*, *percent*, *national*, *were*, *united*, *we*, *states*, *american*, *york*, *their*, *her*, *students*, *washington*, *million*, *political*, *new*, *government*, *among*, *state*, *john*, *bush*, *three*, *economic*, *from*, *toward*, *war*, *clinton*, *university*, *center*

D. Pipeline: Threshold-Based Decision Rules

This appendix provides the rule-based framework used to make final predictions, as introduced in the Pipeline section.

1. If $\bar{p_k}$, the top-k-averaged probability of the model, is below the lower-bound threshold $t_{lower} = 0.4$, the model seems not confident at all and we decide that the decision is 0 (no diagnosis; we do not check any additional features).
2. If $\bar{p_k}$ is higher than a pre-defined threshold ($t_{upper} = 0.8$) we deem the model confident enough and decide for 1 (diagnosis; again, we do not check additional features).
3. If $t_{lower} \leq \bar{p_k} \leq t_{upper}$:
   a) If the user has $n \leq 5$ messages *or* less than two rounds of conversation are available, we decide 0 due to insufficient data.
   b) If the user has $n > 5$ messages *but* we processed fewer than 2 rounds of conversation for this user, we decide 0 due to insufficient data.
   c) If the user has $n > 5$ messages and there are more than 2 rounds of conversation available, we check the additional linguistic and metadata features. If relevant thresholds are exceeded (e.g., frequent night-time activity, presence of diagnostic verbs), we decide 1, otherwise 0.