Natural Language Processing in the Detection and Treatment of Mental Health Issues

Alba María Mármol-Romero

Computer Science Department, SINAI, CEATIC, Universidad de Jaén, 23071, Spain

Abstract

Mental health issues are an increasing global public health concern. With the rise of social media, there is growing interest in the early detection of mental disorders by analyzing user posts. This research project aims to leverage Artificial Intelligence (AI) to enhance mental wellness. We operate initially under two primary objectives: first, that Natural Language Processing (NLP) can effectively identify signs of mental disorders or emotional distress in user messages over time; and second, that Large Language Models (LLMs) can provide high-quality information to mental health professionals. To test these hypotheses, we explore key research questions, including the feasibility of detecting emotional symptoms in text messages and the capability of chatbots to collect high-quality data for professional use. This study focuses primarily on the Spanish language.

Keywords

Mental Health, Mental Disorders Detection, Large Language Model, Early Risk Detection, Dialogue systems

1. Introduction

According to the World Health Organization (WHO), mental health is a state of mental well-being that enables individuals to cope with life's stresses, realize their abilities, learn and work effectively, and contribute to their communities. The absence of mental health could carry on mental health problems such as eating disorder (ED), depression or anxiety disorder, being the last two most prevalent nowadays. In 2019, one in eight people in the world was diagnosed with one or more mental health issues and it is estimated that only one year before COVID-19 the number of people suffering from anxiety or depression has increased by about 30% [1, 2]. Because of these data, both effective treatment and prevention or early detection of signs of mental health problems are important social branches [3].

In addition, due to the increasing use of social media, people often express their emotional problems or thoughts on the Internet in search of comfort, support or to unwind [4, 5]. Given the large amount of information in text format (natural language) of this type accessible on social media, NLP plays an important role in detecting sentiments and emotions [6] or hate speech [7]. Moreover, the evaluation and treatment of mental health issues also heavily rely on natural language which makes NLP and LLMs potentially valuable tools for interpreting users' mental and emotional states through their written communication [8]. In recent years, research in NLP and computational social science has increasingly focused on detecting mental health issues through online text data, such as social media content [9, 10]. For instance, studies have shown NLP can analyze large datasets from social media platforms like Twitter or Facebook to detect subtle cues of mental health conditions by examining language patterns over time [11].

For this reason, our research focuses on applying NLP and AI to identify in an early way the risk of a user suffering from a disorder in social media and to develop systems and tools to help professionals with mental health in their work. By harnessing the potential of these technologies, we strive not only to improve our understanding of the dynamics of mental health in digital spaces but also to provide scientific and professional communities with the results of our mental well-being efforts. Ultimately, this work aims to contribute to the advancement of the development of tools in this field for Spanish speakers, since the research carried out so far has focused on the English language.

Doctoral Symposium on Natural Language Processing, 26 September 2024, Valladolid, Spain.

amarmol@ujaen.es (A. M. Mármol-Romero)

0000-0001-7952-4541 (A. M. Mármol-Romero)

2. Related work

The detection of whether a person is suffering from a mental disorder and the implementation of treatment are two of the important tasks the scientific community has been tackling in recent years. Still, the speed factor is one of the most important factors that is being taken into account. The early detection of signs of mental disorders is important, since, undetected, mental disorders can develop into more serious consequences, constituting a major predictive factor of suicide [12]. Therefore, both the early detection of a person's risk of suffering from a mental health problem and the rapid implementation of treatment are two growing tasks.

2.1. Early risk detection

The concept of early risk detection involves promptly identifying potential signs of mental health issues, hate speech, and other concerns on social networks [13]. This concept is central to the well-known eRisk shared task eRisk[1] which began in 2017 and is hosted annually at the Conference and Labs of the Evaluation Forum (CLEF). Many studies on early risk prediction originate from eRisk, which has addressed early risk detection of gambling [14], self-harm [15], and disorders such as anorexia or depression [16]. These tasks consist of sequentially processing subjects' posts on Reddit. This simulates a real-time analysis of social media to evaluate the early risk detection capabilities of various systems teams.

So far, several datasets have been developed to identify mental health problems like stress, depression or anxiety [17, 18, 19], as well as the severity of depression [20, 21] or suicide risk [22]. However, to carry out the task of early detection it is necessary to get a user-level dataset that permits the application of measures such as Early Risk Detection Error (ERDE) [23]. Twitter and Reddit are the most studied platforms for mental health research [24], especially in early detection. All eRisk datasets come from Reddit users but there are other datasets annotated at the user level extracted from other social networks like Twitter-STMHD [25] which is a collection of user Twitter profiles suffering from mental health disorders. These datasets often focus on English-language data, but there are exceptions. For instance, Villa-Pérez et al. [26] created two datasets in English and Spanish that comprise the timeline of Twitter users who explicitly reported in one or more of their posts having been diagnosed with one disorder. Additionally, the dataset MentalRiskES [27] is a new open-sourced corpus for the early detection of mental disorders in Spanish, focusing on eating disorders, depression, and anxiety. It consists of user messages posted on groups within the Telegram message platform. This corpus was used by the MentalRiskES shared task [28] organized at the Iberian Languages Evaluation Forum (IberLEF). This task follows the same structure as the eRisk task focusing on the early risk detection of mental disorders.

However, one of the emerging branches in recent years comes from the need for explainability of a risk prediction [29, 30]. To prove a person by their message thread may be at risk of suffering from a mental problem the latest research is focusing not so much on predicting whether or not they suffer from a mental disorder but on whether or not they suffer from certain symptoms. eRisk 2023 edition [31] introduced a novel task that consisted of ranking phrases according to their relevance to standardised symptoms of depression. They used the BDI-Sen [32] dataset which is a corpus label at the Reddit post sentences level and covers all the symptoms present in the Beck Depression Inventory-II (BDI-II). As such, PsySym [33] is another annotated symptom identification corpus of multiple psychiatric disorders for Reddit post sentences in the English language too.

2.2. Dialogue systems for mental health

In the medical field, conversational agents (CAs) are gaining popularity, particularly for addressing health-related inquiries, a functionality that can also be expanded to mental health concerns [34, 35]. So far, dialogue systems have mainly assisted users in performing self-reporting methods, however, it is

[1] https://erisk.irlab.org/

interesting to see how current studies take advantage of the capabilities of conversational agents and smartphones to evaluate these more efficiently [36].

Replika[2] [37] is an AI companion chatbot designed to engage users in meaningful conversations. While not exclusively focused on mental health, many users find comfort in talking to Replika about their feelings and emotions. Replika adapts to the user's conversational style and provides emotional support, making it a versatile tool for mental well-being. But nowadays some popular chatbots such as Woebot[3] [38] provide support for emotional problems using NLP and AI to understand users' emotions. It was originally developed as a study to mitigate symptoms of anxiety and depression in adolescents and, as of now, Woebot is only available to new users in the United States who are participating in the study. Youper[4] [39] is another popular AI-powered emotional health assistant that uses CBT and other therapeutic techniques to guide users through conversations aimed at improving their mental well-being. Youper helps users track their moods, understand their emotions, and develop healthier mental habits. For Spanish speakers, Wysa[5] [40] is a significant chatbot in this space. Like Woebot, Wysa uses NLP and AI to understand users' emotions but extends its support by integrating cognitive behavioural therapy (CBT), mindfulness, and Dialectical Behaviour Therapy (DBT). Available in both English and Spanish, Wysa serves users in 65 countries worldwide.

Studies have shown chatbots can significantly benefit both young people and adults by reducing symptoms of anxiety and depression and improving overall mood [38, 41, 40]. One key factor contributing to these positive outcomes is the concept of self-disclosure [42, 43]. Self-disclosure involves sharing personal and intimate information, which plays a crucial role in building intimacy and trust between individuals [44]. In the context of chatbots, when these digital companions engage in self-disclosure, it not only enhances the perceived intimacy and enjoyment users experience but also fosters a deeper emotional connection. This increased trust can make users feel more understood and supported, thereby boosting their emotional well-being. Furthermore, users often feel more comfortable sharing their concerns and feelings with a chatbot that demonstrates openness, which can lead to a more meaningful and supportive interaction [45, 46]. This reciprocal sharing creates a sense of mutual understanding and empathy, contributing positively to the user's mental health and overall sense of connection.

Despite their benefits, the deployment of chatbots in mental health care presents challenges and ethical considerations. Although people who frequently use these tools trust them and their security more than people who have never used any of them, these chatbots can also be disruptive and introduce risks for users with sensitive questions or disclosure of information [47]. Privacy concerns, the accuracy of the chatbots' responses, and the need for human oversight are significant issues that researchers and developers must address to ensure chatbots provide reliable and safe support.

However, chatbots do not necessarily have to be emotionally involved with the individual, sometimes they are simply useful tools for gathering information beyond applying a self-reported test [48] since users find language is more precise in communicating their mental health issues, preferring it to rating scales [49].

3. Hypotheses and objectives

Given the large number of existing applications of NLP and the use of more advanced techniques such as LLM in the mental health field, our research elaborates on the following premises as scientific hypotheses:

- H1: NLP techniques allow identifying and tracking signs of mental disorders or emotional problems in user-generated text messages over time.
- H2: LLM can provide high-quality, contextually relevant information and support to mental health professionals, enhancing their ability to diagnose and treat patients effectively.

[2] https://replika.ai/
[3] https://woebothealth.com/
[4] https://www.youper.ai/
[5] https://www.wysa.com/

To prove these hypotheses we asked ourselves the following research questions:

- Q1: Can NLP models accurately identify symptoms of emotional problems in text messages?
- Q2: Are there identifiable linguistic features or patterns that are most indicative of these symptoms in the Spanish language?
- Q3: Can we utilize LLMs to induce user self-disclosure in mental health?
- Q4: Can we assess the feasibility of collecting high-quality, clinically relevant data through interactions with chatbots?
- Q5: Can we build chatbots to elicit meaningful mental-health-related information from users while maintaining user trust and engagement?

4. Methodology

As detailed below, in previous research work I developed resources, such as a corpus or a basic dialogue system, which I have been able to develop in this pre-doctoral period. These resources will serve as the foundational elements and support for future work. Additionally, several experiments have been conducted to validate and extend these initial developments.

4.1. Dialogue system

As a preliminary work associated with a research project called BigHug[6], focused on the early detection of disorders and misbehaviours in online social networks, a dialogue system was developed. For this project, the author of this paper developed a novel chatbot [50] to chat about several mental disorders for young Spanish on the Telegram Platform. The most novel aspect of this chat, apart from its ability to converse in Spanish, is that it allows for both closed and open dialogue and also does not present itself as a therapist but as one more teenager who wants to talk about his or her problems. For the open dialogue, we integrated the Generative Pre-trained Transformer (GPT-3) trained mostly on English texts, so we also used DeepL[7] to translate. For the controlled dialogue, we used some questions and sentences established by psychologists and specialists in mental disorders in teenagers.

The dialogue system creation involved a major collaboration, where the full development of the dialogue system was carried out by the author of this paper and now forms a key part of the thesis. It will be a central component of ongoing and future research work related to H2 above. The basic corpus obtained from this experimentation has been used in initial experiments to test and refine the functionalities of the system and to detect needs and strengths. Moreover, nowadays there are generative models of language with greater capacity, which is why we propose a more updated development focused on the needs of a therapist, focusing on the ability to contextualise and synthesise useful information.

4.2. Developed dataset

A new extensive dataset entitled MentalRiskES [27] was developed which contains threads of messages in Spanish. Three collections of data for evaluating early risk detection in three mental disorders (ED, depression, and anxiety) contain more than 45,000 messages sent by over 1,300 subjects from various Telegram groups. This data was annotated crowdsourcing according to the definition of these disorders by remarkable organizations such as WHO and the symptoms of that disorder. So 10 annotators labelled each subject (their last 50 or 100 messages sent to the platform) according to the annotation guideline. This dataset was used in the shared task with the same name MentalRiskES [28] hosted at IberLEF (editions 39 and 40) and is available upon request via GitHub[8].

[6] https://bighug.ujaen.es/
[7] https://www.deepl.com/en/docs-api/
[8] https://github.com/sinai-uja/corpusMentalRiskEs

The creation of this collection of Spanish-language message threads aimed at early risk detection for mental health disorders is directly relevant to H1 of the thesis. The author of the thesis has been directly involved in all phases of the development of this corpus. As future work, it is planned to work in more depth and analyse the dataset created and to develop a new dataset that will allow the detection of symptoms for certain disorders and messages for Spanish.

4.3. Participation in Shared Tasks

We have engaged in the shared task eRisk organized during CLEF conferences to test the hypothesis and gain access to annotated data. We plan to continue to participate in future editions of the task.

- **eRisk 2022** [51]. Two of the proposed tasks were addressed: early detection of signs of pathological gambling, and measuring the severity of the signs of eating disorders. The approach presented for the first task is based on the use of sentence embeddings from Transformers with features related to volumetry, lexical diversity, complexity metrics, and emotion-related scores, while the approach for the second task is based on text similarity estimation using contextualized word embeddings from Transformers [52].
- **eRisk 2023** [31]. One of the proposed tasks was addressed: early detection of signs of pathological gambling. The approach presented is based on pre-trained models from Transformers architecture with comprehensive preprocessing data and data balancing techniques. Moreover, we integrate Long-short Term Memory (LSTM) architecture with automodels from Transformers [53].
- **eRisk 2024** [54]. Two of the proposed tasks were addressed: search for symptoms of depression and early detection of signs of anorexia. The approach presented in the first task is based on the use of a two-step detection approach using a transformer-based model, while the approach for the second is based on calculating perplexity using two transformer-based models trained with causal language modelling. [55].

On the other hand, I was part of the organising committee of MentalRiskES, a shared task organized at IberLEF (edition 39) as part of the International Conference of the Spanish Society for Natural Language Processing (SEPLN). MentalRiskES aim to promote the early detection of mental risk disorders in Spanish.

- **MentalRiskES 2023** [28]. We outline three detection tasks: Task 1 on eating disorders, Task 2 on depression, and Task 3 on an undisclosed disorder during the competition (anxiety) to observe the transfer of knowledge among the different disorders proposed. To establish a baseline benchmark, we performed experiments using three different Transformer-based models. In this edition, 37 teams were registered from 8 different countries, 17 sent their submission and 16 wrote their working notes.
- **MentalRiskES 2024** [56]. We propose three detection tasks: Task 1 to detect risk for depression or anxiety, Task 2 for depression and anxiety but determining contextual risk factors and Task 3 to identify whether a subject is at risk for suicidal ideation. To establish a baseline benchmark, we performed experiments using three different Transformer-based models. In this edition, 28 teams were registered from 10 different countries, 12 sent their submission and 10 wrote their working notes.

5. Research Elements Proposed for Discussion

My research is still at the beginning of the way so there are a lot of questions to address and elements to be proposed and discussed. Some of them are the following:

- Early detection of mental disorders and emotional problems: What are the challenges and limitations of detecting signs of mental health problems in text data? Can early detection methods

be generalised to different mental health conditions or should they be disease-specific? Is it possible to identify symptoms of emotional problems in text messages? How accurately can NLP models detect early signs of specific mental health disorders in text communications?
- Effectiveness of dialogue systems in mental health support: What are the key features and functionalities that should be included in the dialogue system to ensure it is both supportive and safe for users? Can a chatbot be designed to collect high-quality data that is valuable and reliable for mental health professionals? How can the impact of the dialogue system on users' mental health and well-being be measured accurately? In what ways can a LLM be designed to encourage user self-disclosure in a manner that promotes well-being? How do users perceive the use of chatbots and automated systems in their mental health care? What factors influence their acceptance and trust, and how can these systems be designed to align with the standards and practices of certified mental health diagnostics and treatments?
- Ethical and legal considerations: What are the ethical implications of using automated systems for detecting and responding to mental health issues? Is it possible to ensure the system does not inadvertently cause harm or distress to users?

Acknowledgments

My sincere thanks to my thesis tutors, Arturo Montejo-Raéz, Miguel Ángel García-Cumbreras and Manuel García-Vega, for guiding me along this process, to the doctoral programme of the University of Jaén and the Centre for Advanced Studies in Information and Communication Technologies (CEATIC for its acronym in Spanish) for their support in this research experience.

This work has been supported by project MODERATES (TED2021-130145B-I00) funded by Plan Nacional I+D+i from the Spanish Government.

References

[1] World Health Organization, Mental disorders, https://www.who.int/es/news-room/fact-sheets/detail/mental-disorders, 2022. Accessed: 06.06.2024.

[2] A. Kumar, K. R. Nayar, Covid 19 and its mental health consequences, Journal of Mental Health 30 (2021) 1–2. doi:10.1080/09638237.2020.1757052, pMID: 32339041.

[3] J. Rehm, K. D. Shield, Global burden of disease and the impact of mental and addictive disorders, Current psychiatry reports 21 (2019) 1–7.

[4] 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.

[5] 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) 1–13.

[6] S. Zad, M. Heidari, H. James Jr, O. Uzuner, Emotion detection of textual data: An interdisciplinary survey, in: 2021 IEEE World AI IoT Congress (AIIoT), IEEE, 2021, pp. 0255–0261.

[7] F. M. Plaza-del Arco, M. D. Molina-González, L. A. Urena-López, M. T. Martín-Valdivia, Comparing pre-trained language models for spanish hate speech detection, Expert Systems with Applications 166 (2021) 114120.

[8] X. Xu, B. Yao, Y. Dong, S. Gabriel, H. Yu, J. Hendler, M. Ghassemi, A. K. Dey, D. Wang, Mental-llm: Leveraging large language models for mental health prediction via online text data, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 8 (2024) 1–32.

[9] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, J. C. Eichstaedt, Detecting depression and mental illness on social media: an integrative review, Current Opinion in Behavioral Sciences 18 (2017) 43–49.

[10] M. Malgaroli, T. D. Hull, J. M. Zech, T. Althoff, Natural language processing for mental health interventions: a systematic review and research framework, Translational Psychiatry 13 (2023) 309.

[11] S. Henry, M. Yetisgen, O. Uzuner, Natural Language Processing in Mental Health Research and Practice, Springer International Publishing, Cham, 2021, pp. 317–353. URL: https://doi.org/10.1007/978-3-030-70558-9_13. doi:10.1007/978-3-030-70558-9_13.
[12] World Health Organization, Depressive disorder (depression), https://www.who.int/news-room/fact-sheets/detail/depression, 2023. Accessed: 07.06.2024.
[13] D. E. Losada, F. Crestani, J. Parapar, erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings 8, Springer, 2017, pp. 346–360.
[14] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, Overview of erisk at clef 2021: Early risk prediction on the internet (extended overview)., CLEF (Working Notes) (2021) 864–887.
[15] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk 2019 early risk prediction on the internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10, Springer, 2019, pp. 340–357.
[16] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk: early risk prediction on the internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings 9, Springer, 2018, pp. 343–361.
[17] D. Owen, J. C. Collados, L. Espinosa-Anke, Towards preemptive detection of depression and anxiety in twitter, arXiv preprint arXiv:2011.05249 (2020).
[18] A. Haque, V. Reddi, T. Giallanza, Deep learning for suicide and depression identification with unsupervised label correction, in: Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part V 30, Springer, 2021, pp. 436–447.
[19] S. Ji, X. Li, Z. Huang, E. Cambria, Suicidal ideation and mental disorder detection with attentive relation networks, Neural Computing and Applications 34 (2021) 10309–10319. doi:10.1007/s00521-021-06208-y.
[20] I. Pirina, Ç. Çöltekin, Identifying depression on Reddit: The effect of training data, in: G. Gonzalez-Hernandez, D. Weissenbacher, A. Sarker, M. Paul (Eds.), Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 9–12. doi:10.18653/v1/W18-5903.
[21] U. Naseem, A. G. Dunn, J. Kim, M. Khushi, Early identification of depression severity levels on reddit using ordinal classification, in: Proceedings of the ACM Web Conference 2022, Association for Computing Machinery, New York, NY, USA, 2022, p. 2563–2572. doi:10.1145/3485447.3512128.
[22] M. Gaur, A. Alambo, J. P. Sain, U. Kursuncu, K. Thirunarayan, R. Kavuluru, A. Sheth, R. Welton, J. Pathak, Knowledge-aware assessment of severity of suicide risk for early intervention, in: The World Wide Web Conference, Association for Computing Machinery, New York, NY, USA, 2019, p. 514–525. doi:10.1145/3308558.3313698.
[23] 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, Springer International Publishing, Cham, 2016, pp. 28–39.
[24] K. Harrigian, C. Aguirre, M. Dredze, On the state of social media data for mental health research, arXiv preprint arXiv:2011.05233 (2020).
[25] A. K. Singh, U. Arora, S. Shrivastava, A. Singh, R. R. Shah, P. Kumaraguru, et al., Twitter-stmhd: An extensive user-level database of multiple mental health disorders, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 16, 2022, pp. 1182–1191.
[26] M. E. Villa-Pérez, L. A. Trejo, M. B. Moin, E. Stroulia, Extracting mental health indicators from english and spanish social media: A machine learning approach, IEEE Access 11 (2023) 128135–

128152.
[27] A. M. Mármol Romero, A. Moreno Muñoz, F. M. Plaza-del Arco, M. D. Molina González, M. T. Martín Valdivia, L. A. Ureña-López, A. Montejo Ráez, MentalRiskES: A new corpus for early detection of mental disorders in Spanish, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 11204–11214. URL: https://aclanthology.org/2024.lrec-main.978.
[28] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del Arco, M. D. Molina-González, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Raéz, Overview of MentalriskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish, Procesamiento del Lenguaje Natural 71 (2023) 329–350.
[29] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, L. Kagal, Explaining explanations: An overview of interpretability of machine learning, in: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA), IEEE, 2018, pp. 80–89.
[30] A. S. Uban, B. Chulvi, P. Rosso, On the explainability of automatic predictions of mental disorders from social media data, in: International Conference on Applications of Natural Language to Information Systems, Springer, 2021, pp. 301–314.
[31] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, Overview of erisk 2023: Early Risk Prediction on the Internet, in: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2023, pp. 294–315.
[32] A. Pérez, J. Parapar, Á. Barreiro, S. Lopez-Larrosa, Bdi-sen: A sentence dataset for clinical symptoms of depression, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023, pp. 2996–3006.
[33] Z. Zhang, S. Chen, M. Wu, K. Zhu, Symptom identification for interpretable detection of multiple mental disorders on social media, in: Y. Goldberg, Z. Kozareva, Y. Zhang (Eds.), Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 9970–9985. doi:10.18653/v1/2022.emnlp-main.677.
[34] S. Siddique, J. C. L. Chow, Machine learning in healthcare communication, Encyclopedia 1 (2021) 220–239. doi:10.3390/encyclopedia1010021.
[35] J. C. Chow, V. Wong, L. Sanders, K. Li, Developing an ai-assisted educational chatbot for radiotherapy using the ibm watson assistant platform, in: Healthcare, volume 11, MDPI, 2023, p. 2417.
[36] A. I. Jabir, L. Martinengo, X. Lin, J. Torous, M. Subramaniam, L. Tudor Car, Evaluating conversational agents for mental health: scoping review of outcomes and outcome measurement instruments, Journal of Medical Internet Research 25 (2023) e44548.
[37] M. Skjuve, A. Følstad, K. I. Fostervold, P. B. Brandtzaeg, My chatbot companion-a study of human-chatbot relationships, International Journal of Human-Computer Studies 149 (2021) 102601.
[38] K. K. Fitzpatrick, A. Darcy, M. Vierhile, Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial, JMIR Ment Health 4 (2017) e19. doi:10.2196/mental.7785.
[39] A. Mehta, A. N. Niles, J. H. Vargas, T. Marafon, D. D. Couto, J. J. Gross, Acceptability and effectiveness of artificial intelligence therapy for anxiety and depression (youper): Longitudinal observational study, Journal of medical Internet research 23 (2021) e26771.
[40] B. Inkster, S. Sarda, V. Subramanian, An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study, JMIR Mhealth Uhealth 6 (2018) e12106. doi:10.2196/12106.
[41] O. Romanovskyi, N. Pidbutska, A. Knysh, Elomia chatbot: The effectiveness of artificial intelligence in the fight for mental health., in: COLINS, 2021, pp. 1215–1224.
[42] A.-K. Reuel, S. Peralta, J. Sedoc, G. Sherman, L. Ungar, Measuring the language of self-disclosure across corpora, in: S. Muresan, P. Nakov, A. Villavicencio (Eds.), Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, Dublin, Ireland,

2022, pp. 1035–1047. URL: https://aclanthology.org/2022.findings-acl.83. doi:10.18653/v1/2022.findings-acl.83.
[43] T. Blose, P. Umar, A. Squicciarini, S. Rajtmajer, Privacy in crisis: A study of self-disclosure during the coronavirus pandemic, 2020. URL: https://arxiv.org/abs/2004.09717. arXiv:2004.09717.
[44] D. Catona, K. Greene, Self-Disclosure, 2015. doi:10.1002/9781118540190.wbeic162.
[45] Y.-C. Lee, N. Yamashita, Y. Huang, W. Fu, "i hear you, i feel you": Encouraging deep self-disclosure through a chatbot, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–12. doi:10.1145/3313831.3376175.
[46] J. Meng, N. Dai, Emotional support from ai chatbots: Should a supportive partner self-disclose or not?, Journal of Computer-Mediated Communication 26 (2021) 207–222. doi:10.1093/jcmc/zmab005.
[47] P. Chametka, S. Maqsood, S. Chiasson, Security and privacy perceptions of mental health chatbots, in: 2023 20th Annual International Conference on Privacy, Security and Trust (PST), IEEE, 2023, pp. 1–7.
[48] A. Schick, J. Feine, S. Morana, A. Maedche, U. Reininghaus, Validity of chatbot use for mental health assessment: experimental study, JMIR mHealth and uHealth 10 (2022) e28082.
[49] V. Varadarajan, S. Sikström, O. Kjell, H. Schwartz, ALBA: Adaptive language-based assessments for mental health, 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 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 2466–2478. URL: https://aclanthology.org/2024.naacl-long.136.
[50] A. M. Mármol-Romero, M. García-Vega, M. Á. García-Cumbreras, A. Montejo-Ráez, An empathic gpt-based chatbot to talk about mental disorders with spanish teenagers, International Journal of Human–Computer Interaction (2024) 1–17.
[51] P. Martın-Rodilla, D. E. Losada, F. Crestani, Overview of eRisk 2022: Early Risk Prediction on the Internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings, volume 13390, Springer Nature, 2022, p. 233.
[52] A. M. Mármol-Romero, S. M. J. Zafra, F. M. P. del Arco, M. D. Molina-González, M. T. M. Valdivia, A. Montejo-Ráez, Sinai at erisk@ clef 2022: Approaching early detection of gambling and eating disorders with natural language processing., in: Working Notes of CLEF), 2022, pp. 961–971.
[53] A. M. Mármol-Romero, F. del Arco, A. Montejo-Ráez, Sinai at erisk@ clef 2023: Approaching early detection of gambling with natural language processing, Working Notes of CLEF (2023) 18–21.
[54] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, erisk 2024: Depression, anorexia, and eating disorder challenges, in: Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part V, Springer-Verlag, Berlin, Heidelberg, 2024, p. 474–481. doi:10.1007/978-3-031-56069-9_65.
[55] A. M. Mármol-Romero, A. Moreno Muñoz, P. Álvarez-Ojeda, K. M. Valencia-Segura, M.-C. Eugenio, M. García-vega, A. Montejo-Ráez, Sinai at erisk@ clef 2024: Approaching the search for symptoms of depression and early detection of anorexia signs using natural language processing., in: Working Notes of CLEF, 2024. To appear.
[56] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del Arco, M. D. Molina-González, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Raéz, Overview of MentalriskES at IberLEF 2024: Early Detection of Mental Disorders Risk in Spanish, Procesamiento del Lenguaje Natural 73 (2024). To appear.