Foundational Models for Personalised Mental Health

Geoffrey Chern-Yee Tan
Department of Mood and Anxiety Institute of Mental Health
Singapore, Singapore
Psychotherapy_Research@imh.com.sg

Beverly Zhiyu Wang
Data aNalytics & Ai Synapxe
Singapore, Singapore
beverly.wang@synapxe.sg

Hong Ming Tan
Institute of Operations Research and Analytics National University of Singapore
Singapore, Singapore
thm@nus.edu.sg

Laurence S. Pe
Institute of Molecular Biosciences Mahidol University
Salaya, Thailand
laurence.pe13@gmail.com

Alvin Keen Peng Yuen
Data aNalytics & Ai Synapxe
Singapore, Singapore
alvin.yuen1@synapxe.sg

Yuguang Mu
School of Biological Sciences Nanyang Technological University
Singapore, Singapore
YGMu@ntu.edu.sg

Bernett Lee
Centre of Biomedical Informatics Lee Kong Chian School of Medicine
Singapore, Singapore
bernett.lee@ntu.edu.sg

Sharon Lu Huixian
Department of Psychology Institute of Mental Health
Singapore, Singapore
sharon_hx_lu@imh.com.sg

Ethel Siew Ee Tan
Department of Mood and Anxiety Institute of Mental Health
Singapore, Singapore
ethel_SE_TAN@imh.com.sg

Kah Vui Fong
Faculty of Science National University of Singapore
Singapore, Singapore
e0773790@u.nus.edu

Jussi Keppo
Business School National University of Singapore
Singapore, Singapore
keppo@u.nus.edu

Han Leong Goh
Data aNalytics & Ai Synapxe
Singapore, Singapore
hanleong.goh@synapxe.sg

Abstract*—Mental health disorders are heterogenous in presentation and treatment response. For example, only one third of patients started on an antidepressant will achieve remission and each trial of medication can take several weeks. Additionally side effects and the development of chronic conditions such as diabetes or high cholesterol are common. We discuss the potential application of foundation models as developed from electronic medical records (EMRs), large language models (LLMs) and for pharmacogenetics drawing potential links and applications in mental health. In terms of EMRs, the concept of a patient representation has been used across applications such as disease prediction and personalised treatment. These approaches have been applied in mental health to label diseases such as depression and bipolar disorder as well as to predict suicide in risk assessment. We discuss a range of applications for LLMs, from supporting the preprocessing of EMRs for FEMRs, therapy support through transcription and assessment and patient monitoring, and psychoeducation. We discuss the potential applications of biomedical foundation models to precision medicine with pharmacogenetics. Finally, we touch on ways of integrating broad sources of data and outputs from various models.

***Keywords*—Healthcare, Personalised Medicine, Mental Health, Foundation models, Electronic Health Records, Pharmacogenetics**

I. Introduction

Mental health disorders are a leading cause of disability globally. Mental health disorders are heterogenous in presentation and treatment response. For example, only one third of patients started on an antidepressant will achieve remission and each trial of medication can take several weeks [1]. Additionally side effects and the development of chronic conditions such as diabetes or high cholesterol, are common [2].

In psychiatry, there are significant differences between medications in their potential long-term side effects. For example, in first-generation antipsychotics, motor side effects are prominent, while in second generation antipsychotics, metabolic side effects are prominent. Thus, the clinician must weigh the cost-benefit from a combination of available information and medical intuition. Additionally, in recent years, there has been a greater emphasis on a collaborative approach involving a discussion of potential risks with patients. The ability to provide more detailed probabilities from local data and the medical literature can make a significant difference to medical decision-making and the complex discussion about treatment options with the patient. In addition, mental health treatments conform to a biopsychosocial model that may involve multiple interventions that address different facets of the mental health condition.

This research was supported by IMH Research Seed Funding (642–2018), LKCMed-NUSMed-NHG Collaborative Mental Health Research Pilot Grant Call 2020 (MHRPG/2003) FY21 PRENATAL / EARLY CHILDHOOD GRANT CALL (H22P0M0005), and National Healthcare Group through NHG Centre for Medical Technologies & Innovations (NHG CMTi) Medtech Grant (CMTi-23-01-02).

We review the potential application of foundation models (FMs) in the areas of electronic medical records (EMRs), unstructured text data generated in various clinical contexts, generation of text for clinicians and patients, and pharmacogenetics, and discuss approaches to translating them for clinical decision-support in the area of mental health.

II. Foundation models for EMRs

EMRs include comprehensive medical histories of patients, providing researchers promising resources for conducting large-scale, data-driven studies to explore disease patterns and progression. Miotto et al. (2016) introduced a novel approach called "Deep Patient" which utilises deep learning and EMRs to generate a general-purpose patient representation [3]. It can be further applied in personalized prescriptions, disease predictions, drug targeting and clinical trial recruitment. Subsequently, several foundation models have been developed for clinical contexts. One of the challenges to utilising EMR data in modelling is time patterns in longitudinal data. Variations in data generation frequencies and patients' health statuses can lead to instances of data sparsity. A strategy to mitigate this challenge involves the introduction of time intervals, facilitating the conversion of data into standardized time series to be easily consumed by machine learning models.

The utilization of EMR data in modelling is challenging due to high dimensionality, heterogeneity and mixture of structured and unstructured data formats. Direct inputting of raw EMR data into machine learning models requires intensive computation to derive insights and complicates model explainability. An alternative and effective strategy involves the utilization of commonly-used hierarchical classification schemas in the clinical domain, such as the International Statistical Classification of Diseases and Related Health Problems (ICD) in diagnosis data [4], and the Anatomical Therapeutic Chemical Classification (ATC) in medication data [5]. Based on clinical domain expertise, these schemas offer a hierarchical structure enabling further information processing.

Wornow, et al (2023) published a narrative review on clinical foundation models (FM) examining 84 FMs trained on non-imaging EMR data [6]. They categorise clinical FMs into Foundation models for EMRs (FEMRs) and Clinical language models (CLaMs). They define FEMRs as models trained on the timeline of events in a patient's medical history that will output a machine-understandable 'patient representation' known as patient embedding, comprising a fixed length high-dimensional vector condensing patient information.

For example, Med-BERT was based on a structured approach using coded diagnoses from International Classifications of Diseases (ICD) [7]. It utilizes Cerner Health Facts (v17) with a collection of databases from over 600 hospitals and clinics from the United States as the input source of prediction. It was able to predict diagnosis in smaller samples, identify temporal relationships based on visits and perform contextual embedding.

A number of FEMRs have been applied to mental health prediction. Meng et al (2021) applied a temporal deep learning model to perform bidirectional representation learning on EMR (BRLTM) to predict future diagnoses of depression [8]. The model aggregated five heterogeneous and high-dimensional data sources from EMR and processed them in a temporal manner at various prediction windows with a precision-recall area under the curve of 0.76. Zeng et al (2022) applied Claim-PT to predict suicide risk [9]. They first conducted CPT pretraining on next visit prediction, category prediction pooling demographic, medical, pharmacy, utilisation and date information before performing population-specific fine-tuning to achieve an AUC of 0.84 on suicide claims. Li et al (2020) developed multimodal bayesian topic modelling (MixEHR) that was able to predict a bipolar disease label in a Mayo Clinic cohort [10].

III. Large Language Models

Large-language models (LLMs) are AI systems, typically utilising neural network architectures and pre-trained on large amounts of text data to learn associative relationships between words. Clinical language models are like general-purpose LLMs in that they have the ability to perform natural language processing (NLP) tasks and language generation, but they have been trained on clinical and/or biomedical text [6], [11].

ChatGPT is a type of language trained model which entails a large amount of text data that possesses a prospective progression in patient care and therapeutic outcomes. We discuss a range of applications for ChatGPT in the area of mental health from supporting the generation of meaningful structured features or tokens for FEMRs, to various aspects of documentation, assessment, clinical decision support, and patient interfaces such as counselling [12].

In regards to generating features, a number of FMs have the ability to convert diagnoses and medication records into their corresponding classification schemas, particularly when involving free-text data. To address this challenge, various Natural Language Processing (NLP) models such as Long Short-Term Memory (LSTM), BERT, clinicalBERT, and BioBERT, offer viable avenues for the development of auto-encoders. An example of this has been in the use of LLMs to identify social determinants of health (SDoH) such as employment, housing, transportation, parental status, relationship, and social support, which have an important role in influencing health outcomes. However, such information is often documented in the free text of clinical notes. Guevara et al, were able to use models including GPT3.5 and GPT4, in zero- and few-shot settings and FLAN-T5 XL to accurately extract SDoH and identify over 90% of patients with adverse SDoH as compared to ICD codes that only captured 2% [13].

There is a large treatment gap in mental health with many suffering from mental illness not accessing mental health services due in part to a lack of resources. The promise of automated systems like ChatGPT that can provide mental health advice to primary physicians, patients and caregivers is enticing [14]. Attempts have been made to assess LLMs such as ChatGPT across a range of tasks important in psychiatry. Using imposed clinical vignettes, ChatGPT is able to provide differential diagnoses including the correct diagnosis with 93.3% accuracy [15].

In support of documentations in medicine, LLMs can generate clinical notes in a significantly short amount of time with a high level of accuracy, saving much time for physicians to focus on patients [16]. Moreover, LLMs could also be beneficial in intelligent question-answering, providing reliable data and resources for health management, triage, disease screening , and even professional training. On the other hand, Thus it has been suggested that LLMs have promise in interfaces with clinicians by summarising evidence for specific situations and with patients and caregivers for psychoeducation [17].

However there are several limitations to LLMs in psychiatry. LLMs such as ChatGPT, may provide inconsistent and sometimes wrong or made-up answers. Additionally it cannot replicate all aspects of the clinical interaction such as nonverbal communication. Nevertheless, the ability of ChatGPT to integrate electronic records, scientific literature, and updated practice guidelines could extract valuable insight to be used in clinical decision making and personalised treatment plan for individual patients which depicts its promising advancement in diagnosis, drug treatment, and therapeutic outcome [18, 19]. We propose that a fundamental tenet in the use of these tools where outputs can be inconsistent and sometimes unexpected or wrong is that the clinical service provider must ultimately be accountable. This requires that interfaces and the use of LLM outputs are developed in a way that encourages the clinician to exercise their own judgement and to take steps to verify information and suggestions received.

A transformer PaLM2 fine-tuned with medical domain (Med-PaLM2), without prior training, showed the ability to diagnose depressive disorder and PTSD according to the criteria and cases of DSM-5 and Distress Analysis Interview Corpus Wizard of Oz (DAIC-WOZ) as the main source of analysis [20]. In fact, this finding of Med-PaLM2 showed high accuracy and indistinguishable from that of the findings of human rater. Med-PaLM2 is a type of LLM which is trained in large data of general medical knowledge. Hence, it labels categories of possible diagnoses and validates description of the disorder using the words that are likely associated with the disorder and provides a summary of rationale for the model decision. This is particularly beneficial in describing heterogenic psychiatric symptoms when linguistic descriptions are ambiguous.

Additionally, there is the potential for Med-PaLM2's applications to extend beyond initial diagnosis. One such application is in the monitoring of ongoing treatment progress. By continuously analysing patient data, it can help track symptom progression, allowing clinicians to provide timely interventions. Med-PaLm2 can also support personalised treatment plans, for instance, by suggesting specific therapeutic approaches or medication adjustments based on a patient's specific profile and treatment history. This capability can enhance the precision of mental health care, potentially making treatments more targeted and effective.

There are various methods on keeping medical records of patients. However, current clinical notes are sparse and high dimensional limiting its applicability to machine learning. BERT can extract data to generate contextualised representation of medical notes that use distant words which typically present vagueness [21]. The use of NLP models demonstrated improvement in addressing medical concerns from diagnosis to treatment. ClinicalBERT utilises the pre-trained BERT and employs bidirectional transformers allowing more accurate findings. In this context, it focuses on the risk assessment for readmission. Using the Medical Information Mart for Intensive Care III (MIMIC-III), clinicalBERT was able to accurately assess cases based on clinical texts and provide semantic relationships similar to ones postulated by physicians [22]. Additionally, it provided precise prediction of readmission within 30 days using data including early clinical notes 24-72 hrs after admission and discharge summaries. ClinicalBERT displays high recall ability at a fixed rate of false alarm with a precision set at 80% or 20% false positive, therefore, minimising false positive rate. Furthermore, when the Unified Medical Language System was added to clinicalBERT, the medical natural language inference (MedNLP) showed 84% accuracy on predicting whether the assumption was true [23]. Therefore, clinicalBERT can predict whether the treatment was effective for the diagnosis and can aid in the high relapse rate and readmission in psychiatry.

IV. Speech Recognition and Natual Language Understanding

There are several sources of unstructured data in mental health. Counselling and psychotherapy in particular can generate large amounts of audio and video speech data. Large amounts of time can be spent documenting and understanding the various factors that have caused and led to their condition. The ability to use natural language understanding (NLU) to identify themes in a particular patient and a particular session and to make causal inferences can be a powerful application.

Furthermore, aside from being time-consuming, manual transcription may lead to errors due to linguistic aberrations. The use of automated analysis of speech can identify patterns to be measured which is beneficial in various mental illnesses that present disorganized speech such as psychosis [24]. This is particularly useful in the early phase to track the stage and progression of the disorders [25]. Given the subjective approach in psychiatry, the application of speech recognition could distinguish and quantitate formal thought disorders and be managed more objectively and proficiently. In fact, this automation in speech recognition is applicable to non-English language and still able to identify and classify schizophrenia [26]. Therefore, NLU can aid in early detection of mental health symptoms and automate real-time transcription of therapy sessions. This increases the efficiency of documentation, helping to extract the most salient themes, which allows psychiatrists to focus more on patient care.

Kintsugi Voice utilises vocal biomarkers to detect “signs of clinical depression and anxiety from short clips of free-form speech” [27]. This tool likely uses speech recognition to accurately translate speech to text, while NLU is used for sentiment analysis to understand speech patterns. Despite its subtlety, speech patterns are more objective and could provide additional insights into a patient’s mental state beyond what they report. Thus, Kintsugi Voice could allow for more objective assessment of one’s mental health. Individuals can record short clips of their voice to be analyzed through Kintsugi Voice. This could facilitate remote monitoring of mental health symptoms without needing in-person appointments. This could help to alleviate the workload of medical professionals and provide a more comprehensive method of assessment.

V. Data Security and Privacy

A major barrier to the application of foundational models (FMs) and GenAI to the healthcare context relates to data security. Analysis of sensitive information such as unstructured medical notes and audio or video recordings must be compliant to the Personal Data Protection Act (PDPA). It includes de-identifying personally identifiable information (PII) by removing or masking identifiers like names and addresses. PDPA also requires strong data protection measures like encryption, access controls, and regular audits. In healthcare, this may involve using advanced technologies like differential privacy to analyse data while keeping individuals' identities private. Thus there is a need to be able to provide data for the fine-tuning of models without weights and data being stored outside of the system. Several strategies can be employed to enhance data security in healthcare AI applications. One way is the masking of personal information with pseudonyms that can only be reversed by authorised personnel. Additionally, PDPA compliance often involves displaying data only in aggregate form, rather than individually, to further protect individuals' privacy.

VI. Biomedical Text-Mining

Despite the number of NLP systems, the majority of deep base learning models are trained on general texts which are distinguishable from biomedical corpora [28]. Nevertheless, BERT shows promising performance on different NLP tasks with the same structure, hence, adapting it in biomedical concepts may potentially provide advantages on biomedical text mining.

For example, Bio-BERT was trained on PMC and PubMed to perform biomedical text mining tasks including name entity recognition (NER), relation extraction (RE), and question answering [29]. It has been applied to a drug-drug interaction (DDI) corpus to detect pharmacokinetics and pharmacodynamics of drugs and to provide knowledge on possible interactions of medications [30]. These relationships could in turn be integrated with genetics data in order to inform multi-modal models estimating response to medication treatments. Bio-BERT has also been shown to be able to accurately predict human phenotypes-gene relationships [31].

VII. Integrating Inputs from a Range of Models

In healthcare, the development of FMs, leveraging a diverse array of machine learning techniques and neural network architectures, has emerged as a promising avenue for advancing clinical decision support and predictive analytics. The following shows an example of how neural networks can be applied to construct foundation models in the clinical domain. This model entails the utilization of pre-processed EMR data at the input, with the output layer designed to accommodate multi-label predictions such as prognostic assessment of various diseases and treatment outcomes.

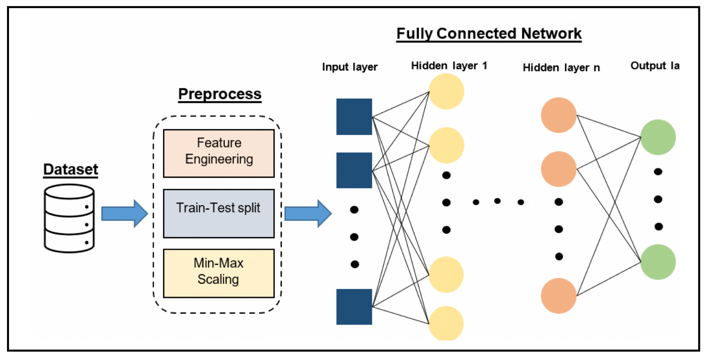

Fig. 1. Framework for pre-processing and fine-tuning of multimodal outputs from foundation models for specific tasks.

This model structure may extend to the integration of heterogeneous data sources beyond EMR, such as genomic information and medical imaging data like X-rays and MRI scans. To accommodate the different nature of this data, distinct neural network architectures, such as Convolutional neural networks (CNNs) and Recurrent neural networks (RNNs), could be employed for pre-processing tasks. This integration with more data sources offers promise with enhanced predictive capabilities across a spectrum of healthcare scenarios.

VIII. Impact and Conclusions

We provide an overview of the range of applications of foundation models in healthcare and draw links for how they may be used in the context of mental health.

Acknowledgment

This research was supported by IMH Research Seed Funding (642–2018), LKCMed-NUSMed-NHG Collaborative Mental Health Research Pilot Grant Call 2020 (MHRPG/2003) FY21 PRENATAL / EARLY CHILDHOOD

GRANT CALL (H22P0M0005), and National Healthcare Group through NHG Centre for Medical Technologies & Innovations (NHG CMTi) Medtech Grant (CMTi-23-01-02).

References

[1] H. B. Maier *et al.*, "Validation of the predictive value of BDNF -87 methylation for antidepressant treatment success in severely depressed patients—a randomized rater-blinded trial," *Trials*, vol. 25, no. 1, p. 247, Apr. 2024, doi: 10.1186/s13063-024-08061-5.

[2] C. U. Correll, J. Detraux, J. De Lepeleire, and M. De Hert, "Effects of antipsychotics, antidepressants and mood stabilizers on risk for physical diseases in people with schizophrenia, depression and bipolar disorder," *World Psychiatry*, vol. 14, no. 2, pp. 119–136, 2015, doi: 10.1002/wps.20204.

[3] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," *Sci. Rep.*, vol. 6, no. 1, p. 26094, May 2016, doi: 10.1038/srep26094.

[4] J. H. B. Masud, C.-C. Kuo, C.-Y. Yeh, H.-C. Yang, and M.-C. Lin, "Applying Deep Learning Model to Predict Diagnosis Code of Medical Records," *Diagnostics*, vol. 13, no. 13, Art. no. 13, Jan. 2023, doi: 10.3390/diagnostics13132297.

[5] I. Reinecke *et al.*, "Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation," *JMIR Med. Inform.*, vol. 11, p. e40312, Jan. 2023, doi: 10.2196/40312.

[6] M. Wornow *et al.*, "The shaky foundations of large language models and foundation models for electronic health records," *Npj Digit. Med.*, vol. 6, no. 1, pp. 1–10, Jul. 2023, doi: 10.1038/s41746-023-00879-8.

[7] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," *Npj Digit. Med.*, vol. 4, no. 1, pp. 1–13, May 2021, doi: 10.1038/s41746-021-00455-y.

[8] Y. Meng, W. Speier, M. K. Ong, and C. W. Arnold, "Bidirectional Representation Learning From Transformers Using Multimodal Electronic Health Record Data to Predict Depression," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 8, pp. 3121–3129, Aug. 2021, doi: 10.1109/JBHI.2021.3063721.

[9] X. Zeng, S. L. Linwood, and C. Liu, "Pretrained transformer framework on pediatric claims data for population specific tasks," *Sci. Rep.*, vol. 12, no. 1, p. 3651, Mar. 2022, doi: 10.1038/s41598-022-07545-1.

[10] Y. Li *et al.*, "Inferring multimodal latent topics from electronic health records," *Nat. Commun.*, vol. 11, no. 1, p. 2536, May 2020, doi: 10.1038/s41467-020-16378-3.

[11] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nat. Med.*, vol. 29, no. 8, pp. 1930–1940, Aug. 2023, doi: 10.1038/s41591-023-02448-8.

[12] S.-W. Cheng *et al.*, "The now and future of ChatGPT and GPT in psychiatry," *Psychiatry Clin. Neurosci.*, vol. 77, no. 11, pp. 592–596, 2023, doi: 10.1111/pcn.13588.

[13] M. Guevara *et al.*, "Large language models to identify social determinants of health in electronic health records," *Npj Digit. Med.*, vol. 7, no. 1, pp. 1–14, Jan. 2024, doi: 10.1038/s41746-023-00970-0.

[14] R. S.-Y. Wong, "ChatGPT in psychiatry: promises and pitfalls," *Egypt. J. Neurol. Psychiatry Neurosurg.*, vol. 60, no. 1, p. 14, Jan. 2024, doi: 10.1186/s41983-024-00791-2.

[15] A. Rao *et al.*, "Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study," *J. Med. Internet Res.*, vol. 25, no. 1, p. e48659, Aug. 2023, doi: 10.2196/48659.

[16] J. Liu, C. Wang, and S. Liu, "Utility of ChatGPT in Clinical Practice," *J. Med. Internet Res.*, vol. 25, p. e48568, Jun. 2023, doi: 10.2196/48568.

[17] D. Kalla, N. Smith, F. Samaah, and S. Kuraku, "Study and Analysis of Chat GPT and its Impact on Different Fields of Study." Rochester, NY, Mar. 01, 2023. Accessed: Apr. 30, 2024. [Online]. Available: https://papers.ssrn.com/abstract=4402499

[18] Y. Wei, L. Guo, C. Lian, and J. Chen, "ChatGPT: Opportunities, risks and priorities for psychiatry," *Asian J. Psychiatry*, vol. 90, p. 103808, Dec. 2023, doi: 10.1016/j.ajp.2023.103808.

[19] A. Le Glaz *et al.*, "Machine Learning and Natural Language Processing in Mental Health: Systematic Review," *J. Med. Internet Res.*, vol. 23, no. 5, p. e15708, May 2021, doi: 10.2196/15708.

[20] I. R. Galatzer-Levy, D. McDuff, V. Natarajan, A. Karthikesalingam, and M. Malgaroli, "The Capability of Large Language Models to Measure Psychiatric Functioning." arXiv, Aug. 03, 2023. doi: 10.48550/arXiv.2308.01834.

[21] Y. Si, J. Wang, H. Xu, and K. Roberts, "Enhancing clinical concept extraction with contextual embeddings," *J. Am. Med. Inform. Assoc. JAMIA*, vol. 26, no. 11, pp. 1297–1304, Nov. 2019, doi: 10.1093/jamia/ocz096.

[22] K. Huang, J. Altosaar, and R. Ranganath, "ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission." arXiv, Nov. 28, 2020. doi: 10.48550/arXiv.1904.05342.

[23] B. Hao, H. Zhu, and I. Paschalidis, "Enhancing Clinical BERT Embedding using a Biomedical Knowledge Base," in *Proceedings of the 28th International Conference on Computational Linguistics*, D. Scott, N. Bel, and C. Zong, Eds., Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 657–661. doi: 10.18653/v1/2020.coling-main.57.

[24] García Molina, José Tomás, Pablo A. Gaspar, and Alicia Figueroa-Barra. "Automatic Speech Recognition in Psychiatric Interviews: A Rocket to Diagnostic Support in Psychosis." *Revista Colombiana de Psiquiatría*, February 7, 2024. https://doi.org/10.1016/j.rcp.2023.12.002.

[25] Boer, J. N. de, M. van Hoogdalem, R. C. W. Mandl, J. Brummelman, A. E. Voppel, M. J. H. Begemann, E. van Dellen, F. N. K. Wijnen, and I. E. C. Sommer. "Language in Schizophrenia: Relation with Diagnosis, Symptomatology and White Matter Tracts." *NPJ Schizophrenia* 6, no. 1 (April 20, 2020): 10. https://doi.org/10.1038/s41537-020-0099-3.

[26] Figueroa-Barra, Alicia, Daniel Del Aguila, Mauricio Cerda, Pablo A. Gaspar, Lucas D. Terissi, Manuel Durán, and Camila Valderrama. "Automatic Language Analysis Identifies and Predicts Schizophrenia in First-Episode of Psychosis." *Schizophrenia (Heidelberg, Germany)* 8, no. 1 (June 1, 2022): 53. https://doi.org/10.1038/s41537-022-00259-3.

[27] "Voice Biomarker Technology — Kintsugi." Accessed June 1, 2024. https://www.kintsugihealth.com/solutions/kintsugivoice.

[28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.

[29] J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/bioinformatics/btz682.

[30] Y. Zhu, L. Li, H. Lu, A. Zhou, and X. Qin, "Extracting drug-drug interactions from texts with BioBERT and multiple entity-aware attentions," *J. Biomed. Inform.*, vol. 106, p. 103451, Jun. 2020, doi: 10.1016/j.jbi.2020.103451.

[31] D. Sousa, A. Lamurias, and F. M. Couto, "A Silver Standard Corpus of Human Phenotype-Gene Relations," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1487–1492. doi: 10.18653/v1/N19-1152.