A Comprehensive Review of Datasets for Clinical Mental Health AI Systems

Aishik Mandal[1,2]**, Prottay Kumar Adhikary**[3,4]**, Hiba Arnaout**[1]**, Iryna Gurevych**[1]**, and Tanmoy Chakraborty**[3,4,*]

[1]Ubiquitous Knowledge Processing Lab (UKP Lab), Department of Computer Science and Hessian Center for AI (hessian.AI), Technische Universität Darmstadt
[2]National Research Center for Applied Cybersecurity ATHENE, Germany
[3]Department of Electrical Engineering, Indian Institute of Technology Delhi, India
[4]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, India
[*]Corresponding author: Tanmoy Chakraborty (tanchak@iitd.in)

ABSTRACT

Mental health disorders are rising worldwide. However, the availability of trained clinicians has not scaled proportionally, leaving many people without adequate or timely support. To bridge this gap, recent studies have shown the promise of Artificial Intelligence (AI) to assist mental health diagnosis, monitoring, and intervention. However, the development of efficient, reliable, and ethical AI to assist clinicians is heavily dependent on high-quality clinical training datasets. Despite growing interest in data curation for training clinical AI assistants, existing datasets largely remain scattered, under-documented, and often inaccessible, hindering the reproducibility, comparability, and generalizability of AI models developed for clinical mental health care. In this paper, we present the first comprehensive review of clinical mental health datasets relevant to the training and development of AI-powered clinical assistants. We categorize these datasets by mental disorders (e.g., depression, schizophrenia), data modalities (e.g., text, speech, physiological signals), task types (e.g., diagnosis prediction, symptom severity estimation, intervention generation), accessibility (public, restricted or private), and sociocultural context (e.g., language and cultural background). Along with these, we also investigate synthetic clinical mental health datasets. We identify critical gaps such as a lack of longitudinal data, limited cultural and linguistic representation, inconsistent collection and annotation standards, and a lack of modalities in synthetic data. We conclude by outlining key challenges in curating and standardizing future datasets and provide actionable recommendations to facilitate the development of more robust, generalizable, and equitable mental health AI systems.

Introduction

Mental health disorders are a growing global concern, affecting millions and placing increasing pressure on already strained healthcare systems. Effective treatment typically involves trained clinicians conducting assessments, developing individualized care plans, and engaging in ongoing therapeutic sessions. However, such clinician-centered care is time-intensive, and a persistent global shortage of mental health professionals leaves many individuals without adequate or timely support. To address this care gap, there is growing interest in applying artificial intelligence (AI) to augment mental health diagnosis, monitoring, and intervention. AI systems offer the potential to improve clinical decision-making, scale access to care, and personalize treatment. However, the development of reliable and generalizable mental health AI tools critically depends on access to high-quality training and evaluation datasets.

The collection and dissemination of clinical mental health data is hampered by stringent privacy regulations, such as the General Data Protection Regulation (GDPR)[1] in the European Union and the Health Insurance Portability and Accountability Act (HIPAA)[2] in the United States. These frameworks place strict limitations on the use and sharing of sensitive health information, complicating the creation of large, representative datasets that meet ethical and legal standards. Early data-driven approaches often turned to social media platforms such as Reddit[3], Twitter[4], and YouTube[5] to compile mental health datasets[6]. While these sources offer scale and accessibility, they generally lack clinical validation and may not accurately reflect diagnosed populations, limiting their clinical applicability.

In recent years, researchers have begun constructing clinically grounded datasets through collaborations with mental health institutions, often under strict privacy-preserving protocols and with informed consent. These datasets, derived from electronic health records, clinician notes, therapy transcripts, and physiological or neuroimaging data, are often shared *only* upon request or remain private, following extensive pseudonymization procedures. Despite increased interest in such resources, there is no comprehensive review focused specifically on clinical mental health datasets for AI development. Previous reviews have

Review	Focus	Data Source	#Datasets	Aspects covered					Synthetic Data
				D	A	T	M	C	
Garg et al.[20]	Algorithms, Datasets	SM	17	✓	✓	✗	✗	✗	✗
Ahmed et al.[21]	Datasets	Clinical, SM	23	✓	✓	✓	✗	✓	✗
Thieme et al.[22]	HCI	Clinical, SM	NA	✓	✗	✓	✓	✗	✗
El-Sappagh et al.[23]	Responsible AI	Clinical	49	✓	✓	✓	✓	✗	✗
Uyanik et al.[24]	Multimodal Algorithms	Clinical	15	✓	✗	✗	✗	✗	✗
Tyagi et al.[25]	Multimodal Algorithms	Clinical	8	✗	✗	✗	✗	✗	✗
Al Sahili et al.[26]	Multimodal Algorithms, Datasets	YT, Movie	26	✓	✗	✗	✓	✓	✗
Graham et al.[12]	AI Algorithms	EHR, SM	28	✓	✗	✓	✓	✗	✗
Dehbozorgi et al.[13]	AI Algorithms	Apps, Chatbots	15	✓	✗	✓	✓	✗	✗
Shatte et al.[14]	ML Algorithms	EHR, SM	300	✓	✗	✓	✓	✗	✗
Su et al.[15]	DL Algorithms	Clinical, SM	57	✓	✗	✓	✓	✗	✗
Iyortsuun et al.[16]	ML/DL Algorithms	EHR, SM	33	✓	✓	✓	✓	✗	✗
Scherbakov et al.[17]	NLP Algorithms	Notes, SM	1768	✓	✓	✓	✓	✗	✗
Laricheva et al.[18]	NLP Algorithms	Transcripts	41	✓	✗	✓	✓	✗	✗
Malgaroli et al.[19]	NLP Algorithms	Clinical, Apps	102	✓	✓	✓	✓	✗	✗
Ours	Datasets	Clinical	89	✓	✓	✓	✓	✓	✓

Table 1. Compared to prior reviews, our review uniquely focuses on clinical mental health datasets, offering a systematic analysis by disorder type, data accessibility, task type, modality, and cultural context. We are also the first to review synthetic data generation for clinical mental health applications. Here D = Mental Disorders, A = Access Level, T = Task type, M = Modalities and C = Cultural background. SM represent Social Media and EHR denotes Electronic Health Record.

addressed broader applications of AI in mental health[7], or specific methodological areas such as machine learning[8], deep learning[9], natural language processing (NLP)[10], and large language models (LLMs)[11].

Although some recent review articles (see Table 1) have incorporated data-centric perspectives, they often treat datasets as secondary to model performance or application domains. Reviews have examined resources spanning electronic health records, imaging, social media, wearables, and digital platforms[12–16], or focused on natural language processing[17–19], yet few provide a systematic analysis of dataset characteristics. Key considerations, such as accessibility (public, restricted, private), cultural and linguistic representation, and the emerging role of synthetic data, remain largely unaddressed. In contrast, our review provides a dedicated and comprehensive examination of clinical mental health datasets for AI development. We categorize datasets by disorder type, access level, task formulation, data modality, and sociocultural context, then identify gaps that limit model generalizability and clinical relevance. We further highlight current efforts in synthetic data generation to address privacy and data scarcity. By shifting the focus from algorithms to data foundations, our work offers actionable guidance for curating and evaluating datasets that support robust, ethical, and equitable mental health AI systems (see Figure 2).

To identify relevant resources for this survey, we conducted a systematic search across Google Scholar, DBLP, and the ACL Anthology, using a set of targeted keywords covering both general and disorder-specific domains. These included terms such as "Mental Health AI Datasets," "Multimodal Mental Health AI Datasets," "Mental Health AI Datasets for Depression," "Multimodal Mental Health AI Datasets for Depression," "Mental Health AI Datasets for Anxiety," "Multimodal Mental Health AI Datasets for Anxiety," "Mental Health AI Datasets for Bipolar Disorder," "Multimodal Mental Health AI Datasets for Bipolar Disorder," "Mental Health AI Datasets for PTSD," "Multimodal Mental Health AI Datasets for PTSD," "Mental Health AI Datasets for Schizophrenia," "Multimodal Mental Health AI Datasets for Schizophrenia," "MRI Mental Health AI Datasets," and "EEG Mental Health AI Datasets". This search initially yielded 560 records, which, after deduplication and screening for relevance, were narrowed down to 89 datasets that met our selection criteria of containing real-world clinical mental health data applicable to AI research.

Dataset	Modalities	Disorder	Task	Size (P)	Imb (Dis./HC)	Culture	Access
LabWriting[27]	T	SZ	BC	373 (188)	93/95	US	Private

Study	Modality	Disorder	Task	Samples	Class	Country	Access
Gehrmann et al.[28]	T	Dep. & SZ	BC	41,000 (41,000)	NA	US	Restricted
WorryWords[29]	T	Anx.	BC	44,450 (44,450)	NA	Canada	Public
Hou et al.[30]	T	Dep.	Analysis	376 (376)	128/248	China	Private
CHSN[31]	T	Dep. & Anx. & BD	BC, Analysis	6M (1M)	NA	US	Private
PRIORI[32]	A	Dep. & BD	MC	34,830 (37)	NA	US	Private
FME Hospital Dataset[33]	A	PTSD	MC	200 (200)	150/50	Taiwan	Public
PTSD Speech Corpus[34]	A	PTSD	BC	26 (26)	8/18	US	Private
German LMU Dataset[35]	A	PTSD	BC	15 (15)	7/8	Germany	Private
Vergyri et al.[36]	A	PTSD	BC	39 (39)	15/24	US	Private
Marmar et al.[37]	A	PTSD	BC	129 (129)	52/77	US	Private
Hu et al.[38]	A	PTSD	BC/QS	136 (136)	76/60	China	Private
RADAR-MDD[39]	A	Dep.	Analysis	585 (585)	NA	UK/ Spain/ Netherlands	Restricted
Chang et al.[40]	A	Dep.	Analysis	62 (25)	25/0	US	Private
Broek et al.[41]	A	PTSD	MC	24 (24)	24/0	Netherlands	Private
Salekin et al.[42]	A	Dep. & Anx.	BC	105 (105)	60/45	US	Private
Abbas et al.[43]	V	SZ	BC/QS	27 (27)	18/9	US	Private
Shafique et al.[44]	V	Anx.	MC	50 (50)	40/10	Pakistan	Public
Langer et al.[45]	V	Anx.	Analysis	114 (114)	65/49	US	Private
Pampouchidou et al.[46]	V	Dep. & Anx.	BC/QS	322 (65)	20/45	Greece	Private
Jiang et al.[47]	V	Dep.	BC	365 (12)	NA	US	Private
Gilanie et al.[48]	V	BD	BC	502 (502)	310/192	Pakistan	Private
EATD[49]	T, A	Dep.	BC	162 (162)	30/132	China	Public
Aloshban et al.[50]	T, A	Dep.	BC	59 (59)	29/30	Italy	Private
MMPsy[51]	T, A	Dep.& Anx.	BC	11,983 (11,983)	1,557/10,426	China	Public
Tang et al.[52]	T, A	SZ	BC	31 (31)	20/11	US	Restricted
Wawer et al.[53]	T, A	SZ	BC	94 (94)	47/47	Poland	Private
Hong et al.[54]	T, A	SZ	BC	201 (39)	23/16	US	Private
Allende-Cid et al.[55]	T, A	SZ	BC	189 (63)	13/50	Chile	Private
Iter et al.[56]	T, A	SZ	BC	14 (14)	9/5	US	Private
Elvevåg et al.[57]	T, A	SZ	Analysis	83 (83)	53/30	US	Private
Bedi et al.[58]	T, A	SZ	BC	34 (34)	5/29	US	Private
Elvevag et al.[59]	T, A	SZ	Analysis	51 (51)	26/25	US	Private
Li et al.[60]	T, A	SZ	BC/QS	63 (63)	38/25	China	Private
Ciampelli et al.[61]	T, A	SZ	BC/QS	163 (163)	93/70	Netherlands	Private
Cabuk et al.[62]	T, A	SZ	Analysis	76 (76)	38/38	Turkey	Private
Xu et al.[63]	T, A	SZ	BC/QS	75 (75)	50/25	Singapore	Private
Jeong et al.[64]	T, A	SZ	Analysis	22 (7)	7/0	Canada	Private
Parola et al.[65]	T, A	SZ	Analysis	387 (387)	187/200	Denmark/ China/ Germany	Private
Aich et al.[66]	T, A	SZ & BD	MC	1288 (644)	247 SZ 286 BD 110 HC	US	Public

Name	Modality	Disorder	Task	Size	Split	Country	Availability
Arslan et al.[67]	T, A	BD	MC	143 (143)	53 FEP 40 FEBD 50HC	Turkey	Private
DEPAC[68]	T, A	Dep. & Anx.	Analysis	2674 (571)	NA	Canada	Private
Ex-ray[69]	T, A	SZ	BC	56 (56)	47/9	Australia	Private
Hayati et al.[70]	T, A	Dep.	BC	53 (53)	11/42	Malaysia	Private
Jiang et al.[71]	A, V	Dep. & Anx.	BC	73 (73)	51/22	US	Private
AViD Corpus[72]	A, V	Dep.	QS	340 (292)	NA	Germany	Restricted
Pittsburgh[73]	A, V	Dep.	MC	130 (49)	49/0	US	Restricted
Black Dog Institute[74]	A, V	Dep.	BC	60 (60)	30/30	Australia	Private
Lin et al.[75]	A, V	Dep. & Anx.	BC/QS	35 (35)	18/17	UK	Restricted
Guo et al.[76]	A, V	Dep.	BC	208 (208)	104/104	China	Restricted
E-DAIC[77]	T, A, V	Dep.& PTSD	QS	275 (275)	66/209	US	Restricted
CMDC[78]	T, A, V	Dep.	BC/QS	78 (78)	26/52	China	Public
VH DAIC[79]	T, A, V	Dep. & PTSD	BC	53 (53)	22/31(PTSD) & 17/36(Dep.)	US	Private
Schultebraucks et al.[80]	T, A, V	Dep. & PTSD	BC	81 (81)	NA	US	Private
MEDIC[81]	T, A, V	Coun.	BC	38 (10)	NA	China	Restricted
BDS[82]	T, A, V	BD	BC/MC	95 (95)	46/49	Turkey	Restricted
Chuang et al.[83]	T, A, V	SZ	QS	37 (26)	26/0	Taiwan	Private
Premananth et al.[84]	T, A, V	SZ	MC	140 (40)	30/16	US	Private
Zhang et al.[85]	T, A, V	SZ	BC	160 (40)	20/20	China	Restricted
Tao et al.[86]	T, A, V	Dep. & Anx.	BC	139 (139)	64/75	China	Private
MODMA[87]	A, EEG	Dep.	QS	52 (52)	23/29	China	Restricted
VerBIO[88]	A, EEG	Anx.	Analysis	344 (344)	NA	US	Restricted
COBRE[89]	sMRI, fMRI	SZ	BC	146 (146)	72/74	US	Public
Park et al.[90]	EEG	Dep. & Anx. & SZ	BC	945 (945)	NA	South Korea	Public
RepOD[91]	EEG	SZ	Analysis	28 (28)	14/14	Poland	Public
SchizConnect[92]	EEG	SZ	MC	1392 (1392)	NA	US	Public
UCLA[93]	fMRI, sMRI	SZ & BD	MC	229 (229)	50 SZ 49 BD 130 HC	US	Restricted
NUSDAST[94]	MRI	SZ	BC	341 (341)	171/170	US	Public
MCIC[95]	fMRI, sMRI	SZ	BC, Analysis	331 (331)	162/169	US	Public
MLSP2014[96]	fMRI, sMRI	SZ	BC	144 (144)	69/75	US	Public
FBIRN[97]	fMRI, sMRI	SZ	BC	256 (256)	128/128	US	Public
Cai et al.[98]	EEG	Dep.	BC	213 (213)	92/121	China	Private
Pend et al.[99]	EEG	Dep.	BC	55 (55)	27/28	China	Private
Zhu et al.[100]	A, V, EEG	Dep.	BC	51 (51)	24/27	China	Private
Mumtaz et al.[101]	EEG	Dep.	BC	64 (64)	34/30	Malaysia	Public
Cavanagh et al.[102]	EEG	Dep.	BC	122 (122)	46/76	US	Public
Luo et al.[103]	EEG	Dep.	BC	40 (40)	18/22	China	Private
Garg et al.[104]	EEG	Dep.	BC	120 (120)	62/58	Malaysia	Public
Li et al.[105]	EEG	Dep.	BC	140 (140)	70/70	China	Restricted

Shen et al.[106]	EEG	Dep.	BC	35 (35) 170 (170) 214(214)	15/20 81/89 105/109	China	Private
Chung et al.[107]	EEG	Dep.	BC	214 (67)	49/18	Taiwan	Restricted
PRED + CT[108]	EEG	Dep.	BC	119 (119)	44/75	US	Public
Ros et al.[109]	EEG	PTSD	BC	50 (50)	20/30	Canada	Private
Nicholson et al.[110]	EEG	PTSD	BC	73 (73)	41/32	Canada	Private
Kim et al.[111]	EEG	SZ	BC	238 (238)	119/119	South Korea	Private
Barros et al.[112]	EEG	SZ	BC	128 (128)	65/63	US	Public
Borisov et al.[113]	EEG	SZ	BC	84 (84)	45/39	Russia	Public
MPRC[114]	EEG	SZ	BC	78 (78)	46/32	US	Public
SRPBS[115]	MRI	Dep. & SZ	Analysis	2030 (2030)	450 BD 159 SZ 1421 HC	Japan	Restricted

Table 2. We provide a detailed summary of clinical mental health datasets, outlining key characteristics: modality (T: text, A: audio, V: video), associated disorder (SZ: schizophrenia, Dep.: depression, Anx.: anxiety, BD: bipolar disorder, PTSD: post-traumatic stress disorder), task type (BC: binary classification, MC: multi-class classification, QS: questionnaire score prediction, Reg.: regression), dataset size and number of participants (P), data imbalance (Imb.; Dis/HC = participants with disorder / healthy controls, FEP: first episode psychosis, FEBD: first episode bipolar disorder), cultural context, and accessibility level. Coun. stands for counseling.

Mental Disorders

According to the World Health Organization (WHO), mental disorders are clinically significant disturbances in cognition, emotion regulation, or behavior, often resulting in distress or impaired functioning. A 2019 WHO report [1] estimated one in eight people globally experiencing a mental disorder, with depression, anxiety, bipolar disorder, post-traumatic stress disorder (PTSD), and schizophrenia being the most prevalent. A more recent study from the National Institute of Mental Health (NIMH) shows one in five adults in America struggling with mental health [2]. Figure 1(a) illustrates the distribution of these conditions across existing clinical mental health datasets. Most datasets focus on schizophrenia, PTSD, and depression, with relatively fewer addressing anxiety and bipolar disorder. Given the distinct symptom profiles of each disorder, condition-specific features, diagnostic criteria, and validated clinical questionnaires are essential. Table 3 summarizes the primary psychological questionnaires used for assessing the severity of these conditions.

Depression. Depression is the most frequently represented disorder in clinical datasets, particularly those incorporating multimodal data. Many datasets focus on identifying depressive symptoms through text, speech, and facial expressions[77–80,86]. These datasets typically include severity ratings based on standardized scales such as the Hamilton Depression Rating Scale (HDRS)[116], Patient Health Questionnaire. Depression (PHQ-9)[117], Beck Depression Inventory (BDI-II)[118], Self rating depression scale (SDS)[119], Quick Inventory of Depressive Symptomatology Self Report (QIDS-SR)[120], Center for Epidemiologic Studies of Depression Scale (CES-D)[121], Hospital Anxiety and Depression Scale Depression sub scale (HADS-D)[122] and Inventory of Depressive Symptomatology – Self Report (IDS-SR)[123] making them valuable for both classification and regression tasks. Some other datasets try to diagnose depression through EEG[98,99,101–108] or EEG combined with audio and video[87,100].

Anxiety. Clinical datasets targeting anxiety are less common and often smaller in size. Most rely on video recordings to capture visual cues like facial tension, restlessness, and gaze aversion[44–46]. These datasets are typically collected in clinical interview or exposure settings, and may include self-reported or clinician-administered scores from instruments such as Liebowitz Social Anxiety Scale (LSAS)[124], State-Trait Anxiety Inventory (STAI)[125], Generalized Anxiety Disorder (GAD-7)[126], Hospital Anxiety and Depression Scale Anxiety sub scale (HADS-A)[122], Social Interaction Anxiety Scale (SIAS) and Social Phobia Scale (SPS)[127].

Bipolar Disorder. Bipolar disorder is underrepresented in publicly-available clinical datasets. Existing resources often include longitudinal data capturing both depressive and manic episodes using audio, text, and video modalities. Speech datasets focus on prosodic rhythm and vocal markers[32], while text-based corpora capture linguistic markers of mood shifts[66,67]. A smaller
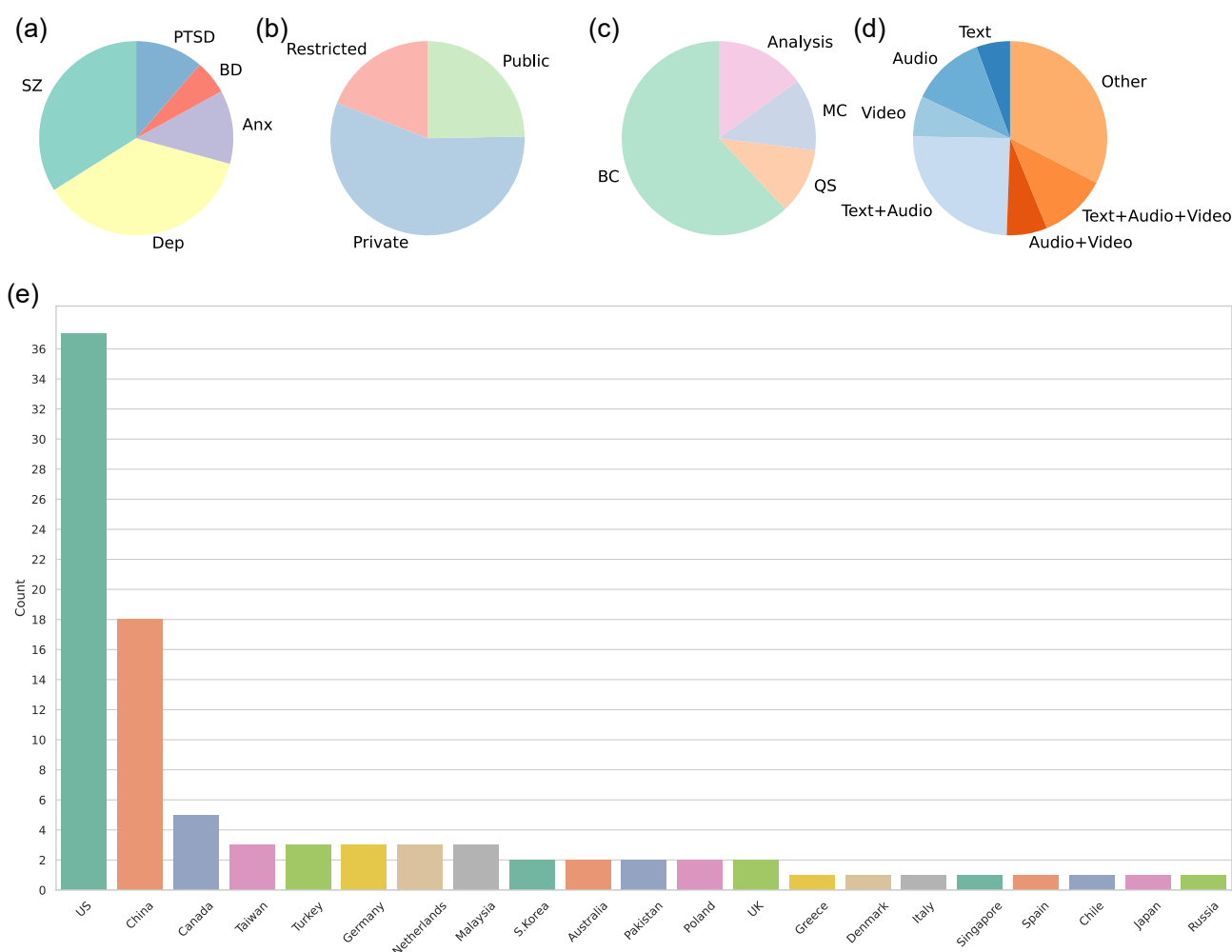
Figure 1. An overview of the distribution of clinical mental health datasets reviewed in this work. **(a)** Distribution by mental health disorder, including schizophrenia (SZ), depression (Dep), anxiety (Anx), bipolar disorder (BD), and post-traumatic stress disorder (PTSD). **(b)** Distribution by dataset accessibility level: public, restricted, and private. **(c)** Distribution by task type: binary classification (BC), multi-class classification (MC), questionnaire score prediction (QS), and analysis tasks. **(d)** Distribution by modality and modality combinations: text, audio, video, text+audio, audio+video, text+audio+video, and others (e.g., EEG, MRI). **(e)** Distribution showing the number of datasets collected in each country.

subset includes annotated video data reflecting affective expression[48]. The Young Mania Rating Scale (YMRS)[128] is typically used for severity labeling when available.

Post-Traumatic Stress Disorder (PTSD). PTSD is moderately represented in clinical datasets, with a majority focusing on speech-based markers of trauma-related stress. Audio datasets commonly include features such as vocal tension, pitch variability, and temporal disfluencies[33–38,41]. Multimodal datasets incorporating text and video also exist, particularly in clinical interview settings[77,79,80]. Some datasets also use EEG for diagnosing PTSD[109,110]. Severity annotations often rely on the Subjective Unit of Distress (SUD)[129], PTSD Checklist for DSM-5 (PCL-5)[130] or its civilian variant, the PTSD Checklist – Civilian Version (PCL-C)[131].

Schizophrenia. Schizophrenia is among the most comprehensively represented disorders in clinical datasets. Many corpora focus on speech disorganization, including spontaneous speech, narrative recall, or interview responses[27,52–66]. Several also incorporate visual modalities such as head movement[43,85] and facial affect[84]. In addition, datasets using EEG[90–92,111–114] or MRI [89,93–97,115] data provide access to neurophysiological correlates of psychosis. Commonly used assessments include Scale for the assessment of Positive symptoms (SAPS)[132], Scale for the assessment of Negative symptoms (SANS)[133], Negative Symptom Assessment (NSA-16)[134], Positive and Negative Syndrome scale (PANSS)[135], and Assessment of Thought, Language

and Communication (TLC)[136], allowing for fine-grained labeling of symptom dimensions.

Data Accessibility

Clinical mental health datasets vary widely in their accessibility, directly shaping their utility for research and the risks they pose to patient privacy. We categorize accessibility into three levels: public, restricted, and private, each occupying a different point along the privacy–usability trade-off. While public datasets maximize accessibility, they pose the greatest risk of sensitive information exposure. Private datasets offer the strongest privacy safeguards but are often inaccessible to the broader research community. Restricted datasets represent a compromise, supporting research use under controlled conditions. The distribution of dataset accessibility in our review is shown in Figure 1(b).

Public Datasets. Public datasets are freely available for use in training, evaluation, and analysis. While many social media–based mental health datasets are openly released, clinical datasets are rarely public due to the sensitive nature of the data. Even anonymized text, audio, or video samples carry re-identification risks, especially when multiple modalities are combined. Multimodal datasets are particularly vulnerable to cross-modal privacy breaches, where patterns across data types can inadvertently reveal personal identities. Public release also requires explicit participant consent, which is difficult to obtain given understandable concerns about data misuse. As a result, publicly available clinical mental health datasets, especially those with multimodal data, remain scarce (see Table 2). Only some select multimodal datasets from China like EATD[49], MMPsy[51] and CMDC[78] are made public after rigorous anonymization. Most of the publicly available datasets contain MRI[89,94–97] and EEG[90–92,101,102,104,108,112–114] data. A further concern is the misuse of these datasets by actors lacking clinical or ethical oversight, such as unregulated commercial applications of mental health AI.

Restricted Datasets. Restricted datasets are available to qualified researchers through formal application processes that often include institutional review, ethical approval, and a clearly defined research purpose. This model enables controlled sharing that balances privacy protection with research utility. Although access procedures can delay use, they help ensure that data are not repurposed for non-clinical or commercial ends. Restricted datasets are particularly common for sensitive modalities such as audio and video. For example, the AViD Corpus[72], Pittsburgh dataset[73] and datasets from Lin et al.[75] and Guo et al.[76] contain audio and video modalities and are restricted. Similarly datasets containing text, audio and video modalities like E-DAIC[77], MEDIC[81], BDS[82] and dataset from Zhang et al.[85] have restricted access. While most datasets with EEG and MRI data are made public, some of them are kept restricted[93,105,107].

Private Datasets. Private datasets are held exclusively by the institutions that collected them and are typically not shared beyond internal researchers or a small number of vetted collaborators. They are stored in secure environments, often under strict governance protocols. This model offers the highest level of privacy and is more likely to receive participant consent, but it limits transparency, reproducibility, and the potential for secondary research. As shown in Figure 1, private datasets represent the majority of clinical mental health data resources today. Among text-only datasets LabWriting[27] is kept private since it contains information about personal lives of participants. Most Audio-only[32,34–38] and Video-only[43,45–48] datasets are kept private since they contain biometric and identifiable features like voice and face of a person respectively. Similarly multimodal datasets[50,53–65,67,71,74,79,80,83,84] including audio or video data is also kept private for the same reason. A few datasets with EEG[98,99,103,106,109–111] are also kept private due to the risk of leaking private information.

Task Types

Clinical mental health datasets are typically developed with specific downstream tasks in mind, reflecting key stages of diagnosis, symptom assessment, and treatment planning. Based on our analysis, these tasks fall into four primary categories: binary classification, multi-class classification, questionnaire score prediction, and therapeutic response generation. The task distribution across reviewed datasets is shown in Figure 1(c). Some datasets are also used for analysing the features that are present in patients with certain mental health issues and correlation of these features with severity prediction[30,39,40,45,57,59,62,64,65,68,88,91,95,115].

Binary Classification. This is the most prevalent task across clinical datasets, where the goal is to determine whether an individual has a specific mental health condition, typically producing a yes/no output. These datasets are commonly used to train AI models that classify individuals as either affected by a particular disorder or as healthy controls. Most datasets try to classify depression[28,31,42,46,47,49–51,70,71,74–76,78–80,86,90,98–108] and schizophrenia[27,28,43,52–56,58–61,63,69,85,89,90,94–97,111–114]. Other datasets also classify anxiety[29,31,42,46,51,71,75,86,90], bipolar disorder[31,48,82], and PTSD[34–38,79,80,109,110]. Many datasets support binary classification either directly or by design – particularly those comparing affected individuals with healthy controls, making it well-suited for early diagnostic research. In addition to classification, these datasets are often used to analyze linguistic, acoustic, and visual markers associated with different disorders. However, binary labels offer limited clinical nuance and do not reflect symptom severity or heterogeneity.
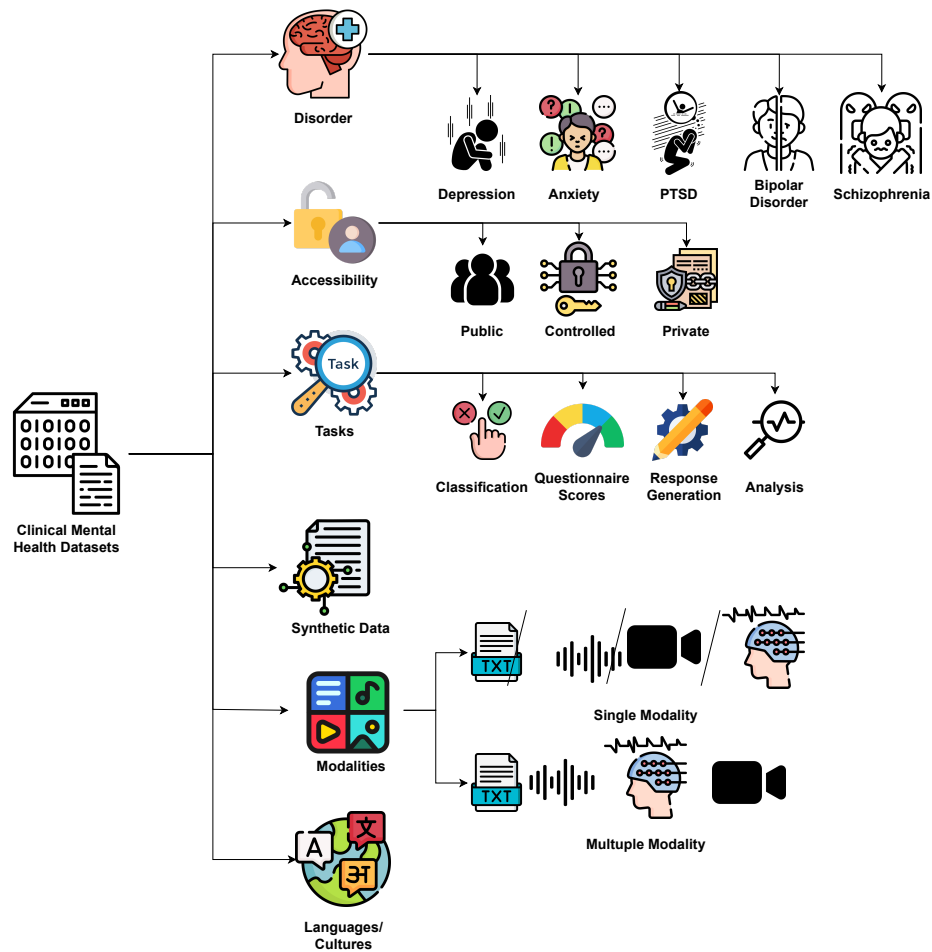
Figure 2. Schematic overview of this review. This diagram illustrates the key dimensions covered in our review, including mental health disorder categories, dataset accessibility levels, task types, synthetic data sources, data modalities, and cultural or linguistic representation. The taxonomy offers a structured framework for understanding the landscape of clinical mental health datasets and serves as a guide for navigating available resources in this domain.

Multi-Class Classification. A smaller set of datasets support multi-class classification, which distinguishes between different severity levels, such as mild, moderate, or severe, of a given disorder[32,33,44,66,67,73,82,84,92]. This granularity is useful for tracking disease progression or informing treatment decisions. These datasets typically include clinician-assigned severity ratings or annotated questionnaire scores that allow stratification into multiple classes. Although more clinically relevant than binary classification, this task still provides a coarse approximation of symptom complexity. Some other multi-class classification datasets also classify between several disorders. PRIORI[32] classifies participants into suffering from depression or bipolar disorder or as healthy controls. Similarly, Aich et al.[66] classify between schizophrenia, bipolar disorder and healthy controls. UCLA[93] uses MRI to classify participants into schizophrenia, bipolar disorder and healthy control classes.

Questionnaire Score Prediction. Several datasets aim to predict total or item-wise scores on standardized psychological questionnaires, which are widely used in clinical settings for diagnosis and monitoring. These datasets provide rich supervision signals by aligning model outputs with instruments used in real-world practice. Table 3 provides a summary of the different questionnaires used in the datasets reviewed in this paper. For depression, datasets commonly include annotations for the Hamilton Depression Rating Scale (HDRS)[32,40,47,67,73,103] and the Patient Health Questionnaire (PHQ-9)[39,51,68,71,75–79,87,98], with others using BDI-II[46,70,72,100–102,108], SDS[30,49,105], QIDS-SR[74], CES-D[80], HADS-D[101] or IDS-SR[39]. Social anxiety datasets incorporate LSAS[44], STAI[46], GAD-7[68,71,75,98], HADS-A or SIAS and SPS[42], while bipolar disorder datasets use YMRS[32,67,82]. PTSD-related datasets include labels based on SUD[41], PCL-5[38,80] or PCL-C[77,79]. Schizophrenia datasets support prediction of PANSS scores[43,60,61,83,111], SAPS[64], SANS[64], NSA-16[63], and TLC[64,83]. These datasets are typically multimodal and annotated by clinical professionals, making them highly valuable for translational AI research. Questionnaire prediction tasks also provide more interpretable and clinically actionable outputs than raw classification.

Questionnaire	Disorder	Completed By	Items	Range
HDRS[116]	Depression	Clinician	17	0-52
PHQ-9[117]	Depression	Patient	9	0-27
BDI-II[118]	Depression	Patient	21	0-63
SDS[119]	Depression	Patient	20	20-80
QIDS-SR[120]	Depression	Patient	9	0-27
CES-D[121]	Depression	Patient	20	0-60
HADS-D[122]	Depression	Patient	7	0-21
IDS-SR[123]	Depression	Patient	30	0-84
LSAS[124]	Anxiety	Clinician/ Patient	24	0-144
STAI[125]	Anxiety	Patient	40	40-160
GAD-7[126]	Anxiety	Patient	7	0-21
HADS-A[122]	Anxiety	Patient	7	0-21
SIAS[127]	Anxiety	Patient	20	0-80
SPS[127]	Anxiety	Patient	20	0-80
YMRS[128]	Bipolar Disorder	Clinician	11	0-60
SUD[129]	PTSD	Patient	1	0-10
PCL-5[130]	PTSD	Patient	20	0-80
PCL-C[131]	PTSD	Patient	17	17-85
SAPS[132]	Schizophrenia	Clinician	34	0-170
SANS[133]	Schizophrenia	Clinician	25	0-125
NSA-16[134]	Schizophrenia	Clinician	16	16-96
PANSS[135]	Schizophrenia	Clinician	30	30-210
TLC[136]	Schizophrenia	Clinician	18	0-72

Table 3. Overview of clinical questionnaires for assessing mental health disorders, including target disorder, respondent type (clinician or self-report), number of items, and score range.

Therapeutic Response Generation. An emerging frontier in computational mental health is therapeutic response generation, where models are trained to produce contextually appropriate and empathetic responses in therapy-like settings. Among the datasets surveyed, only MEDIC[81] offers the necessary modality and structure to support this task. Although originally developed for empathy detection, MEDIC includes audio-visual recordings of counseling sessions, making it a valuable resource for response generation research. In contrast, existing datasets such as[77,78] primarily consist of clinical interviews aimed at diagnostic or severity assessment, lacking the dynamic and interactive nature of real therapy sessions. This highlights a critical gap: the scarcity of datasets capturing authentic therapeutic interactions. To advance the development of AI systems that can meaningfully augment therapy or assist clinicians, there is a pressing need for the collection and dissemination of high-quality, recorded therapy session datasets.

Synthetic Datasets

Synthetic data generation has emerged as a promising strategy to address longstanding challenges in mental health AI research, including data scarcity, privacy concerns, and limited demographic diversity. By simulating clinically realistic data, these methods enable the creation of training and evaluation resources without exposing sensitive patient information. We show a summary of existing synthetic datasets in Table 4.

A common approach uses large language models (LLMs), such as ChatGPT[137], in zero- or few-shot settings to generate synthetic conversations derived from real clinical data. For example, Wu et al.[138] extend the E-DAIC PTSD dataset[77] using ChatGPT-generated interview transcripts. Other efforts, such as Psych8k[139] and MentalChat16k[140], produce synthetic question-answer (QA) pairs by prompting LLMs with anonymized interview data. MDD-5k[141] scales this further by generating thousands of synthetic diagnostic dialogues using a neuro-symbolic multi-agent LLM framework trained on real interactions. D4[142] and Mousavi et al.[143] also generate simulated clinician–patient conversations using portraits or textual profiles of real individuals. While these methods reduce direct privacy risks, they still rely on real-world data and may be vulnerable to information leakage.

To address this, recent work avoids real data altogether. Datasets such as Thousand Voices of Trauma[144] and PSYCON[145]

Dataset	Language	Modality	Size	Data types
PTSD Synth[138]	English	Text	3281(ZS)/392(FS)	Multi-turn Dialogue
Psych8k[139]	English	Text	8,187	QA Pairs
MentalChat16k[140]	English	Text	6,338	QA Pairs
CpsyCounD[153]	Chinese	Text	3,134	Multi-turn Dialogue
HealMe[146]	English	Text	1,300	Multi-turn Dialogue
SMILE[148]	Chinese	Text	55,165	Multi-turn Dialogue
SoulChat[149]	Chinese	Text	2,300,248	Multi-turn Dialogue
CBT QA[151]	Chinese	Text	22,327	QA Pairs
CACTUS[152]	English	Text	31,577	Multi-turn Dialogue
Thousand Voices of Trauma[144]	English	Text	500	Multi-turn Dialogue
PSYCON[145]	English	Text	1020	Multi-turn Dialogue
D^4[142]	Chinese	Text	1,339	Multi-turn Dialogue
Mousavi et al.[143]	English	Text	800	Multi-turn Dialogue
MDD-5k[141]	Chinese	Text	5,000	Multi-turn Dialogue
M2CoSC[154]	English	Text, Image	429	Multi-turn Dialogue
MIRROR[155]	English	Text, Image	3,073	Multi-turn Dialogue

Table 4. Overview of synthetic mental health datasets, detailing language, modalities, dataset size (ZS: zero-shot, FS: few-shot), and data type (multi-turn dialogues or single-turn QA pairs).

construct diverse synthetic profiles of clients and therapists – varying in age, gender, behavior, and disorder, and generate counseling dialogues across a wide spectrum of conditions, including depression, PTSD, schizophrenia, and bipolar disorder. Similarly, HealMe[146] simulates clients from the PATTERNREFRAME dataset[147], with dialogues generated between a ChatGPT client and therapist. However, HealMe follows a rigid, fixed-step counseling strategy, limiting conversational diversity. SMILE[148] and SoulChat[149] aim to increase variability by generating multi-turn dialogues from QA pairs in PsyQA[150] or crowdsourced sources, though the lack of clinical grounding in these responses can limit psychological fidelity.

To address this, CBT-LLM[151] enriches QA interactions using prompts informed by cognitive behavioral therapy (CBT), while CACTUS[152] builds on PATTERNREFRAME with more detailed client profiles and dynamic CBT-based conversations, improving both grounding and variability. CPsyCoun[153] follows a similar pattern, using forum-sourced memos as seeds for synthetic counseling dialogues, though expert oversight remains limited.

While most synthetic datasets focus on text, some begin to address multimodal aspects of mental health. M2CoSC[154] augments textual dialogues with static client images generated via GPT-4V. However, the visual content remains unchanged during interaction, and dialogue generation follows fixed strategies. In contrast, MIRROR[155] simulates dynamic facial expressions in response to conversational cues and integrates a CBT-driven dialogue model, representing a significant step toward psychologically grounded, multimodal simulation.

Despite these advances, there remains a substantial gap in synthetic generation for speech and video modalities. Non-verbal signals, such as tone, pause patterns, facial expressions, and body posture, are critical for clinical judgment and therapeutic engagement. Future research should prioritize multimodal synthetic data generation that captures these dimensions to more accurately model real-world therapeutic interactions.

Dataset Modalities

Single-Modality Datasets. While multimodal learning offers a richer representation of mental health cues, single-modality approaches remain widely used due to their lower cost, simpler deployment, and reduced privacy risks. In some cases, disorders can be reliably detected from just one modality, making unimodal systems a viable alternative.

Text-only datasets are relatively rare and often limited in diagnostic utility due to the context-dependent nature of language. For instance, WorryWords[29] links over 44,000 words to anxiety associations, but lacks conversational context. Other studies leverage clinical notes to detect disorders such as schizophrenia and depression[28], or analyze written language in patients with schizophrenia[27]. Hou et al.[30] analyze the correlation between depression and reading habits among university students. They analyze the text of the book read by a person and try to predict if they suffer from depression. The College Health Surveillance Dataset (CHSN)[31] contains a large amount of data from EHRs of 1 Million university students over 6 Million visits covering mental disorders like depression, anxiety and bipolar disorder. The Ex-Ray[69] dataset also contains psychological reports for clustering patients. Several others explore text classification of schizophrenia[52–57], though many also collect audio to analyze

speech patterns.

Audio data provides valuable diagnostic signals, especially for PTSD through features like prosody, spectral characteristics, and vocal tract dynamics[33–38,41,42], often extracted using tools like OpenSMILE[35,38] or Wav2Vec. On the other hand, PRIORI[32] collects Smart phone conversations to predict depression and bipolar disorder through extracted rhythm features. Similarly, RADAR-MDD[39] contains smartphone speech recordings of depressed patients from the U.K., Spain and Netherlands to find depression markers from a set of 28 speech features. Similarly, Chang et al.[40] ask psychologists to analyze speech cues in smartphone recordings to understand speech features that are important for diagnosing mental illness. Despite its diagnostic value, audio data collection presents challenges, including equipment requirements and acoustic control.

Video introduces the greatest privacy risks due to identifiable facial features, but is particularly effective in assessing disorders involving non-verbal cues. Gaze aversion and facial expressions are key indicators of social anxiety[44,45], depression[46,47], bipolar disorder[48], and schizophrenia[43]. Abbas et al.[43], for instance, use head movement rates from smartphone cameras to detect schizophrenia.

Physiological modalities like EEG and MRI also show their utility in diagnosing various mental disorders. Datasets with EEG are mainly used to diagnose depression[90,98,99,101–108] and schizophrenia[90–92,111–114] but can also help in anxiety[90] and PTSD[109,110] diagnosis. Meanwhile, MRI data is useful for schizophrenia classification[89,93–97,115].

Multimodal Datasets. Many psychiatric symptoms manifest across multiple behavioral channels, making multimodal analysis essential for accurate diagnosis. For depression, relevant cues span speech semantics[156], prosody[157], and facial expressions[158]. Evidence from prior works further supports the efficacy of multimodal approaches. The AVEC-2017[159] and AVEC-2019[160] challenges demonstrated that models incorporating text, audio, and visual information outperform unimodal baselines in depression severity prediction. Similarly, the AVEC-2018 challenge[161] showed improved performance in bipolar disorder diagnosis through multimodal fusion.

The most common combination is text and audio, seen in 22 clinical datasets targeting conditions like depression, anxiety, schizophrenia, bipolar disorder, and PTSD[49,50,58–63]. These datasets typically comprise interview transcripts and audio recordings from clinical or semi-structured interactions. Some, like MMPsy[51] and Jeong et al.[64], collect responses to structured prompts. Following a similar approach, Ex-Ray[69] records participants speaking and writing a passage about a particular topic with specific target words. Other datasets embed diagnostic tasks, including Theory of Mind assessments[65], social role-play[66], and narrative responses to TAT images[67]. Other tasks include phoneme task, testing phonemic and semantic fluency, picture description and prompted narrating[68]. Hayati et al.[70] collect and transcribe semi-structured interviews for depression detection in different Malay dialects.

Ten public datasets include all three modalities – text, audio, and video, while seven include audio and video. To protect privacy, raw video is often withheld, and only extracted features like facial landmarks or action units are shared[71,73–75,77–81,83,84]. Most multimodal datasets are collected from interviews with trained professionals or virtual agents, though several involve task-based paradigms. AViD[72], for instance, uses PowerPoint prompts to guide participants through reading, storytelling, and TAT-inspired tasks. Guo et al.[76] balance emotional content across stimuli, while BDS[82] and Zhang et al.[85] incorporate structured speaking tasks such as counting and passage reading. In another study, Tao et al.[86] collect video recordings of patients talking with a ChatGPT-based virtual character in real time mental health conversation.

Beyond behavioral data, some datasets integrate physiological or neuroimaging signals. MODMA[87] includes paired EEG and audio for depression detection, while Zhu et al.[100] use audio, video and EEG for depression detection. VerBIO[88] uses EEG and audio to analyze anxiety disorder.

Language and Cultural Diversity in Datasets

Culture profoundly shapes how individuals experience, express, and interpret psychological distress[162]. Across cultural and linguistic contexts, people employ different metaphors for mental illness[163], and their symptoms may manifest in divergent physical or behavioral forms[164]. Ignoring these cultural nuances risks misdiagnosis, inappropriate treatment, and reduced efficacy of AI models in mental health care.

For instance, individuals with depression in South Asian cultures often attribute their distress to supernatural or moral causes, whereas White British individuals more frequently cite biological explanations[165]. Cultural interpretations of anxiety also vary: Americans commonly report fear of heart attacks, while Cambodians describe sensations such as "limb blockage", tightness or soreness in arms and legs[166]. Emotional expression is likewise shaped by cultural norms. In individualistic societies such as the United States and United Kingdom, anxiety often stems from guilt or self-blame, whereas in East and Southeast Asian cultures, it is more commonly rooted in embarrassment[166].

Cultural differences also influence symptomatology in psychotic and mood disorders. Although auditory hallucinations are a near-universal feature of schizophrenia, visual hallucinations are more frequently reported in Africa, Asia, and the Caribbean than in Europe or North America[167], and the content of hallucinations varies by region. In bipolar disorder, Tunisian patients
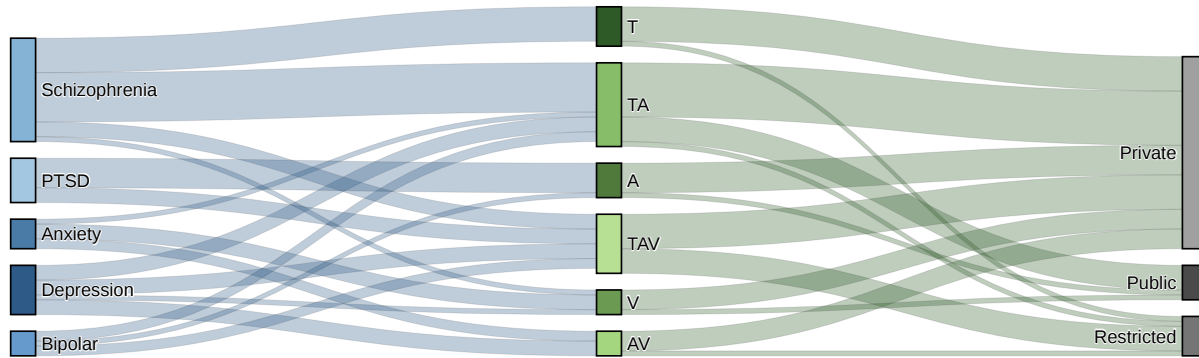
Figure 3. Sankey diagram depicting the distribution of dataset modalities across mental health disorders and their accessibility levels. The figure visualizes the types and combinations of modalities present in clinical mental health datasets for various disorders, along with their accessibility status (public, restricted, or private), highlighting gaps in data availability and openness across different disorders and formats.

often present initially with manic symptoms, while French patients more commonly exhibit depressive episodes[168]. For PTSD, symptoms in populations such as the Kalahari Bushmen or Vietnamese communities often diverge from Western norms like emotional numbing or avoidance, resulting in underdiagnosis[169].

Even LLMs, such as ChatGPT, have shown limitations in multicultural therapeutic contexts, tending by default to culturally neutral or Western-centric advice, rather than offering context-specific responses[170]. Zahran et al.[171] test eight different LLMs including multi-lingual LLMs in the Arabic context and find that current LLMs are inadequate in culture specific contexts. Hayati et al.[70] demonstrate that even within the same language (Malay), the performance of LLMs in depression detection can vary significantly across regional dialects, underscoring the need for cultural and linguistic sensitivity. These examples highlight the necessity for culturally competent AI systems, trained on diverse datasets that reflect the lived experiences of varied populations.

Yet, as illustrated in Figure 1(e), most clinical mental health datasets remain concentrated in English-speaking (e.g., U.S., U.K., Canada, Australia) and Chinese-speaking (e.g., China, Taiwan, Taipei) settings. Some additional representation exists from Germany (German), Turkey (Turkish), Pakistan (Urdu), Greece (Greek), Poland (Polish), Denmark (Danish), Italy (Italian), the Netherlands (Dutch), and Singapore (English or Mandarin), Spain (Spanish), Malaysia, Korea, Japan, Chile, Russia. However, these are typically limited to one or two datasets per region. Large global regions, including South and Southeast Asia, the Middle East, Africa, Central and South America, are still largely missing. Similarly, widely spoken languages such as Hindi, Arabic, Bengali, Portuguese, and major African languages are either severely underrepresented or absent entirely.

This lack of cultural and linguistic diversity inherently limits the generalizability of current AI models. To ensure equitable, effective, and globally relevant mental health technologies, there is an urgent need to collect and curate datasets from underrepresented cultures and languages.

Challenges

Small Datasets. Despite growing interest in clinical mental health datasets, most remain small, often under 200 participants (Table 2), due to high collection costs, logistical challenges, and the ethical and legal complexities of handling sensitive data. Such limited scale hinders robust AI training, leading to overfitting and poor generalization. While methods like transfer learning and regularization offer partial relief, they cannot replace scale, highlighting the persistent tension between protecting privacy and enabling research. To address these limitations, we propose federated learning and synthetic data generation. Federated learning[172] enables model training across decentralized, institution-specific datasets (Table 2) without sharing raw data, thus preserving privacy while increasing effective scale. Synthetic data, grounded in the statistical and clinical properties of real datasets, can further augment training, but must capture the nuanced complexities of mental health rather than superficial correlations.

Limited Diversity. Current datasets often lack diversity in geography, language, culture, and clinical setting, being mostly collected in English- or Chinese-speaking countries and within single institutions. This narrow scope limits model generalizability and risks cultural and demographic bias. Data sharing across regions is constrained by privacy regulations, but federated learning can enable training on distributed datasets without moving sensitive data. Synthetic data generation, using diverse client and therapist profiles[141,142,144,145,152], can further reflect varied linguistic, cultural, and demographic contexts, capturing complex contextual–psychological interactions to support more inclusive mental health technologies.

Lack of Standardization. Wide variation in protocols, annotations, modalities, and task definitions hampers training and makes evaluation inconsistent. Without shared schemas, collaboration remains fragile. Addressing this requires unified annotation, metadata, and evaluation standards, alongside developing AI systems flexible enough to handle missing modalities and heterogeneous inputs via multi-task and modular architectures.

Synthetic Data: Missing Modalities, Languages, and Depth. While synthetic data can address privacy concerns and expand datasets, most current resources are single-turn English or Chinese textual dialogues, omitting the multimodal, multilingual, and longitudinal nature of real psychotherapy. They rarely model temporal dynamics, underrepresented disorders, or diverse cultural contexts. Clinically useful datasets should incorporate multi-session, multimodal (audio, visual, physiological) data, multiple languages, and diverse therapist–client profiles, grounded in psychological theory and enabled by multimodal LLMs[173]. Achieving this requires sustained collaboration among clinicians, computational scientists, and cultural experts. Without such interdisciplinary effort, synthetic data will remain an elegant but shallow solution to a much deeper problem.

Federated Learning. Federated learning (FL) can address the scale and diversity limits of mental health datasets but remains vulnerable to gradient leakage attacks[174]. Combining FL with local differential privacy (LDP)[174] adds protection by locally injecting noise, yet often degrades performance on small clinical datasets[175]. Advancing secure, effective deployment requires novel LDP-FL methods that better balance privacy and utility, validated in realistic, resource-constrained settings.

Privacy. As shown in Table 2 and Figure 1 (b), most clinical mental health datasets remain private due to a lack of robust, scalable privacy mechanisms. Traditional anonymization, often single-modality and without formal guarantees[176], is inadequate, especially for multimodal data, where one modality can reveal sensitive attributes from another. Broader access will require multimodal anonymization methods with theoretical guarantees to prevent cross-modal leakage and ensure equitable protection across populations.

Evaluation Benchmark. The value of a mental health dataset is ultimately determined by the performance of models trained on it in clinically relevant tasks. Existing efforts like CounselBench[177] are useful but focus mainly on single-turn counseling. There is a need for broader benchmarks covering tasks such as disorder classification, symptom severity prediction, and multi-turn therapy assessment, while reflecting cultural, linguistic, and demographic diversity. Developing such resources will require close collaboration among clinicians, computational scientists, and cultural experts to ensure both clinical validity and fairness.

Future Directions

Advancing mental health research through AI requires a shift toward standardized, ethically sound, and privacy-preserving data collection practices. Future initiatives must adopt clear, field-wide data collection principles that not only ensure compliance with ethical and legal standards but also facilitate effective and secure use of the data for AI development. We highlight three complementary strategies for responsibly leveraging collected mental health data: (i) Federated learning with local differential privacy (LDP-FL), where data collected across geographically distributed sites, each representing diverse cultural and demographic populations remains stored locally, with model training conducted in a privacy-preserving, decentralized manner; (ii) Multimodal synthetic data generation grounded in psychological theory, using real-world clinical data as reference points and drawing upon diverse therapist and client profiles, to supplement existing datasets and enhance robustness of trained models; (iii) Public release of anonymized multimodal datasets, made possible through advanced anonymization techniques with theoretical privacy guarantees to minimize the risk of re-identification while retaining research utility.

Data Collection. To build large, diverse, and clinically valuable datasets, data collection must follow standardized protocols with prior approval from institutional ethics committees and informed consent from participants, clearly detailing intended data use, storage procedures, and privacy safeguards. Collection should capture the multimodal nature of therapy sessions through high-quality audio-visual recordings of patient–therapist interactions, with video documenting facial expressions, gaze, head movements, and body posture of patients, and audio recorded via separate microphones to allow accurate speaker diarization. Text transcripts should be created manually by authorized researchers or via secure, local automatic speech recognition (ASR) systems to avoid external data transmission. Additionally, physiological measures such as EEG and MRI should be included

where relevant, especially for conditions like schizophrenia and depression, where they have shown diagnostic value. Adhering to these guidelines is critical for producing clinically meaningful, ethical, and reproducible datasets that support inclusive AI research in mental health.

Data Utilization. To responsibly leverage mental health data while protecting privacy, we identify three complementary pathways. First, federated learning combined with local differential privacy (LDP-FL) provides a privacy-preserving framework for using distributed datasets across institutions and regions. This approach enables integration of culturally and linguistically diverse data by keeping it locally stored. However, current LDP-FL methods often underperform on small clinical datasets[175], highlighting the need for improved techniques balancing privacy and utility. Alternatively, data collected within a single institution can be augmented via synthetic data generation. Multimodal large language models[173] can create high-quality synthetic datasets guided by structured client and therapist profiles reflecting diverse cultural, linguistic, and demographic attributes. These synthetic interactions, should be grounded in psychological frameworks such as cognitive behavioral therapy (CBT)[152], and enhanced with few-shot examples from real, ethically collected data, improving realism and clinical fidelity. Crucially, synthetic data is free of real patient identifiers and can be publicly shared to promote transparency and accelerate research while preserving privacy. Finally, public release of real-world mental health data would greatly improve research utility and remains an essential long-term goal but requires robust multimodal anonymization methods. Existing anonymization techniques are usually modality-specific and fail to address cross-modal privacy risks (e.g., audio revealing text-based sensitive information). Future frameworks must provide theoretical privacy guarantees and handle complex multimodal data. Once achieved, safe public release of diverse datasets can enable large-scale aggregation and foster more generalizable, inclusive mental health AI systems.

Author Contributions

T.C. and I.G. conceptualized the idea; A.M. and P.K.A. developed the work; A.M., P.K.A., T.C. and H.A. wrote the manuscript; T.C., I.G., A.M., P.K.A. and H.A. revised the manuscript.

Funding Information

Competing Interests

The authors declare no competing interests.

Additional Information

Materials & Correspondence should be emailed to Tanmoy Chakraborty (`tanchak@iitd.ac.in`).

References

1. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation) (text with eea relevance) (2016).

2. Act, A. Health insurance portability and accountability act of 1996. *Public law* **104**, 191 (1996).

3. Shen, J. H. & Rudzicz, F. Detecting anxiety through reddit. In Hollingshead, K., Ireland, M. & Loveys, K. (eds.) *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality, CLPsych@ACL 2017, Vancouver, Canada, August 3, 2017*, 58–65, DOI: 10.18653/V1/W17-3107 (Association for Computational Linguistics, 2017).

4. Mitchell, M., Hollingshead, K. & Coppersmith, G. Quantifying the language of schizophrenia in social media. In Mitchell, M., Coppersmith, G. & Hollingshead, K. (eds.) *Proceedings of the 2nd Workshop on Computational Linguistics and*

Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsych@NAACL-HLT 2015, June 5, 2015, Denver, Colorado, USA, 11–20, DOI: 10.3115/V1/W15-1202 (The Association for Computational Linguistics, 2015).

5. Yoon, J., Kang, C., Kim, S. & Han, J. D-vlog: Multimodal vlog dataset for depression detection. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, 12226–12234, DOI: 10.1609/AAAI.V36I11.21483 (AAAI Press, 2022).

6. Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H. & Eichstaedt, J. C. Detecting depression and mental illness on social media: an integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017).

7. Dehbozorgi, R. *et al.* The application of artificial intelligence in the field of mental health: a systematic review. *BMC psychiatry* **25**, 132 (2025).

8. Shatte, A. B. R., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: a scoping review of methods and applications. *Psychol. Medicine* **49**, 1426–1448, DOI: 10.1017/S0033291719000151 (2019).

9. Iyortsuun, N. K., Kim, S.-H., Jhon, M., Yang, H.-J. & Pant, S. A review of machine learning and deep learning approaches on mental health diagnosis. *Healthcare* **11**, DOI: 10.3390/healthcare11030285 (2023).

10. Scherbakov, D. A., Hubig, N. C., Lenert, L. A., Alekseyenko, A. V. & Obeid, J. S. Natural language processing and social determinants of health in mental health research: Ai-assisted scoping review. *JMIR Ment Heal.* **12**, e67192, DOI: 10.2196/67192 (2025).

11. Demszky, D. *et al.* Using large language models in psychology. *Nat. Rev. Psychol.* **2**, 688–701 (2023).

12. Graham, S. *et al.* Artificial intelligence for mental health and mental illnesses: an overview. *Curr. psychiatry reports* **21**, 1–18 (2019).

13. Dehbozorgi, R. *et al.* The application of artificial intelligence in the field of mental health: a systematic review. *BMC psychiatry* **25**, 132 (2025).

14. Shatte, A. B., Hutchinson, D. M. & Teague, S. J. Machine learning in mental health: a scoping review of methods and applications. *Psychol. medicine* **49**, 1426–1448 (2019).

15. Su, C., Xu, Z., Pathak, J. & Wang, F. Deep learning in mental health outcome research: a scoping review. *Transl. psychiatry* **10**, 116 (2020).

16. Iyortsuun, N. K., Kim, S.-H., Jhon, M., Yang, H.-J. & Pant, S. A review of machine learning and deep learning approaches on mental health diagnosis. *Healthcare* **11**, DOI: 10.3390/healthcare11030285 (2023).

17. Scherbakov, D. A., Hubig, N. C., Lenert, L. A., Alekseyenko, A. V. & Obeid, J. S. Natural language processing and social determinants of health in mental health research: Ai-assisted scoping review. *JMIR Mental Heal.* **12**, e67192 (2025).

18. Laricheva, M., Liu, Y., Shi, E. & Wu, A. Scoping review on natural language processing applications in counselling and psychotherapy. *Br. J. Psychol.* (2024).

19. Malgaroli, M., Hull, T. D., Zech, J. M. & Althoff, T. Natural language processing for mental health interventions: a systematic review and research framework. *Transl. Psychiatry* **13**, 309 (2023).

20. Garg, M. Mental health analysis in social media posts: a survey. *Arch. Comput. Methods Eng.* **30**, 1819–1842 (2023).

21. Ahmed, A. *et al.* Overview of the role of big data in mental health: A scoping review. *Comput. Methods Programs Biomed. Updat.* **2**, 100076, DOI: https://doi.org/10.1016/j.cmpbup.2022.100076 (2022).

22. Thieme, A., Belgrave, D. & Doherty, G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. *ACM Trans. Comput. Hum. Interact.* **27**, 34:1–34:53, DOI: 10.1145/3398069 (2020).

23. El-Sappagh, S. H., Nazih, W., Alharbi, M. & AbuHmed, T. Responsible artificial intelligence for mental health disorders: Current applications and future challenges. *J. Disabil. Res.* **4**, 20240101 (2025).

24. Uyanik, H. *et al.* Automated detection of neurological and mental health disorders using eeg signals and artificial intelligence: A systematic review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **15**, e70002 (2025).

25. Tyagi, A., Singh, V. P. & Gore, M. M. Towards artificial intelligence in mental health: a comprehensive survey on the detection of schizophrenia. *Multimed. Tools Appl.* **82**, 20343–20405 (2023).

26. Sahili, Z. A., Patras, I. & Purver, M. Multimodal machine learning in mental health: A survey of data, algorithms, and challenges. *CoRR* **abs/2407.16804**, DOI: 10.48550/ARXIV.2407.16804 (2024). 2407.16804.

27. Sarioglu Kayi, E., Diab, M., Pauselli, L., Compton, M. & Coppersmith, G. Predictive linguistic features of schizophrenia. In Ide, N., Herbelot, A. & Màrquez, L. (eds.) *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)*, 241–250, DOI: 10.18653/v1/S17-1028 (Association for Computational Linguistics, Vancouver, Canada, 2017).

28. Gehrmann, S. *et al.* Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PloS one* **13**, e0192360 (2018).

29. Mohammad, S. M. WorryWords: Norms of anxiety association for over 44k English words. In Al-Onaizan, Y., Bansal, M. & Chen, Y.-N. (eds.) *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 16261–16278, DOI: 10.18653/v1/2024.emnlp-main.910 (Association for Computational Linguistics, Miami, Florida, USA, 2024).

30. Hou, Y., Xu, J., Huang, Y. & Ma, X. A big data application to predict depression in the university based on the reading habits. In *2016 3rd International Conference on Systems and Informatics (ICSAI)*, 1085–1089, DOI: 10.1109/ICSAI.2016.7811112 (2016).

31. Turner, J. C. & Keller, A. College Health Surveillance Network: Epidemiology and Health Care Utilization of College Students at US 4-Year Universities. *J. Am. Coll. Heal.* **63**, 530–538, DOI: 10.1080/07448481.2015.1055567 (2015). Publisher: Informa UK Limited.

32. Gideon, J., Provost, E. M. & McInnis, M. G. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016, Shanghai, China, March 20-25, 2016*, 2359–2363, DOI: 10.1109/ICASSP.2016.7472099 (IEEE, 2016).

33. Islam, K. A., Pérez, D. & Li, J. A transfer learning approach for the 2018 FEMH voice data challenge. In Abe, N. *et al.* (eds.) *IEEE International Conference on Big Data (IEEE BigData 2018), Seattle, WA, USA, December 10-13, 2018*, 5252–5257, DOI: 10.1109/BIGDATA.2018.8622447 (IEEE, 2018).

34. Banerjee, D. *et al.* A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. *Knowl. Inf. Syst.* **60**, 1693–1724 (2019).

35. Kathan, A. *et al.* The effect of clinical intervention on the speech of individuals with PTSD: features and recognition performances. In Harte, N., Carson-Berndsen, J. & Jones, G. (eds.) *24th Annual Conference of the International Speech Communication Association, Interspeech 2023, Dublin, Ireland, August 20-24, 2023*, 4139–4143, DOI: 10.21437/INTERSPEECH.2023-1668 (ISCA, 2023).

36. Vergyri, D. *et al.* Speech-based assessment of PTSD in a military population using diverse feature classes. In *16th Annual Conference of the International Speech Communication Association, INTERSPEECH 2015, Dresden, Germany, September 6-10, 2015*, 3729–3733, DOI: 10.21437/INTERSPEECH.2015-740 (ISCA, 2015).

37. Marmar, C. R. *et al.* Speech-based markers for posttraumatic stress disorder in us veterans. *Depress. anxiety* **36**, 607–616 (2019).

38. Hu, J., Zhao, C., Shi, C., Zhao, Z. & Ren, Z. Speech-based recognition and estimating severity of ptsd using machine learning. *J. Affect. Disord.* **362**, 859–868, DOI: https://doi.org/10.1016/j.jad.2024.07.015 (2024).

39. Cummins, N. *et al.* Multilingual markers of depression in remotely collected speech samples: A preliminary analysis. *J. Affect. Disord.* **341**, 128–136, DOI: 10.1016/j.jad.2023.08.097 (2023).

40. Chang, K.-h., Chan, M. K. & Canny, J. AnalyzeThis: unobtrusive mental health monitoring by voice. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, 1951–1956, DOI: 10.1145/1979742.1979859 (ACM, Vancouver BC Canada, 2011).

41. van den Broek, E. L., van der Sluis, F. & Dijkstra, T. Cross-validation of bimodal health-related stress assessment. *Pers. Ubiquitous Comput.* **17**, 215–227 (2013).

42. Salekin, A., Eberle, J. W., Glenn, J. J., Teachman, B. A. & Stankovic, J. A. A Weakly Supervised Learning Framework for Detecting Social Anxiety and Depression. *Proc. ACM on Interactive, Mobile, Wearable Ubiquitous Technol.* **2**, 1–26, DOI: 10.1145/3214284 (2018). Publisher: Association for Computing Machinery (ACM).

43. Abbas, A. *et al.* Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. *Digit. Biomarkers* **5**, 29–36 (2021).

44. Shafique, S. *et al.* Towards automatic detection of social anxiety disorder via gaze interaction. *Appl. Sci.* **12**, 12298 (2022).

45. Langer, J. K., Lim, M. H., Fernandez, K. C. & Rodebaugh, T. L. Social anxiety disorder is associated with reduced eye contact during conversation primed for conflict. *Cogn. therapy research* **41**, 220–229 (2017).

46. Pampouchidou, A. *et al.* Automated facial video-based recognition of depression and anxiety symptom severity: cross-corpus validation. *Mach. Vis. Appl.* **31**, 30 (2020).

47. Jiang, Z. *et al.* Classifying major depressive disorder and response to deep brain stimulation over time by analyzing facial expressions. *IEEE Trans. Biomed. Eng.* **68**, 664–672, DOI: 10.1109/TBME.2020.3010472 (2021).

48. Gilanie, G. *et al.* A robust method of bipolar mental illness detection from facial micro expressions using machine learning methods. *Intell. Autom. & Soft Comput.* **39** (2024).

49. Shen, Y., Yang, H. & Lin, L. Automatic depression detection: an emotional audio-textual corpus and A gru/bilstm-based model. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23-27 May 2022*, 6247–6251, DOI: 10.1109/ICASSP43922.2022.9746569 (IEEE, 2022).

50. Aloshban, N., Esposito, A. & Vinciarelli, A. Language or paralanguage, this is the problem: Comparing depressed and non-depressed speakers through the analysis of gated multimodal units. In Hermansky, H. *et al.* (eds.) *22nd Annual Conference of the International Speech Communication Association, Interspeech 2021, Brno, Czechia, August 30 - September 3, 2021*, 2496–2500, DOI: 10.21437/INTERSPEECH.2021-928 (ISCA, 2021).

51. Qin, J. *et al.* Mental-perceiver: Audio-textual multi-modal learning for estimating mental disorders. In Walsh, T., Shah, J. & Kolter, Z. (eds.) *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, 25029–25037, DOI: 10.1609/AAAI.V39I23.34687 (AAAI Press, 2025).

52. Tang, S. X. *et al.* Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *npj Schizophr.* **7**, 25 (2021).

53. Wawer, A., Chojnicka, I., Okruszek, L. & Sarzynska-Wawer, J. Single and cross-disorder detection for autism and schizophrenia. *Cogn. Comput.* **14**, 461–473 (2022).

54. Hong, K. *et al.* Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. *Psychiatry Res.* **225**, 40–49, DOI: https://doi.org/10.1016/j.psychres.2014.10.002 (2015).

55. Allende-Cid, H., Zamora, J., Alfaro-Faccio, P. & Alonso-Sánchez, M. F. A machine learning approach for the automatic classification of schizophrenic discourse. *IEEE Access* **7**, 45544–45553, DOI: 10.1109/ACCESS.2019.2908620 (2019).

56. Iter, D., Yoon, J. & Jurafsky, D. Automatic detection of incoherent speech for diagnosing schizophrenia. In Loveys, K., Niederhoffer, K., Prud'hommeaux, E., Resnik, R. & Resnik, P. (eds.) *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, CLPsych@NAACL-HTL, New Orleans, LA, USA, June 2018*, 136–146, DOI: 10.18653/V1/W18-0615 (Association for Computational Linguistics, 2018).

57. Elvevåg, B., Foltz, P. W., Rosenstein, M. & DeLisi, L. E. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *J. Neurolinguistics* **23**, 270–284, DOI: https://doi.org/10.1016/j.jneuroling.2009.05.002 (2010). Language, Communication and Schizophrenia.

58. Bedi, G. *et al.* Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophr.* **1**, 1–7 (2015).

59. Elvevåg, B., Foltz, P. W., Weinberger, D. R. & Goldberg, T. E. Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. *Schizophr. Res.* **93**, 304–316, DOI: https://doi.org/10.1016/j.schres.2007.03.001 (2007).

60. Li, R. *et al.* Deciphering language disturbances in schizophrenia: A study using fine-tuned language models. *Schizophr. Res.* **271**, 120–128, DOI: https://doi.org/10.1016/j.schres.2024.07.016 (2024).

61. Ciampelli, S., Voppel, A., de Boer, J., Koops, S. & Sommer, I. Combining automatic speech recognition with semantic natural language processing in schizophrenia. *Psychiatry Res.* **325**, 115252, DOI: https://doi.org/10.1016/j.psychres.2023.115252 (2023).

62. Çabuk, T. *et al.* Natural language processing for defining linguistic features in schizophrenia: A sample from turkish speakers. *Schizophr. Res.* **266**, 183–189, DOI: https://doi.org/10.1016/j.schres.2024.02.026 (2024).

63. Xu, S. *et al.* Automatic verbal analysis of interviews with schizophrenic patients. In *23rd IEEE International Conference on Digital Signal Processing, DSP 2018, Shanghai, China, November 19-21, 2018*, 1–5, DOI: 10.1109/ICDSP.2018.8631830 (IEEE, 2018).

64. Jeong, L. *et al.* Exploring the use of natural language processing for objective assessment of disorganized speech in schizophrenia. *Psychiatr. Res. Clin. Pract.* **5**, 84–92, DOI: 10.1176/appi.prcp.20230003 (2023). https://doi.org/10.1176/appi.prcp.20230003.

65. Parola, A. *et al.* Speech disturbances in schizophrenia: Assessing cross-linguistic generalizability of nlp automated measures of coherence. *Schizophr. Res.* **259**, 59–70, DOI: https://doi.org/10.1016/j.schres.2022.07.002 (2023). Language and Speech Analysis in Schizophrenia and Related Psychoses.

66. Aich, A. *et al.* Towards intelligent clinically-informed language analyses of people with bipolar disorder and schizophrenia. In Goldberg, Y., Kozareva, Z. & Zhang, Y. (eds.) *Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, 2871–2887, DOI: 10.18653/V1/2022.FINDINGS-EMNLP.208 (Association for Computational Linguistics, 2022).

67. Arslan, B. *et al.* Computational analysis of linguistic features in speech samples of first-episode bipolar disorder and psychosis. *J. Affect. Disord.* **363**, 340–347, DOI: https://doi.org/10.1016/j.jad.2024.07.102 (2024).

68. Tasnim, M., Ehghaghi, M., Diep, B. & Novikova, J. DEPAC: a Corpus for Depression and Anxiety Detection from Speech, DOI: 10.48550/arXiv.2306.12443 (2023). ArXiv:2306.12443 [eess].

69. Diederich, J., Al-Ajmi, A. & Yellowlees, P. Ex-ray: Data mining and mental health. *Appl. Soft Comput.* **7**, 923–928, DOI: 10.1016/j.asoc.2006.04.007 (2007).

70. Hayati, M. F. M., Ali, M. A. M. & Rosli, A. N. M. Depression Detection on Malay Dialects Using GPT-3. In *2022 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 360–364, DOI: 10.1109/IECBES54088.2022.10079554 (2022).

71. Jiang, Z. *et al.* Multimodal mental health digital biomarker analysis from remote interviews using facial, vocal, linguistic, and cardiovascular patterns. *IEEE J. Biomed. Heal. Informatics* **28**, 1680–1691, DOI: 10.1109/JBHI.2024.3352075 (2024).

72. Valstar, M. F. *et al.* AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. In Schuller, B. W., Valstar, M. F., Cowie, R., Krajewski, J. & Pantic, M. (eds.) *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, AVEC@ACM Multimedia 2013, Barcelona, Spain, October 21, 2013*, 3–10, DOI: 10.1145/2512530.2512533 (ACM, 2013).

73. Dibeklioglu, H., Hammal, Z. & Cohn, J. F. Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE J. Biomed. Heal. Informatics* **22**, 525–536, DOI: 10.1109/JBHI.2017.2676878 (2018).

74. Alghowinem, S. *et al.* Multimodal depression detection: Fusion analysis of paralinguistic, head pose and eye gaze behaviors. *IEEE Trans. Affect. Comput.* **9**, 478–490, DOI: 10.1109/TAFFC.2016.2634527 (2018).

75. Lin, W., Orton, I., Li, Q., Pavarini, G. & Mahmoud, M. Looking at the body: Automatic analysis of body gestures and self-adaptors in psychological distress. *IEEE Trans. Affect. Comput.* **14**, 1175–1187, DOI: 10.1109/TAFFC.2021.3101698 (2023).

76. Guo, W., Yang, H., Liu, Z., Xu, Y. & Hu, B. Deep neural networks for depression recognition based on 2d and 3d facial expressions under emotional stimulus tasks. *Front. neuroscience* **15**, 609760 (2021).

77. DeVault, D. *et al.* Simsensei kiosk: a virtual human interviewer for healthcare decision support. In Bazzan, A. L. C., Huhns, M. N., Lomuscio, A. & Scerri, P. (eds.) *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, 1061–1068 (IFAAMAS/ACM, 2014).

78. Zou, B. *et al.* Semi-structural interview-based chinese multimodal depression corpus towards automatic preliminary screening of depressive disorders. *IEEE Trans. Affect. Comput.* **14**, 2823–2838, DOI: 10.1109/TAFFC.2022.3181210 (2023).

79. Stratou, G., Scherer, S., Gratch, J. & Morency, L. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender. *J. Multimodal User Interfaces* **9**, 17–29, DOI: 10.1007/S12193-014-0161-4 (2015).

80. Schultebraucks, K., Yadav, V., Shalev, A. Y., Bonanno, G. A. & Galatzer-Levy, I. R. Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. *Psychol. Medicine* **52**, 957–967, DOI: 10.1017/S0033291720002718 (2022).

81. Zhu, Z. *et al.* MEDIC: A multimodal empathy dataset in counseling. In El-Saddik, A. *et al.* (eds.) *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, 6054–6062, DOI: 10.1145/3581783.3612346 (ACM, 2023).

82. Çiftçi, E., Kaya, H., Güleç, H. & Salah, A. A. The turkish audio-visual bipolar disorder corpus. In *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, 1–6, DOI: 10.1109/ACIIAsia.2018.8470362 (2018).

83. Chuang, C.-Y. *et al.* Multimodal assessment of schizophrenia symptom severity from linguistic, acoustic and visual cues. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 3469–3479, DOI: 10.1109/TNSRE.2023.3307597 (2023).

84. Premananth, G. *et al.* A multimodal framework for the assessment of the schizophrenia spectrum. In Lapidot, I. & Gannot, S. (eds.) *25th Annual Conference of the International Speech Communication Association, Interspeech 2024, Kos, Greece, September 1-5, 2024*, DOI: 10.21437/INTERSPEECH.2024-2224 (ISCA, 2024).

85. Zhang, J. *et al.* Automatic schizophrenia detection using multimodality media via a text reading task. *Front. Neurosci.* **Volume 16 - 2022**, DOI: 10.3389/fnins.2022.933049 (2022).

86. Tao, Y. *et al.* Classifying Anxiety and Depression through LLMs Virtual Interactions: A Case Study with ChatGPT. In *Proceedings - 2023 2023 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2023*, 2259–2264, DOI: 10.1109/BIBM58861.2023.10385305 (Institute of Electrical and Electronics Engineers Inc., 2023).

87. Cai, H. *et al.* A multi-modal open dataset for mental-disorder analysis. *Sci. Data* **9**, 178 (2022).

88. Yadav, M. *et al.* Exploring individual differences of public speaking anxiety in real-life and virtual presentations, DOI: 10.1109/TAFFC.2020.3048299 (2022).

89. The mind research network uonm (2012) the center for biomedical research excellence (cobre). (2012).

90. Park, S. M. *et al.* Identification of Major Psychiatric Disorders From Resting-State Electroencephalography Using a Machine Learning Approach. *Front. Psychiatry* **12**, DOI: 10.3389/fpsyt.2021.707581 (2021). Publisher: Frontiers.

91. Olejarczyk, E. & Jernajczyk, W. Graph-based analysis of brain connectivity in schizophrenia. *PLOS ONE* **12**, e0188629, DOI: 10.1371/journal.pone.0188629 (2017). Publisher: Public Library of Science.

92. Wang, L. *et al.* SchizConnect: Mediating Neuroimaging Databases on Schizophrenia and Related Disorders for Large-Scale Integration. *NeuroImage* **124**, 1155–1167, DOI: 10.1016/j.neuroimage.2015.06.065 (2016).

93. Bilder, R. *et al.* UCLA Consortium for Neuropsychiatric Phenomics LA5c Study, DOI: 10.18112/openneuro.ds000030.v1.0.0 (2018).

94. Wang, L. *et al.* Northwestern university schizophrenia data and software tool (nusdast). *Front. neuroinformatics* **7**, 25 (2013).

95. Gollub, R. L. *et al.* The MCIC Collection: A Shared Repository of Multi-Modal, Multi-Site Brain Image Data from a Clinical Investigation of Schizophrenia. *Neuroinformatics* **11**, 367–388, DOI: 10.1007/s12021-013-9184-3 (2013).

96. Silva, R. F. *et al.* The tenth annual mlsp competition: Schizophrenia classification challenge. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6, DOI: 10.1109/MLSP.2014.6958889 (2014).

97. Keator, D. B. *et al.* The Function Biomedical Informatics Research Network Data Repository. *NeuroImage* **124**, 1074–1079, DOI: 10.1016/j.neuroimage.2015.09.003 (2016).

98. Cai, H. *et al.* A Pervasive Approach to EEG-Based Depression Detection. *Complexity* **2018**, 5238028, DOI: 10.1155/2018/5238028 (2018). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2018/5238028.

99. Peng, H. *et al.* Multivariate Pattern Analysis of EEG-Based Functional Connectivity: A Study on the Identification of Depression. *IEEE Access* **7**, 92630–92641, DOI: 10.1109/ACCESS.2019.2927121 (2019).

100. Zhu, J. *et al.* Multimodal Mild Depression Recognition Based on EEG-EM Synchronization Acquisition Network. *IEEE Access* **7**, 28196–28210, DOI: 10.1109/ACCESS.2019.2901950 (2019).

101. Mumtaz, W., Xia, L., Mohd Yasin, M. A., Azhar Ali, S. S. & Malik, A. S. A wavelet-based technique to predict treatment outcome for Major Depressive Disorder. *PLOS ONE* **12**, e0171409, DOI: 10.1371/journal.pone.0171409 (2017). Publisher: Public Library of Science (PLoS).

102. Cavanagh, J. F., Bismark, A. W., Frank, M. J. & Allen, J. J. B. Multiple Dissociations Between Comorbid Depression and Anxiety on Reward and Punishment Processing: Evidence From Computationally Informed EEG. *Comput. Psychiatry (Cambridge, Mass.)* **3**, 1–17, DOI: 10.1162/cpsy_a_00024 (2019).

103. Luo, G. *et al.* Exploring Adaptive Graph Topologies and Temporal Graph Networks for EEG-Based Depression Detection. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 3947–3957, DOI: 10.1109/TNSRE.2023.3320693 (2023).

104. Garg, S., Shukla, U. P. & Cenkeramaddi, L. R. Detection of Depression Using Weighted Spectral Graph Clustering With EEG Biomarkers. *IEEE Access* **11**, 57880–57894, DOI: 10.1109/ACCESS.2023.3281453 (2023).

105. Li, L. *et al.* Construction of a resting EEG-based depression recognition model for college students and possible mechanisms of action of different types of exercise. *BMC psychiatry* **23**, 849, DOI: 10.1186/s12888-023-05352-0 (2023).

106. Shen, J. *et al.* Exploring the Intrinsic Features of EEG Signals via Empirical Mode Decomposition for Depression Recognition. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **31**, 356–365, DOI: 10.1109/TNSRE.2022.3221962 (2023).

107. Chung, K.-H., Chang, Y.-S., Yen, W.-T., Lin, L. & Abimannan, S. Depression assessment using integrated multi-featured EEG bands deep neural network models: Leveraging ensemble learning techniques. *Comput. Struct. Biotechnol. J.* **23**, 1450–1468, DOI: 10.1016/j.csbj.2024.03.022 (2024).

108. Cavanagh, J. F., Napolitano, A., Wu, C. & Mueen, A. The Patient Repository for EEG Data + Computational Tools (PRED+CT). *Front. Neuroinformatics* **11**, DOI: 10.3389/fninf.2017.00067 (2017). Publisher: Frontiers.

109. Ros, T. *et al.* Neurofeedback Tunes Scale-Free Dynamics in Spontaneous Brain Activity. *Cereb. Cortex (New York, N.Y.: 1991)* **27**, 4911–4922, DOI: 10.1093/cercor/bhw285 (2017).

110. Nicholson, A. A. *et al.* A randomized, controlled trial of alpha-rhythm EEG neurofeedback in posttraumatic stress disorder: A preliminary investigation showing evidence of decreased PTSD symptoms and restored default mode and salience network connectivity using fMRI. *NeuroImage: Clin.* **28**, 102490, DOI: 10.1016/j.nicl.2020.102490 (2020).

111. Kim, J.-Y., Lee, H. S. & Lee, S.-H. EEG Source Network for the Diagnosis of Schizophrenia and the Identification of Subtypes Based on Symptom Severity—A Machine Learning Approach. *J. Clin. Medicine* **9**, 3934, DOI: 10.3390/jcm9123934 (2020). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

112. Barros, C., Roach, B., Ford, J. M., Pinheiro, A. P. & Silva, C. A. From Sound Perception to Automatic Detection of Schizophrenia: An EEG-Based Deep Learning Approach. *Front. Psychiatry* **12**, DOI: 10.3389/fpsyt.2021.813460 (2022). Publisher: Frontiers.

113. Borisov, S. V., Kaplan, A. Y., Gorbachevskaya, N. L. & Kozlova, I. A. Analysis of EEG Structural Synchrony in Adolescents with Schizophrenic Disorders. *Hum. Physiol.* **31**, 255–261, DOI: 10.1007/s10747-005-0042-z (2005).

114. Albrecht, M. A., Waltz, J. A., Cavanagh, J. F., Frank, M. J. & Gold, J. M. Increased conflict-induced slowing, but no differences in conflict-induced positive or negative prediction error learning in patients with schizophrenia. *Neuropsychologia* **123**, 131–140, DOI: 10.1016/j.neuropsychologia.2018.04.031 (2019).

115. Tanaka, S. C. *et al.* A multi-site, multi-disorder resting-state magnetic resonance image database. *Sci. Data* **8**, DOI: 10.1038/s41597-021-01004-8 (2021). Publisher: Springer Science and Business Media LLC.

116. Hamilton, M. A rating scale for depression. *J. neurology, neurosurgery, psychiatry* **23**, 56 (1960).

117. Kroenke, K., Spitzer, R. L. & Williams, J. B. The phq-9: validity of a brief depression severity measure. *J. general internal medicine* **16**, 606–613 (2001).

118. Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch. general psychiatry* **4**, 561–571 (1961).

119. Zung, W. W. A self-rating depression scale. *Arch. general psychiatry* **12**, 63–70 (1965).

120. Rush, A. J. *et al.* The 16-item quick inventory of depressive symptomatology (qids), clinician rating (qids-c), and self-report (qids-sr): a psychometric evaluation in patients with chronic major depression. *Biol. psychiatry* **54**, 573–583 (2003).

121. Eaton, W. W., Muntaner, C., Smith, C., Tien, A. & Ybarra, M. Center for epidemiologic studies depression scale: Review and revision. *The use psychological testing for treatment planning outcomes assessment* 363–377 (2004).

122. Zigmond, A. S. & Snaith, R. P. The hospital anxiety and depression scale. *Acta psychiatrica scandinavica* **67**, 361–370 (1983).

123. Rush, A. J., Carmody, T. & Reimitz, P.-E. The inventory of depressive symptomatology (ids): clinician (ids-c) and self-report (ids-sr) ratings of depressive symptoms. *Int. journal methods psychiatric research* **9**, 45–59 (2000).

124. HEIMBERG, R. G. *et al.* Psychometric properties of the liebowitz social anxiety scale. *Psychol. Medicine* **29**, 199–212, DOI: 10.1017/S0033291798007879 (1999).

125. Fountoulakis, K. N. *et al.* Reliability and psychometric properties of the greek translation of the state-trait anxiety inventory form y: preliminary data. *Annals Gen. Psychiatry* **5**, 1–10 (2006).

126. Spitzer, R. L., Kroenke, K., Williams, J. B. W. & Löwe, B. A brief measure for assessing generalized anxiety disorder: The gad-7. *Arch. Intern. Medicine* **166**, 1092–1097, DOI: 10.1001/archinte.166.10.1092 (2006). https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/410326/ioi60000.pdf.

127. Mattick, R. P. & Clarke, J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety11editor's note: This article was written before the development of some contemporary measures of social phobia, such as the social phobia and anxiety inventory (turner et al., 1989). we have invited this article for publication because of the growing interest in the scales described therein. s.t. *Behav. Res. Ther.* **36**, 455–470, DOI: https://doi.org/10.1016/S0005-7967(97)10031-6 (1998).

128. Young, R. C., Biggs, J. T., Ziegler, V. E. & Meyer, D. A. A rating scale for mania: Reliability, validity and sensitivity. *Br. J. Psychiatry* **133**, 429–435, DOI: 10.1192/bjp.133.5.429 (1978).

129. Wolpe, J. Psychotherapy by reciprocal inhibition. *Cond. reflex: a Pavlovian journal research & therapy* **3**, 234–240 (1968).

130. Weathers, F. W. *et al.* The ptsd checklist for dsm-5 (pcl-5) (2013).

131. Blanchard, E. B., Jones-Alexander, J., Buckley, T. C. & Forneris, C. A. Psychometric properties of the ptsd checklist (pcl). *Behav. Res. Ther.* **34**, 669–673, DOI: https://doi.org/10.1016/0005-7967(96)00033-2 (1996).

132. Andreasen, N. C. *Scale for the assessment of positive symptoms (SAPS)* (University of Iowa, 2000).

133. Andreasen, N. C. *Scale for the assessment of negative symptoms (SANS).* (University of Iowa Iowa City, 1981).

134. Andreasen, N. C. Negative symptoms in schizophrenia: definition and reliability. *Arch. general psychiatry* **39**, 784–788 (1982).

135. Kay, S. R., Fiszbein, A. & Opler, L. A. The positive and negative syndrome scale (panss) for schizophrenia. *Schizophr. bulletin* **13**, 261–276 (1987).

136. Andreasen, N. C. Scale for the assessment of thought, language, and communication (tlc). *Schizophr. bulletin* **12**, 473 (1986).

137. OpenAI. Chatgpt (2023). Large language model, accessed July 2025.

138. Wu, Y., Chen, J., Mao, K. & Zhang, Y. Automatic post-traumatic stress disorder diagnosis via clinical transcripts: A novel text augmentation with large language models. In *IEEE Biomedical Circuits and Systems Conference, BioCAS 2023, Toronto, ON, Canada, October 19-21, 2023*, 1–5, DOI: 10.1109/BIOCAS58349.2023.10388714 (IEEE, 2023).

139. Liu, J. M. *et al.* Chatcounselor: A large language models for mental health support. *CoRR* **abs/2309.15461**, DOI: 10.48550/ARXIV.2309.15461 (2023). 2309.15461.

140. Xu, J. *et al.* Mentalchat16k: A benchmark dataset for conversational mental health assistance. *CoRR* **abs/2503.13509**, DOI: 10.48550/ARXIV.2503.13509 (2025). 2503.13509.

141. Yin, C. *et al.* Mdd-5k: A new diagnostic conversation dataset for mental disorders synthesized via neuro-symbolic LLM agents. In Walsh, T., Shah, J. & Kolter, Z. (eds.) *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, 25715–25723, DOI: 10.1609/AAAI.V39I24.34763 (AAAI Press, 2025).

142. Yao, B. *et al.* D4: a chinese dialogue dataset for depression-diagnosis-oriented chat. In Goldberg, Y., Kozareva, Z. & Zhang, Y. (eds.) *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, 2438–2459, DOI: 10.18653/V1/2022.EMNLP-MAIN.156 (Association for Computational Linguistics, 2022).

143. Mousavi, S. M., Cervone, A., Danieli, M. & Riccardi, G. Would you like to tell me more? generating a corpus of psychotherapy dialogues. In Shivade, C. *et al.* (eds.) *Proceedings of the Second Workshop on Natural Language Processing for Medical Conversations*, 1–9, DOI: 10.18653/v1/2021.nlpmc-1.1 (Association for Computational Linguistics, Online, 2021).

144. BN, S. *et al.* Thousand voices of trauma: A large-scale synthetic dataset for modeling prolonged exposure therapy conversations. *CoRR* **abs/2504.13955**, DOI: 10.48550/ARXIV.2504.13955 (2025). 2504.13955.

145. Mishra, K., Priya, P., Burja, M. & Ekbal, A. e-therapist: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts. In Bouamor, H., Pino, J. & Bali, K. (eds.) *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, 13952–13967, DOI: 10.18653/V1/2023.EMNLP-MAIN.861 (Association for Computational Linguistics, 2023).

146. Xiao, M. *et al.* Healme: Harnessing cognitive reframing in large language models for psychotherapy. In Ku, L., Martins, A. & Srikumar, V. (eds.) *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, 1707–1725, DOI: 10.18653/V1/2024.ACL-LONG.93 (Association for Computational Linguistics, 2024).

147. Maddela, M. *et al.* Training models to generate, recognize, and reframe unhelpful thoughts. In Rogers, A., Boyd-Graber, J. L. & Okazaki, N. (eds.) *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, 13641–13660, DOI: 10.18653/V1/2023.ACL-LONG.763 (Association for Computational Linguistics, 2023).

148. Qiu, H., He, H., Zhang, S., Li, A. & Lan, Z. SMILE: single-turn to multi-turn inclusive language expansion via chatgpt for mental health support. In Al-Onaizan, Y., Bansal, M. & Chen, Y. (eds.) *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, 615–636 (Association for Computational Linguistics, 2024).

149. Chen, Y. *et al.* Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations. In Bouamor, H., Pino, J. & Bali, K. (eds.) *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, 1170–1183, DOI: 10.18653/V1/2023.FINDINGS-EMNLP.83 (Association for Computational Linguistics, 2023).

150. Sun, H., Lin, Z., Zheng, C., Liu, S. & Huang, M. Psyqa: A chinese dataset for generating long counseling text for mental health support. In Zong, C., Xia, F., Li, W. & Navigli, R. (eds.) *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, vol. ACL/IJCNLP 2021 of *Findings of ACL*, 1489–1503, DOI: 10.18653/V1/2021.FINDINGS-ACL.130 (Association for Computational Linguistics, 2021).

151. Na, H. CBT-LLM: A chinese large language model for cognitive behavioral therapy-based mental health question answering. In Calzolari, N. *et al.* (eds.) *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, 2930–2940 (ELRA and ICCL, 2024).

152. Lee, S. *et al.* Cactus: Towards psychological counseling conversations using cognitive behavioral theory. In Al-Onaizan, Y., Bansal, M. & Chen, Y. (eds.) *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, 14245–14274 (Association for Computational Linguistics, 2024).

153. Zhang, C. *et al.* Cpsycoun: A report-based multi-turn dialogue reconstruction and evaluation framework for chinese psychological counseling. In Ku, L., Martins, A. & Srikumar, V. (eds.) *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, 13947–13966, DOI: 10.18653/V1/2024.FINDINGS-ACL.830 (Association for Computational Linguistics, 2024).

154. Kim, S., Kim, H., Do, H. & Lee, G. Multimodal cognitive reframing therapy via multi-hop psychotherapeutic reasoning. In Chiruzzo, L., Ritter, A. & Wang, L. (eds.) *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2025 - Volume 1: Long Papers, Albuquerque, New Mexico, USA, April 29 - May 4, 2025*, 4863–4880 (Association for Computational Linguistics, 2025).

155. Kim, S., Kim, H., Lee, J., Jeon, Y. & Lee, G. G. Mirror: Multimodal cognitive reframing therapy for rolling with resistance. *CoRR* **abs/2504.13211**, DOI: 10.48550/ARXIV.2504.13211 (2025). 2504.13211.

156. Chim, J. *et al.* Overview of the CLPsych 2024 shared task: Leveraging large language models to identify evidence of suicidality risk in online posts. In Yates, A. *et al.* (eds.) *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)*, 177–190 (Association for Computational Linguistics, St. Julians, Malta, 2024).

157. Cummins, N. *et al.* A review of depression and suicide risk assessment using speech analysis. *Speech Commun.* **71**, 10–49 (2015).

158. Slonim, D. A. *et al.* Facing change: using automated facial expression analysis to examine emotional flexibility in the treatment of depression. *Adm. Policy Mental Heal. Mental Heal. Serv. Res.* **51**, 501–508 (2024).

159. Ringeval, F. *et al.* Avec 2017: Real-life depression, and affect recognition workshop and challenge. In *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, AVEC '17, 3–9, DOI: 10.1145/3133944.3133953 (Association for Computing Machinery, New York, NY, USA, 2017).

160. Ringeval, F. *et al.* Avec 2019 workshop and challenge: State-of-mind, detecting depression with ai, and cross-cultural affect recognition. In *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, AVEC '19, 3–12, DOI: 10.1145/3347320.3357688 (Association for Computing Machinery, New York, NY, USA, 2019).

161. Ringeval, F. *et al.* AVEC 2018 workshop and challenge: Bipolar disorder and cross-cultural affect recognition. In Ringeval, F., Schuller, B. W., Valstar, M. F., Cowie, R. & Pantic, M. (eds.) *Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop, AVEC@MM 2018, Seoul, Republic of Korea, October 22, 2018*, 3–13, DOI: 10.1145/3266302.3266316 (ACM, 2018).

162. Sue, D. W. *et al.* Position paper: Cross-cultural counseling competencies. *The Couns. Psychol.* **10**, 45–52, DOI: 10.1177/0011000082102008 (1982). https://doi.org/10.1177/0011000082102008.

163. Magaña, D. Cultural competence and metaphor in mental healthcare interactions: A linguistic perspective. *Patient Educ. Couns.* **102**, 2192–2198, DOI: https://doi.org/10.1016/j.pec.2019.06.010 (2019).

164. Kirmayer, L. J. *et al.* Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *J. clinical psychiatry* **62**, 22–30 (2001).

165. Birtel, M. D. & Mitchell, B. L. Cross-cultural differences in depression between white british and south asians: Causal attributions, stigma by association, discriminatory potential. *Psychol. Psychother. Theory, Res. Pract.* **96**, 101–116 (2023).

166. Hofmann, S. G. & Hinton, D. E. Cross-cultural aspects of anxiety disorders. *Curr. psychiatry reports* **16**, 450 (2014).

167. Ceylan, A. C. & YALÇINKAYA ALKAR, Ö. The cross-cultural differences in symptoms of schizophrenia: A systematic review. *J. Cogn. Behav. Psychother. Res.* **12**, 179 (2023).

168. Douki, S., Nacef, F., Triki, T. & Dalery, J. Les aspects culturels du trouble bipolaire: résultats d'une étude comparative entre des patients français et tunisiens. *L'Encéphale* **38**, 194–200, DOI: https://doi.org/10.1016/j.encep.2011.04.003 (2012).

169. Patel, A. R. & Hall, B. J. Beyond the dsm-5 diagnoses: A cross-cultural approach to assessing trauma reactions. *Focus* **19**, 197–203, DOI: 10.1176/appi.focus.20200049 (2021). https://doi.org/10.1176/appi.focus.20200049.

170. Aleem, M., Zahoor, I. & Naseem, M. Towards culturally adaptive large language models in mental health: Using chatgpt as a case study. In Farzan, R. *et al.* (eds.) *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing, CSCW Companion 2024, San Jose, Costa Rica, November 9-13, 2024*, 240–247, DOI: 10.1145/3678884.3681858 (ACM, 2024).

171. Zahran, N., Fouda, A. E., Hanafy, R. J. & Fouda, M. E. A comprehensive evaluation of large language models on mental illnesses in arabic context. *CoRR* **abs/2501.06859**, DOI: 10.48550/ARXIV.2501.06859 (2025). 2501.06859.

172. McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Singh, A. & Zhu, X. J. (eds.) *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, vol. 54 of *Proceedings of Machine Learning Research*, 1273–1282 (PMLR, 2017).

173. Wu, S., Fei, H., Qu, L., Ji, W. & Chua, T. Next-gpt: Any-to-any multimodal LLM. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024* (OpenReview.net, 2024).

174. Nagy, B. *et al.* Privacy-preserving federated learning and its application to natural language processing. *Knowledge-Based Syst.* **268**, 110475, DOI: https://doi.org/10.1016/j.knosys.2023.110475 (2023).

175. Basu, P. *et al.* Benchmarking differential privacy and federated learning for BERT models. *CoRR* **abs/2106.13973** (2021). 2106.13973.

176. Mandal, A., Chakraborty, T. & Gurevych, I. Towards privacy-aware mental health AI models: Advances, challenges, and opportunities. *CoRR* **abs/2502.00451**, DOI: 10.48550/ARXIV.2502.00451 (2025). 2502.00451.

177. Li, Y. *et al.* Counselbench: A large-scale expert evaluation and adversarial benchmark of large language models in mental health counseling. *CoRR* **abs/2506.08584**, DOI: 10.48550/ARXIV.2506.08584 (2025). 2506.08584.