The Applications of Large Language Models in Mental Health: A Scoping Review

Yu Jin, Jiayi Liu, Pan Li, Baosen Wang, Yangxinyu Yan, Huilin Zhang, Chenhao Ni, Jing Wang, Yi Li, Yajun Bu, Yuanyuan Wang

Submitted to: Journal of Medical Internet Research
on: November 26, 2024

Table of Contents

The Applications of Large Language Models in Mental Health: A Scoping Review

Yu Jin[1*]; Jiayi Liu[1*]; Pan Li[1*]; Baosen Wang[1*]; Yangxinyu Yan[2]; Huilin Zhang[1]; Chenhao Ni[1]; Jing Wang[3]; Yi Li[4]; Yajun Bu[4]; Yuanyuan Wang[3]

[1]Beijing Normal University Beijing CN
[2]South China Normal University Guangzhou CN
[3]Guangdong Polytechnic Normal University Guangzhou CN
[4]The People's Hospital of Pingbian County, Honghe Honghe CN
[*]these authors contributed equally

Corresponding Author:
Yuanyuan Wang
Guangdong Polytechnic Normal University
Guangzhou No.55, Zhongshan West Avenue, Tianhe District, Guangzhou City, Guangdong Province, China
Guangzhou
CN

Abstract

Background: Mental health is emerging as an increasingly prevalent public issue, particularly within low- and middle-income nations. There is an urgent need in mental health for efficient detection methods, effective treatments, affordable privacy-focused healthcare solutions, and increased access to specialized psychiatrists. The emergence and rapid development of large language models (LLMs) have been leveraged to address these mental health demands. However, a comprehensive review summarizing the applications, processes, and performance of LLMs in mental health has been lacking up until now.

Objective: To summarize the applications of LLMs in mental health, encompassing trends, research areas, performance comparisons, and prospective future directions.

Methods: A scoping review was conducted to map the landscape of LLM applications in mental health, including trends, application domains, comparative performances, and future trajectories. We conducted a search across seven electronic databases, including Web of Science, PubMed, Cochrane Library, IEEE Xplore, Weipu, CNKI, and WanFang, from January 1, 2019, to August 31, 2024. Subsequent data-charting of eligible articles extracted relevant information on application aspects and performance metrics.

Results: A total of 95 articles were included, drawn from 4,544 studies, employing LLMs for mental health tasks. The applications were categorized into three key areas: screening and detection of mental disorders (n = 67), support for clinical treatments and interventions (n = 31), and assisting in mental health counseling and education (n = 11). The majority of studies utilized LLMs for depression detection and classification (34·7%), clinical treatment support and intervention (14·7%), and suicide risk prediction (12·6%). In comparison with non-transformer models and human experts, LLMs demonstrate superior capabilities in information acquisition and analysis, generating natural-language responses, and addressing complex reasoning problems. Assessments of LLM performances indicate that the majority of LLMs exhibit efficiency and promise in addressing mental health tasks.

Conclusions: This scoping review synthesizes the applications, processes, and performances of LLMs within the mental health field. These findings highlight the substantial potential of LLMs to augment mental health research, diagnostics, and intervention strategies, underscoring the imperative for ongoing development and ethical deliberation in clinical settings.

(JMIR Preprints 26/11/2024:69284)
DOI: https://doi.org/10.2196/preprints.69284

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

1

Review

Enter information for authors (including designations, affiliations, correspondence, contributions) in the online metadata form. Do not use periods after initials, and include degree designations and affiliations for all authors. Trial registration numbers are also filled in on the metadata forms online.

The Applications of Large Language Models in Mental Health: A Scoping Review

Abstract

Background: Mental health is emerging as an increasingly prevalent public issue, particularly within low- and middle-income nations. There is an urgent need in mental health for efficient detection methods, effective treatments, affordable privacy-focused healthcare solutions, and increased access to specialized psychiatrists. The emergence and rapid development of large language models (LLMs) have been leveraged to address these mental health demands. However, a comprehensive review summarizing the applications, processes, and performance of LLMs in mental health has been lacking up until now.
Objective: To summarize the applications of LLMs in mental health, encompassing trends, research areas, performance comparisons, and prospective future directions.
Methods: A scoping review was conducted to map the landscape of LLM applications in mental health, including trends, application domains, comparative performances, and future trajectories. We conducted a search across seven electronic databases, including Web of Science, PubMed, Cochrane Library, IEEE Xplore, Weipu, CNKI, and WanFang, from January 1, 2019, to August 31, 2024. Subsequent data-charting of eligible articles extracted relevant information on application aspects and performance metrics.
Results: A total of 95 articles were included, drawn from 4,544 studies, employing LLMs for mental health tasks. The applications were categorized into three key areas: screening and detection of mental disorders (n = 67), support for clinical treatments and interventions (n = 31), and assisting in mental health counseling and education (n = 11). The majority of studies utilized LLMs for depression detection and classification (34·7%), clinical treatment support and intervention (14·7%), and suicide risk prediction (12·6%). In comparison with non-transformer models and human experts, LLMs demonstrate superior capabilities in information acquisition and analysis, generating natural-language responses, and addressing complex reasoning problems. Assessments of LLM performances indicate that the majority of LLMs exhibit efficiency and promise in addressing mental health tasks.
Conclusions: This scoping review synthesizes the applications, processes, and performances of LLMs within the mental health field. These findings highlight the substantial potential of LLMs to augment mental health research, diagnostics, and intervention strategies, underscoring the imperative for ongoing development and ethical deliberation in clinical settings.

Keywords: Mental health; Large Language Models; Application; Process; Performance; Comparison.

Introduction

Mental health is a growing global public issue, impacting nearly a billion people worldwide, with an estimated one in seven adolescents affected.[1, 2] Nevertheless, more than 70% of individuals with mental health disorders are without access to essential support and services.[3] Furthermore, more than 720,000 people commit suicide annually, with nearly three-quarters of these suicides occurring in low- and middle-income countries.[4] Consequently, there is an urgent need in mental health to facilitate efficient detection from large-scale data, deliver effective treatments and interventions to large populations, and ensure private, affordable healthcare and increased access to specialized psychiatrists in low-income regions. To address inadequate access to effective and equitable mental health care, large-scale, innovative solutions are imperative.

Large language models (LLMs), emerging in 2022, are advanced natural language processing (NLP) models capable of analyzing vast textual data and generating human-like language.[5] Notable LLMs like GPT-3/4,[6] PaLM,[7] and LLaMA,[8] constitute a category of foundational models, each with billions of parameters, trained on extensive textual data.[9] Utilizing the Transformer architecture and self-supervised pre-training, LLMs are adept at tackling a variety of NLP tasks, including information extraction, interaction, content generation, and logical reasoning.[10] In comparison to prior NLP models,[11, 12] LLMs exhibit superior performance in computational efficiency, large-scale data analyses, interaction, and external validity and applicability.[9] Furthermore, LLMs can be fine-tuned to cater to specific domains, including mental health, thereby empowering them to engage in natural language interactions and accomplish mental health-related tasks. LLMs would help address insufficient mental healthcare system capacity, and provide efficient or personalized treatments. Therefore, the application of LLMs within mental health is expanding across diverse domains.[13-16]

Researchers have explored the applications of LLMs in mental health in various areas, encompassing screening or detecting mental disorders,[17-19] supporting the clinical treatments and interventions,[20-22] and assisting in mental health education.[17, 20, 23, 24] Nonetheless, no comprehensive review has yet synthesized these applications, assessed the performance of LLMs, or elucidated their advantages within the mental health domain. Therefore, we conducted this scoping review to address four questions. Firstly, we identified the challenges in mental health and compared the processes adopted by humans, Non-Transformer models, and LLMs. Secondly, we summarized the main areas of LLMs' applications and presented specific processes of these applications. Thirdly, we examined comparative performance studies between LLMs and humans, as well as among different LLMs. Finally, we presented the fine-tuning of LLMs for mental health, which could be directly employed by researchers and psychiatrists. This review aims to furnish a foundational understanding of LLM applications in mental health and offer a road map for future research and clinical practice.

Methods

Protocol registration

We drafted the study protocol based on the relevant items from the scoping review extension of the PRISMA guideline (PRISMA-ScR) (Supplementary Table S1). The final protocol was registered prospectively in the Open Science Framework (https: //osf.io/cxvwy).

Search strategy and selection criteria

A scoping review is a preliminary systematic review that aims to map the existing evidence on a specific topic or field of research.25 It provides a broad overview of the literature by identifying the nature and extent of existing research, including types of studies, variables, and gaps in the evidence base.26 This approach is particularly useful when the body of evidence is large or diverse, or when there is a need to understand the scope of a research area before conducting a more focused systematic review.

This scoping review followed the five-stage framework: (i) identifying the research question; (ii) identifying relevant studies; (iii) study selection; (iv) charting the data; (v) collating, summarizing, and reporting the results. The search terms for mental health include: "psychiatr*", "mental", "psycholog*", "depress*", "anxiety", "posttraumatic stress disorder", "PTSD", "bipolar disorder", "schizophrenia", "obsessive-compulsive disorder", "personality disorder", and "insomnia". The keywords and search terms for LLMs in mental health include: “large language model”, “language *model”, “pretrained model”, “language processing”, “OpenAI language model”, “generative AI”, “generative artificial intelligence”, “AI Chatbot”, “embedding”, “BERT”, “GPT”.

We searched five English-language databases (Web of Science, PubMed, Cochrane Library, and IEEE Xplore) and three Chinese-language databases (Weipu, CNKI, and Wanfang) for peer-reviewed articles published between January 1, 2019 and August 31, 2024. We only included papers in Chinese published in high-quality journals (i.e., Chinese Core Journals). To find other possibly relevant studies and reports that were missed by the automated searches, the reference lists of the included articles and reports were examined.

All studies meeting the following criteria were included in the current review: (i) studies focused on the applications of LLMs in mental health fields; (ii) the LLMs included but not limited to GPT-3/4, ChatGPT, PaLM, LLaMA, and the improved and fine-tuned LLMs; (iii) published in peer-reviewed journals. The general exclusion criteria were: (i) Not focused on mental health or mental disorders; (ii) Not a peer-reviewed research article or review.

The first round of screening was based on titles, keywords, and abstracts. The second round of review was based on the full text of potentially eligible papers and involved the peer-reviewed articles that met the inclusion and exclusion criteria. Six independent researchers (LJY, LP, WBS, YYXY, ZHL, and NCH) conducted both screening and full-text review. Disagreements were discussed with third reviewers (JY and WYY) until a consensus was reached. The search terms for English-language and Chinese-language databases are shown in the Supplementary Tables S2 and S3.

Data extraction, categorization, and labeling

The excluded studies and the final data-collection form used for peer-reviewed articles are shown in Supplementary Table S4 and Table 1. Six researchers (LJY, LP, WBS, YYXY, ZHL, and NCH) independently extracted and double-checked the data items from each included study. The information of each study included category, region, application task, mental condition, data source, sample information, and applied models (Table 1).

At least one reviewer categorized each study manually, by examining the title and abstract to assign categories. The methods and results sections were examined for categories when the studies' categories could not be divided. For this scoping review, we developed a categorization framework based on the applications of LLMs in mental health: (i) the screening or detection of mental disorders; (ii) supporting the clinical treatments and interventions; (iii) assisting in mental health counseling and education.

Statistical Analysis

Descriptive statistics were used to summarize the distribution of studies across different application areas. For each application task, we calculated frequencies and percentages. For the performance comparisons between humans and LLMs, and between various LLMs, we collected the metrics results and plotted them. Calculations and data-charting were performed using R software version 4·4·2 and Canva software.

Results

The initial search was 3,432 records, of which 430 duplicates were removed. Of the remaining 3,002 records, 2736 records were removed due to the irrelevant contents. After the full-text screening, 95 articles fulfilled the inclusion criteria (Fig. 1). Table 1 demonstrates the basic information of each study, including category, region, application task, mental condition, data source, sample information, and applied models.

Comparisons between LLMs, non-transformer models, and humans

The current challenges in mental health fields include difficulty in efficient mental disorders detection from large-scale data, effective treatment or intervention for large populations, private and low-cost healthcare, demand for professional psychiatrists, and so on. Compared with non-transformer models and humans, LLMs present superior capabilities in efficient parallel computing, textual generation, and fine-tuned tasks. Therefore, they could be applied in mental disorders detection or prediction, clinical treatment assistance, preliminarily medical care, and educational materials generation based on the datasets from social media platforms, electronic medical records (EMR), and counseling records. (Fig. 2).

Categorization of studies based on mental health applications

We categorized studies in terms of the applications of LLMs in the mental health field. These applications could be divided into three categories: the screening or detection of mental disorders (n = 67), supporting the clinical treatments and intervention (n = 31), and assisting in mental health counseling and education (n = 11) (Supplementary Tables S5). Each study was categorized with 1 or more applications, thus the percentages sum to more than 100%. Most studies applied LLMs for depression detection and classification (34·7%), supporting clinical treatments and interventions (14·7%), and suicide risk prediction (12·6%) (Supplementary Tables S5). These studies used data from social media platforms like Reddit, Twitter, Facebook, and Weibo, or clinical datasets (Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ), Extended DAIC (E-DAIC)), as well as semi-structured interviews from hospitals. When evaluating the performance of these LLMs, most studies measured the performance of LLMs with various metrics, such as F1 score (56·3%), precision (35·4%), accuracy (46·9%), recall (33·3%).
The number of studies mapped by countries is presented in Fig. 3a. The number of included studies increased year on year, and this trend is shown in Fig.3b. The first application area, mainly focused on depression detection and classification, suicide risk prediction, other mental disorders, and sentiment analysis (Fig. 3c). These studies applied basic LLMs or fine-tuned LLMs to detect or predict depression, suicide risk and other mental disorders (such as anxiety, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD)). The detailed process for these applications of LLMs is presented in Fig. 3c.
In the second application area, most studies explored the capability of LLMs in supporting clinical treatments and interventions, developing conversational virtual humans, and augmenting unbalanced clinical data. These studies applied LLMs to provide treatment advice, assist diagnostic services, and assess prognosis through a question-answering approach. The performance of LLMs was evaluated by professional clinicians and compared with the related performance of humans. Three studies also applied LLMs to develop conversational virtual humans to elicit emotion. Furthermore, to address the imbalance of clinical data and enhance diagnosis and treatment, LLMs could augment clinical data and targeted dialogues in safe and stable ways.
Moreover, LLMs have been applied in mental health counseling by developing conversational agents or chatbots,[27-30] and educational resource supplements. [31, 32] These results showed that the introduction of any interaction (video or chatbot) improved intent to practice and overall experience compared to the baseline.[33] Furthermore, LLMs showed the potential to generate educational materials for training.

Performance comparisons between LLMs and humans, and between various LLMs

Several studies compared the performance between humans and LLMs, the total metrics results of performance were displayed in Supplementary Table S6. Fig. 4 (part a) presents the results of three studies on OCD identification, chatbot efficiency, and treatment support. According to these results, most LLMs showed efficient and promising performance for mental health tasks. Several LLMs such as ChatGPT-4, Claude, and Bard aligned closely with mental health professionals' perspectives.[17, 20, 23, 24]

Fig. 4 (parts b and c) shows the comparisons of model performance between different LLMs in mental disorder diagnosis, clinical treatment assistance, and depression detection and classification. These results found that the latest LLMs (e.g., ChatGPT) perform better than traditional and previous models. The total results were collected in Supplementary Table S7.

Existing fine-tuned LLMs for mental health

Table 2 provides the fine-tuned LLMs for mental health, including availability, base models, the number of parameters, training strategy, and published year. These fine-tuned LLMs could be applied specifically for mental health tasks.

Common advantages and disadvantages

Fig. 5 summarizes the common advantages and disadvantages of LLMs in mental health. These LLMs could be divided into BERT series, GPT series, LLaMA series, and others. The common strengths of BERT series models are helpful for fine-tuning specific mental health issues. However, the BERT series models require large computational resources. The ChatGPT series models can conduct multi-round dialog, even based on small-sample learning. Nevertheless, the accuracy of ChatGPT models should be improved. Furthermore, GPT-4 could receive multimodal data and show more powerful performance in comprehension and generation. As for the LLAMA series models, they are open source for the public, and beneficial for interactive applications, although their complex task performance is inferior to large-scale proprietary models.

Discussion

With the rapid development of LLMs, their extensive natural language generation capabilities enable them to be widely applied in mental health. To our knowledge, this is the first scoping review to explore the applications of LLMs, summarize the process of these applications, and discover the comparative studies between humans and LLMs in mental health. We compared the processes and advantages between LLMs, non-transformer models, and humans. The superior performance of LLMs explains the increasing number of LLMs' applications in the mental health arena. Following a systematic categorization of existing studies based on their areas of application, we further delved into comparative performance studies between LLMs and human counterparts. Concluding our review, we synthesized the fine-tuning practices of LLMs for mental health, offering insights that are readily applicable to future researchers and psychiatrists.

Compared to non-transformer models and humans, LLMs demonstrate superior capabilities in information acquisition, analysis, and the generation of professional responses, particularly when addressing complex reasoning problems. These enhanced capabilities position LLMs as potent tools for the detection and prediction of mental disorders through the analysis of extensive datasets, including social media content,34-36 electronic health records,21, 37 and counseling notes.38, 39 Their application in treatment and intervention is especially noteworthy. LLMs could conceivably quickly assimilate patient clinical records, summarize treatment sessions, and support diagnoses for mental disorders. This potential not only streamlines the workflow for patients and mental health care systems but also saves considerable time and resources. Based on the information interaction and generation ability, LLMs can be instrumental in the development of chatbots designed for initial medical consultation and emotional support. Such applications are helpful in offering discreet and affordable healthcare solutions to individuals who, due to stigma or financial constraints, are reluctant to seek assistance for mental health issues.

In this scoping review, most studies applied LLMs in detecting depression and suicidal risk.40-50 These studies have demonstrated the potential of LLMs like MentaLLaMA,34 PsychBERT,51 RoBERTa,52 and ChatGPT-4,53 in detecting and identifying mental disorders from social media platforms and clinical datasets. These models have been trained to recognize depression and recommend evidence-based treatments, with some, like ChatGPT-4 and Claude, closely aligning with the perspectives of mental health professionals.20 Furthermore, the integration of linguistic features and the appending of mental disorders' background information have been shown to enhance classification performance and calibration of these LLMs.17, 23, 54, 55 LLMs perform comparably to experienced clinicians in identifying suicide ideation (SI) risk, with the addition of suicide attempt history enhancing sensitivity. ChatGPT-4 has demonstrated superior performance over ChatGPT-3.5 in recognizing SI, despite a tendency to underestimate resilience.56 Other studies showed that the usage of LLMs produced effective strategies for predicting suicidal risk with sparsely labeled data.23, 54 This efficiency in labeling and analysis of large datasets is a significant advancement over previous methods, which were often hindered by the requirement of manual annotation and consequently limited by small sample sizes or finite datasets.

Due to their efficient ability to acquire information and generate human-like language, LLMs can help clinicians deliver preliminary and timely care,14 provide treatment guidelines, augment unbalanced clinical datasets,28-30, 57 and assist in the training of professionals. LLMs have shown potential in assisting with Cognitive Behavioral Therapy (CBT) tasks, such as reframing unhelpful thoughts and identifying cognitive biases.53 Mental disorders' knowledge-enhanced pre-training schemes for LLMs, aim to reduce regional disparities in healthcare and deliver tailored diagnostic and therapeutic services.20-22 Furthermore, In the diagnosis of PTSD 15, LLM-augmented datasets have shown improved performance over original datasets, with both zero-shot and few-shot approaches outperforming the original dataset, highlighting the effectiveness of LLMs in enhancing diagnostic accuracy with minimal additional data. According to the results of performance comparisons, LLMs present similar performance as professionals and may even surpass physicians.22 These results demonstrate the potential of LLMs in clinical assistance and support. However, when exploring the quality of LLMs' query responses and educational health materials, several studies suggested that GPT-4 is currently unreliable enough for direct-consumer queries although it has good face validity.32 Therefore, the outputs by LLMs for educational health require cautious human editing and oversight. These findings underscore the growing importance of LLMs in mental health research and practice, offering new avenues for early detection, risk assessment, and intervention strategies. As these models continue to evolve, their role in mental health support is likely to expand, with a focus on enhancing accuracy, sensitivity, and ethical considerations in clinical implementation.

In the future, the application trend of LLMs in mental health is expected to continue to rise, and the aspects of their application will be broader. Initial, most studies based on the LLMs with textual data, multimodal data such as pictures, videos, and sounds also could be integrated with multimodal LLMs. Various data types might further improve the performance of mental disorders detection and identification. Several studies have explored multimodal LLMs in mental health research.58, 59 Moreover, existing studies mainly focus on depression and suicidality; more mental disorders should be investigated with LLMs, especially for rare mental disorders such as borderline personality disorder and bipolar disorder. The applications of LLMs in treatments, interventions, and preliminary care would be beneficial for these patients. Furthermore, although several studies have developed the chatbot for early mental disorders detection or intervention, more exploration should be conducted to improve the accuracy and robustness of LLMs. In addition, it is important to provide open-resource and fine-tuned LLMs for mental health, especially for low- and middle-income countries.

Although LLMs show great performance in various applications in mental health, several areas of LLMs should be improved. First, there is a need for more high-quality, diverse, and representative datasets to train LLMs for mental health applications, ensuring that the models can understand and respond to a wide range of mental health-related queries and scenarios. Second, while LLMs can generate coherent and contextually appropriate responses, they still lag behind human performance in terms of empathy and emotional intelligence, which are crucial in mental health support. Third, LLMs need to improve their ability to reason and understand the context and nuances of mental health dialogues, which often require a deep understanding of human emotions and psychological states. Finally, it is important to establish standards for privacy, safety, and ethical considerations when LLMs process sensitive personal and health information. Several studies underscore the risk of sensitive data exposure and emphasize the prevention of harmful content generation.60, 61 Future advancements depend on collaborative efforts to refine technology, develop standardized evaluations, and ensure ethical applications, aiming to realize LLMs' full potential in supporting mental health care.

Although this is the first scoping review of the applications of LLMs in mental health, several limitations should be considered. First, there is an absence of assessment of the risk of bias due to the unique nature of these included studies. Moreover, we cannot perform a meta-analysis due to the diversity of methods and tasks in the included studies. Second, with the rapid development of LLMs, the results of comparative studies would be cautious. The performance of these LLMs may have significantly improved. Third, the pre-print studies (e.g., from arXiv and medRxiv platforms) were not included in this review due to the lack of peer review, though they were published recently. Finally, we searched for studies in English and Chinese, studies in other languages were not included in this review, which might cause some bias.

Acknowledgments

The authors wish to thank the research and administrative staff at their respective institutions for their support and contributions to this study.This study was supported by the National Natural Science Foundation of China (No. 82201708, 72304040); the Striving for the First-Class, Improving Weak Links, and Highlighting Features (SIH) Key Discipline for Psychology in South China Normal University, and the National Center for Mental Health and Prevention MHSK2024B24; and the Guangdong Planning Office of Philosophy and Social Science Program (GD23YXL01).

Conflicts of Interest

None declared.

Abbreviations

LLM: Large Language Model
NLP: Natural Language Processing
OCD: obsessive-compulsive disorder
PTSD: post-traumatic stress disorder
CBT: Cognitive Behavioral Therapy
SI: suicide ideation

Data Availability

The original contributions presented in the study are included in the article Supplementary Materials, further inquiries can be directed to the corresponding author.

Fig. 1: PRISMA flowchart of studies identified for inclusion in the scoping review about applications of LLMs in mental health.

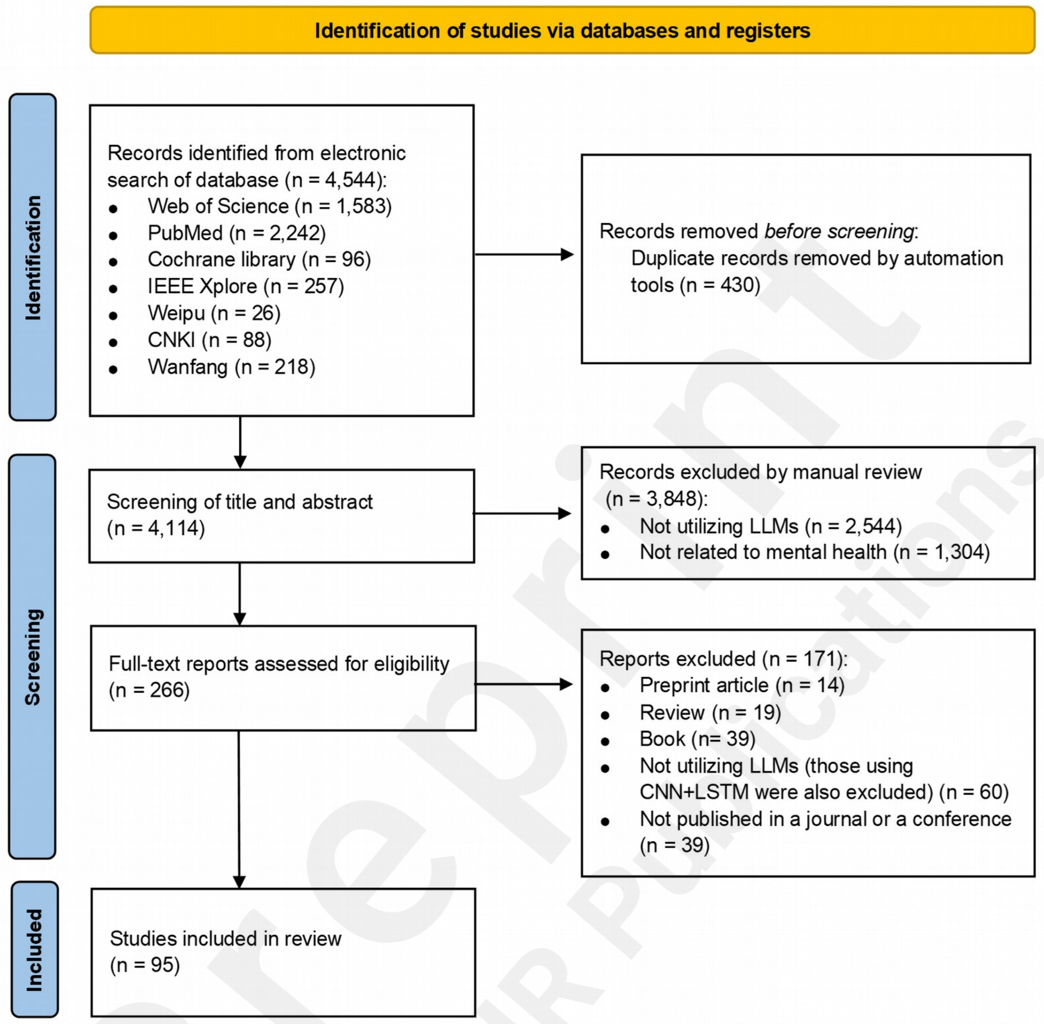

Fig. 2: The current challenges of mental health fields, comparisons between LLMs and non-transformer models, and LLMs.

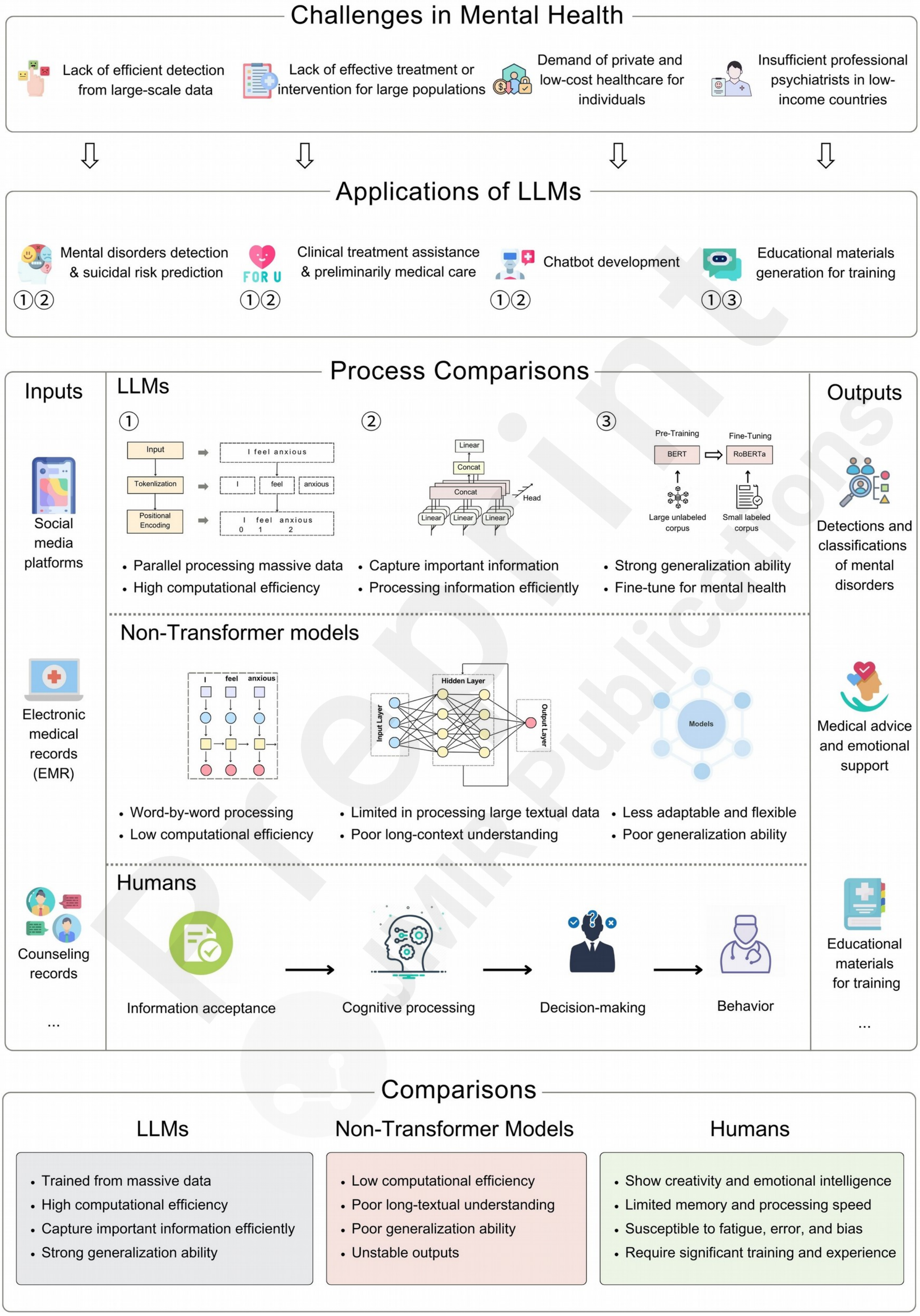

Fig. 3: Trend and applications of LLMs in mental health.
a) Number of studies mapped by countries; b) The trend of included studies published per year; c) The framework of three application categories of LLMs.

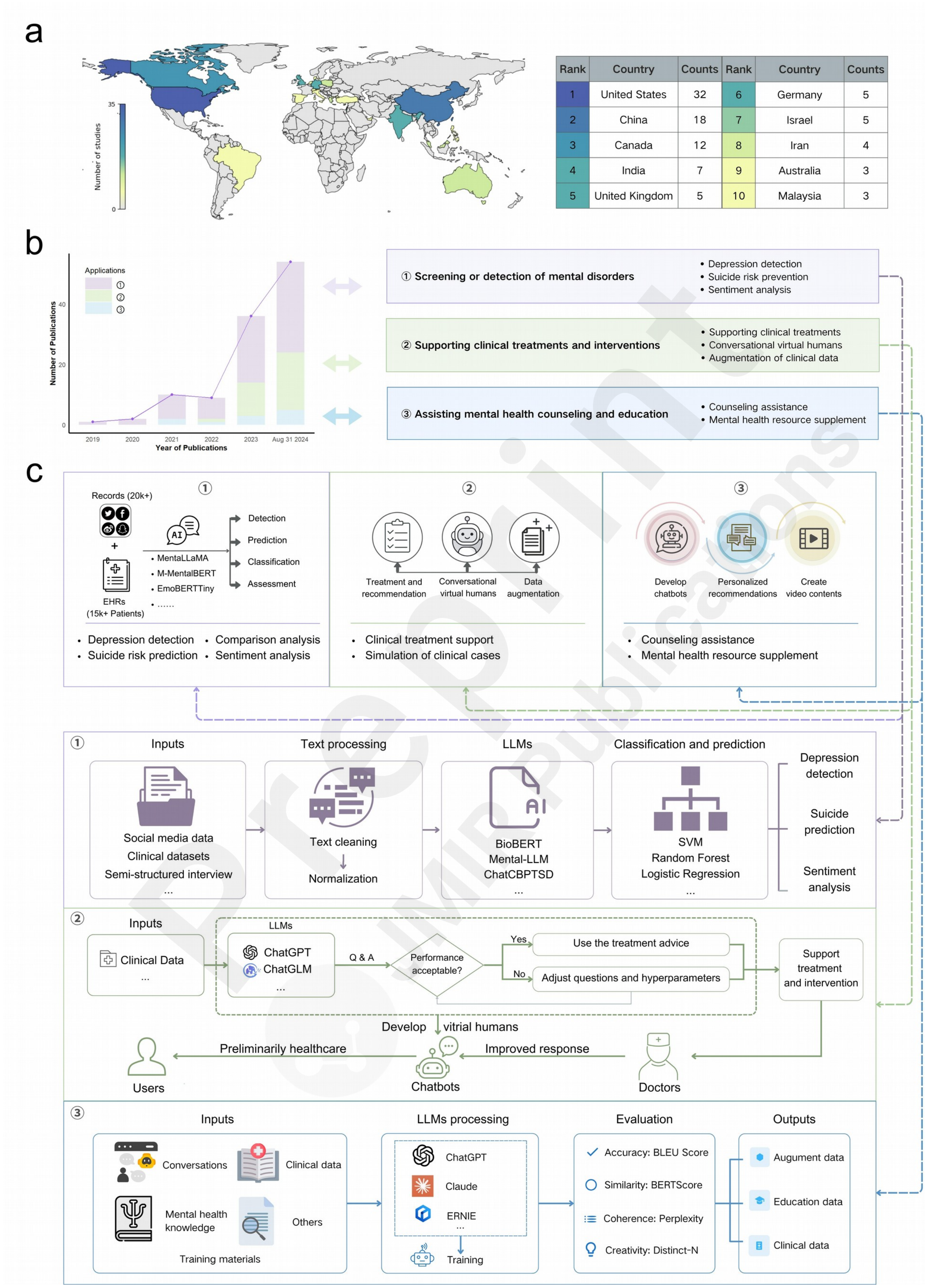

Fig. 4: Performance comparisons between LLMs and humans, and between various LLMs.

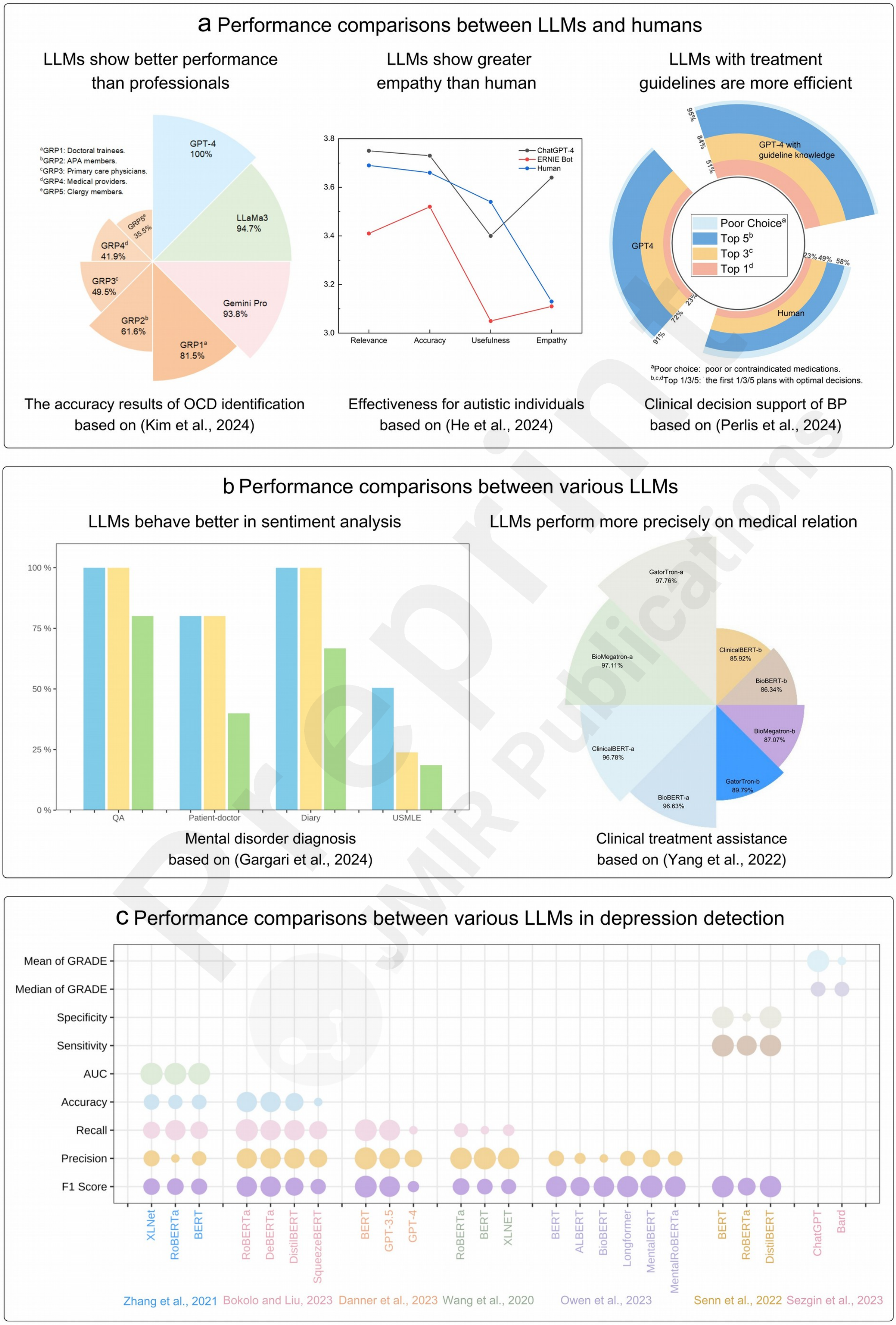

Fig. 5: The common advantages and disadvantages of LLMs.


LLMs
Common Advantages
Common Disadvantages
BERT Series
BERT
ALBERT
RoBERTa
BioBERT
DistilBERT
MentalBERT
● Symptoms classification
● Mental health detection
● Can be fine-tuned for specific mental health issues
● Large computational resources
● Generate treatment suggestions
● Limited handling of non-English languages
ChatGPT Series
GPT-3/ GPT-3.5
GPT-4
● Dialogue generation
● Contextual comprehension
● Small-sample learning
● Mental health advice accuracy
● Weak performance in specific mental disorder
LLAMA Series
LLaMA-7B
LLaMA2
● Open source
● Resource-constrained scenarios
● Interactive applications for mental health
● Complex and specific psychological tasks
Others
XLNet
ERNIE
Claude
Bard (Gemini)
● Long-term textual analysis
● Chinese psychological tasks
●Diverse dialog and text generation
● Tool integration
● Flexible application
● Large computational resources
● Structural complexity
● Limited application domains
● Limited psychological analysis capability
● Inferior performance compared to ChatGPT on certain assessments

Table 1. The basic information of included studies (n = 95).

Category	Basic information				Data information		Models
	Study	Region	Application	Mental condition	Data source	Sample information	Models
Depression detection and classification	34	Singapore	Analysis	Depression, PTSD	Reddit, SMS, Twitter, MHI dataset	105,000 data samples	MentaLLaMA
	51	United States	Detection	Depression, anxiety, SI, etc.	Reddit & Twitter	Conversations	PsychBERT
	35	United States	Detection	Depression	Twitter	Twitter users with depression/2,575 Tweets	BERT, RoBERTa, & XLNet
	62	Germany	Detection	Depression	E-DAIC	275 participants	DepRoBERTa
	20	Israel	Prognosis	Depression	Case vignettes, Previous studies	1,074 experts	GPT-3.5, GPT-4, Claude & Bard
	63	United States	Detection	Depression	DAIC-WOZ, Extended-DAIC & Simulated data	DAIC-WOZ, Extended-DAIC/122 with depression	BERT, GPT-3.5 & GPT-4
	46	China	Detection	Depression & Suicide	Dialogues from real-life scenarios	Depression (64) & Anxiety (75)	GPT-3.5
	50	Malaysia	Detection	Depression	Interviews, Facebook, Reddit & Twitter	53 participants/11 with depression	GPT-3 (ADA model)
	64	China	Prediction	Depression	Weibo	With depression labels/13,993 microblogs	BERT, RoBERTa & XLNET
	49	United States	Detection	Depression	Twitter (sentiment dataset)	632,000 tweets	RoBERTa, DeBERTa, DistilBERT &
	48	China	Detection	Depression	DAIC-WOZ	189 participants	BERT
	65	Canada	Detection	Depression	DTR dataset	Tweets: 42,691, 6,077 (DTR)	Mental-BERT
	66	United Kingdom	Detection	Depression	RSDD, RSDD-Time	Posts (9,210 depressed users)	ALBERT,BioBERT,Longformer,MentalB
	67	Switzerland	Classification	Depression	DAIC-WOZ	Respondents with depression labels	BERT, RoBERTa, DistilBERT
	36	India	Detection	Depression	Twitter	189 interviews	BERT with multimodal frameworks
	45	Malaysia	Detection	Depression	Scraped and Survey PHQ-9	250 users	BERT
	68	Netherlands	Detection	Depression	Clinical data (16,159 patients)	Survey PHQ-9, Scraped	DistilBERT
	69	Greece	Detection	Stress & Depression	Dreaddit dataset	16,159 patients	M-BERT,M-MentalBERT
	24	United States	Prediction	Stress & Depression	Dreaddit,CSSRS-Suicide, etc.	3,553 posts	Mental-Alpaca , Mental-FLAN-T5
	70	United States	Responses	PPD	ACOG, PPD	14 questions	GPT-4, LaMDA
	71	Israel	Evaluations	Depression	Clinical Vignettes	14 questions	GPT-3.5, GPT-4
	72	United States	Screening	Depression	DAIC-WOZ	Diagnosed with depression/8 vignettes	AudiBERT (I, II, III)
	73	Canada	Assessment	Depression	DAIC-WOZ	15 thematic datasets	Prefix-tuned RoBERTa
	74	India	Detection	Depression	Reddit (mental health corpus, depression)	189 subjects	RoBERTa
	43	United Arab	Detection	Depression	E-DAIC	7,650 unique entries	BERT-based custom classifier
	75	Iran	Prediction	Depression	Autodep dataset (Twitter)	219 samples & 20 real participants	DBUFS2E, BBU, MBBU & DRB
	76	United States	Classification	Depression	GLOBEM dataset	Collection of passive sensing/Tweets & bio-	GPT-3.5, GPT-4 & PaLM2
	42	China	Detection	Depression	DAIC-WOZ	Respondents with depression labels	BERT
	77	Russia	Detection	Depression	DAIC-WOZ	Respondents with depression labels	BERT, MentalBERT, MentalRoBERTa,
	78	China	Diagnosis	Depression	Labeled text data	NR	DepGPT
	47	United States	Detection	Depression	DAIC, E-DAIC, EATD	Non-depression and depression/DAIC, E-DAIC,	BERT, RoBERTa
	44	Canada	Prediction	Depression, ADHD, anxiety	Reddit	2,514 users' posts/167,444 clinical posts, 2,987,780	BERT, ROBERTA, Open AI GPT, GPT 2
	79	South Korea	Detection	Depression	Mind station app data	428 diaries	GPT-3.5, GPT-4
	17	United States	Prediction	SI	Brightside Telehealth platform	460 (SI at intake, SI later, without SI)	GPT-4

Suicide	34	Singapore	Analysis	Depression, PTSD	Reddit, SMS, Twitter, MHI dataset	105,000 data samples	MentaLLaMA
	24	United States	Prediction	Stress, depression	Dreaddit,CSSRS-Suicide, etc.	3,553 posts	Mental-Alpaca , Mental-FLAN-T5
	80	Canada	Detection	SI	UMD dataset , LLM Synthetic datasets	Over 100,000 posts & comments	BERT
	54	Canada	Classification	Suicide	Reachout.com Forum Posts, UMD Reddit dataset	Posts in /r/SuicideWatch on Reddit/1,588 labeled	GPT-1
	55	Israel	Assessment	Suicide	Professional Mental Health Assessments	4 vignettes	GPT-3.5, GPT-4
	51	United States	Detection	Depression, anxiety	Reddit & Twitter	148,700 Conversations	PsychBERT
	23	Austria	Detection	Suicide	Twitter	3,202 English tweets	BERT, XLNet
	81	Brazil	Detection	Suicidal ideation	Twitter	suicide-related text/5,699 tweets	Boamente
	82	China	Classification	Suicide	Microblog User data (ZouFan comments)	4,500 pieces	knowledge-perception BERT model
	83	United States	Detection	Suicide	Reddit (SuicideWatch section)	11 depressive & 42 /non-depressive/2.9 million posts	BERT
	40	Morocco	Detection	Suicide	Reddit	Suicide & non-suicide content/232,074 posts	GPT, BERT
Other mental disorders	60	United States	Diagnosis	OCD	Clinical Vignettes	OCD vignettes	GPT-4, Gemini Pro & Llama 3
	34	Singapore	Analysis	Depression, PTSD	Reddit, SMS, Twitter, MHI dataset	105,000 data samples	MentaLLaMA
	51	United States	Detection	Depression, anxiety	Reddit & Twitter	148,700 conversations	PsychBERT
	69	Greece	Detection	Stress & Depression	Dreaddit dataset	16,159 patients	M-BERT,M-MentalBERT
	24	United States	Prediction	Stress, depression	Dreaddit,CSSRS-Suicide, etc.	3,553 posts	Mental-Alpaca , Mental-FLAN-T5
	84	Israel	Diagnosis	BPD & SPD	Emotional scenarios (20 cases)	20 scenarios	GPT-3.5
	85	United States	Rating	Emotion	Psychotherapy transcripts, Interrater dataset	97,497 ratings	BERT
	86	China	Extraction	Psychiatric disorder	Clinical notes (12,006 records)	Human/12,006 anonymous clinical notes	BERT
	87	China	Screening	SCZ, BPD, MDD, DD	EHRs	500 EHRs	BERT, ROBERTa, DistilBERT &
	88	Iran	Diagnosis	Depression, OCD GAD,	DSM-5 based case scenarios	13 case scenarios	GPT-3.5, GPT-4, AYA & Nemotron-3-8B
	89	United States	Text analysis	Sentiment	Tweets, News	Annotated by human/47,925 tweets & news headlines	GPT3.5 Turbo, GPT4 & GPT4 Turbo
	90	United States	Sentiment	Sentiment	Multiple sources	tokens	OPT, GPT-3.5, BERT
	91	United States	Classification	Emotion and sentiment	Social media	417,423 & 1,176,509 samples	EmoBERTTiny
	92	United States	Emotion	Depression	SCPQ (Stress & Coping Process Questionnaire)	100 non-student adults	Text-davinci-003, GPT-3.5, GPT-4
	93	India	Identification	Emotion	GoEmotions dataset, Twitter dataset	27 different emotion categories/Comments & tweets	mobileBERT
	59	Iran	Diagnosis	Mental health disease	Clinical Cases	selected by a medical expert/20 cases	GPT-3.5, GPT-4, Nemotron & Aya
	94	Israel	Emotion	Emotion	RMET and LEAS	36 photos & 20 questions	GPT-4, Bard
	95	United States	Classification	Psychotherapy	Smart home images	7 different environents/10,767 images	GPT-4
	41	India	Detection	Stress, depression suicide	Reddit, Twitter	Stress, Depression, Suicide	GPT-2, GPT-Neo-125M
	52	India	Detection	Stress, anxiety	Reddit	3,553 labeled posts	RoBERTa, XLNet
	96	Canada	Screening	GAD	Prolific platform data	2,000 participants	Longformer
	61	China	Prediction	Mental disorder	Kaggle	16,950 categories/41,851 reviews	MentalBERT
	37	United States	Screening	General mental health issues	EHRs, Clinical notes	2,476,628 patients/290,482,002 clinical notes	GatorTron
	53	United Kingdom	Therapy	Anxiety	Therapist-written Thoughts	20 tasks at each of 3 stages	GPT-4, Bard
	97	United Kingdom	Questionnaires	Depression, anxiety &	C19PRC study data	2,058 adults	Sentence-BERT
	98	Australia	Diagnosis	Various psychiatric	Clinical case vignettes	100 cases	GPT-3.5

Supporting clinical treatments	99	China	Diagnosis	Depression	MedDialog, Metal Real, etc.	NR	LLaMA, ChatGLM & Alpaca
	100	United States	Analysis	Psychoactive experiences	Erowid, PDSP-Ki dataset	11,816 testimonials	BERTowid, BERTiment & CCA
	15	Canada	Diagnosis	PTSD	E-DAIC dataset, ChatGPT-generated Transcripts	severe depression & PTSD/219 participants	GPT-3.5-turbo
	20	Israel	Assessment	Depression	Case vignettes, Previous Studies	1,074 experts	GPT-3.5, GPT-4, Claude & Bard
	21	United States	Disorder	Bipolar depression	EHR-based generated data	50 sets of clinical vignettes	GPT-4
	38	China	Counseling	Stress, LGBTQ issues, etc.	Consultation Websites, Weibo, Zhihu	31 unique questions	ChatGLM, ERNIE Bota & Qianwen
	18	South Korea	Therapy	ADHD, dementia	USPTO Patent data	8,656 patients & 205 DTx patents	BERTopic, PatentSBERTa
	101	India	Prediction	Stress, anxiety	Survey dataset	41,000 entries	Gemini
	39	United States	Counseling	NR	MentalCLOUDS dataset	11,543 utterances	BART, T5, GPT Series, Phi-2,
	16	United States	Detection	CB-PTSD	Participant narratives	1,295 narratives	GPT-3.5-turbo
Chatbots	102	Spain	Emotion	General emotional states	Virtual human conversations	64 participants	GPT-3
	57	United States	Evaluation	Depression & Suicide	Human made	25 conversational agents	GPT-3.5
	20	Israel	Assessment	Depression	Case vignettes, Previous Studies	2 vignettes differed in gender	GPT-3.5, GPT-4, Claude & Bard
	43	United Arab	Detection	Depression	E-DAIC dataset	219 samples & 20 real participants/219 E-DAIC	BERT-based custom classifier
	22	China	Chatbots	Autistic	DXY platform (medical consultation data)	100 consultation samples	GPT-4, ERNIE Bot
	103	Germany	Therapy	ADHD	NR	NR	GPT3.5, GPT-4 Turbo, & Claude-3 Opus
	104	Poland	Sentiment	Mental health	CORTEX and Polish Common Crawl	Sentences & web pages	GPT-3.5
	105	United States	Detection	Loneliness & suicide	Survey data	1,006 users of Replika	Replika
Data augmentation	106	Canada	Augmentation	PTSD	E-DAIC dataset, generated data	219 interview records	CALLM, GPT-4, DistilBERT, BERT
	15	Canada	Diagnosis	PTSD	E-DAIC dataset, ChatGPT-generated Transcripts	Severe depression & PTSD/Augmented data	GPT-3.5
	99	China	Diagnosis	Depression	MedDialog, Metal Real Datasets, etc.	NR	LLaMA-7B, ChatGLM-6B, Alpaca
	80	Canada	Detection	SI	UMD dataset, LLM Synthetic datasets	Over 100,000 posts & comments	BERT
	107	China	Augmentation	Mental health	ChatGPT-generated narratives	3017 instances; 80/20 train-test split	BERT, BLOOM-560M, BLOOMZ-3B,
	108	Turkey	Generation	Various disorders	DAIC-WOZ, ChatGPT-dataset, Bard-dataset	Real patients & synthetic patients	GPT-3.5, Bard
	109	China	Generation	Psychiatry	DSM-5 diagnostic criteria	2000 records	Mistral7B-DSM-5 model
	16	United States	Detection	CB-PTSD	Participant narratives	1,295 Narratives	GPT-3.5-turbo
	62	Germany	Detection	Depression	E-DAIC dataset	275 participants	DepRoBERTa
Assisting in mental health	33	Canada	Chatbot	General mental well-being	Prompts made by author	With mindfulness experience/NR	GPT-3 based chatbots
	27	Philippines	Chatbot	Not suitable	Well-being conversations, PERMA Lexica, etc.	24,850 conversations	VHope
	28	India	Chatbot	Depression, anxiety	Reddit	Questions related to the illness/NR	CareBot
	30	Poland	Chatbot	General mental health	Empathetic Dialogues, DailyDialog datasets	DailyDialog dataset/NR	BERT
	29	United States	Chatbot	General mental health	Reddit	120 posts (2,917 user comments)	Replika
	110	United States	Generation	Depression	Psychiatric questions	4 Questions	GPT-3.5, GPT4
	58	United States	Healthcare	General mental wellbeing	NR	NR	GPT-4o
	111	United Kingdom	Measurement	General mental health	Qwell platform Therapy transcripts	254 conversations	RoBERTa, CTM
ng	105	United States	Detection	Loneliness & suicide	Survey data	1,006 users of Replika	Replika
	32	Australia	Education	Substance use	Mental health portals	"Cracks in the Ice" website	GPT-4

*	31	United States	Education	ADHD, ED	Interview data	With signs of a disorder/102 students	GPT-3

Notes: *: Assisting in mental health education; NR: Not reported; DAIC: Distress Analysis Interview Corpus; E-DAIC: Extended Distress Analysis Interview Corpus; PPD: postpartum depression; BPD: Borderline personality disorder; SPD: Schizoid personality disorder; OCD: Obsessive-compulsive disorder; BPD: Borderline personality disorder; GAD: Generalized anxiety disorder; MDD: Major depressive disorder; SCZ: Schizophrenia; CB-PTSD: Childbirth related Post-traumatic stress disorder; DTR: Depressive Tweets Repository; ADHD: Attention deficit hyperactivity disorder; ACOG: The American College of Obstetricians and Gynecologists; SI: Suicidal ideation; EHRs: Electronic Health Records; CALLM: Clinical interview data Augmentation via Large Language Models; CTM: Contextualized Topic Model; EATD: Emotional Audio-Textual Depression Corpus; CORTEX: CORpus of Translated Emotional teXts.

Table 2. Fine-tuned LLMs for mental health.

Availability	Base model	Models	Parameters	Training strategy	Year
Yes	BERT	MBBU	Unreported	Fine-tuning	2024
		BioBERT	Unreported	Unreported	2023
		MentalRoBERTa	Unreported	Unreported	2023
		PsychBERT	Unreported	Domain adaptation (DAPT)	2021
	LLaMA	MentaLLaMA	7B - 13B	IFT	2024
		ChatCounselor	7B	IFT	2023
	FLAN-T5	Mental-FLAN-T5	7B - 1700B	IFT	2024
		Mental-LLM	7B/11B	IFT	2023
	GPT	LLM-Counselors	Unreported	TFP	2024
		ChatCBPTSD	Unreported	TFP	2023
	Alpaca	Mental-Alpaca	7B - 1700B	IFT	2024
Not specified	BERT	CALLM	Unreported	IFT	2024
		EmoBERTTiny	4.4 M	IFT	2024
		M-MentalBERT	Unreported	IFT	2024
		Boamente	Unreported	IFT	2022
		AudiBERT (I, II, III)	Unreported	IFT	2021
	GPT	Psy-LLM	Unreported	TFP	2023
		CareBot	Unreported	IFT	2021

Notes: MBBU: mental-bert-base-uncased; IFT: Instruction Fine-Tuning; TFP: tuning-free promptin

References

1. Lee C, Mohebbi M, O'Callaghan E, Winsberg M. Large Language Models Versus Expert Clinicians in Crisis Prediction Among Telemental Health Patients: Comparative Study. *JMIR Ment Health.* 2024;11:e58129.
2. Perlis RH, Goldberg JF, Ostacher MJ, Schneck CD. Clinical decision support for bipolar

depression using large language models. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology.* 2024;49(9):1412-1416.

3. He W, Zhang W, Jin Y, Zhou Q, Zhang H, Xia Q. Physician Versus Large Language Model Chatbot Responses to Web-Based Questions From Autistic Patients in Chinese: Cross-Sectional Comparative Analysis. *Journal of medical Internet research.* 2024;26:e54706.
4. Levkovich I, Elyoseph Z. Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study. *JMIR Ment Health.* 2023;10:e51232.
5. Elyoseph Z, Refoua E, Asraf K, Lvovsky M, Shimoni Y, Hadar-Shoval D. Capacity of Generative AI to Interpret Human Emotions From Visual and Textual Data: Pilot Evaluation Study. *JMIR Ment Health.* 2024;11:e54369.
6. Kim J, Leonte KG, Chen ML, et al. Large language models outperform mental and medical health care professionals in identifying obsessive-compulsive disorder. *NPJ Digit Med.* 2024;7(1):193.
7. Pourkeyvan A, Safa R, Sorourkhah A. Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks. *IEEE Access.* 2024;12:28025-28035.
8. Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J. Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study. *JMIR Infodemiology.* 2021;1(1):e26769.
9. Bokolo BG, Liu Q. Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques. *ELECTRONICS.* 2023;12(21).
10. Sood P, Yang X, Wang P. Enhancing Depression Detection from Narrative Interviews Using Language Models. Paper presented at: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 5-8 Dec. 2023, 2023.
11. Firoz N, Berestneva O, Aksyonov SV. Dual Layer Cogni - Insight Deep-Mood Encoder: A Two- Tiered Approach for Depression Detection. Paper presented at: 2024 International Russian Smart Industry Conference (SmartIndustryCon); 25-29 March 2024, 2024.
12. Danner M, Hadzic B, Gerhardt S, et al. Advancing Mental Health Diagnostics: GPT-Based Method for Depression Detection. Paper presented at: 2023 62nd Annual Conference of the Society of Instrument and Control Engineers (SICE); 6-9 Sept. 2023, 2023.
13. Wang X, Chen S, Li T, et al. Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods: Content Analysis. *JMIR medical informatics.* 2020;8(7):e17958.
14. Owen D, Antypas D, Hassoulas A, Pardinas AF, Espinosa-Anke L, Collados JC. Enabling Early Health Care Intervention by Detecting Depression in Users of Web-Based Forums using Language Models: Longitudinal Analysis and Evaluation. *Jmir ai.* 2023;2:e41205.
15. Senn S, Tlachac ML, Flores R, Rundensteiner E. Ensembles of bert for depression classification. Paper presented at: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)2022.
16. Englhardt Z, Ma C, Morris ME, et al. From Classification to Clinical Insights: Towards Analyzing and Reasoning About Mobile and Behavioral Health Data With Large Language Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies -IMWUT.* 2024;8(2).
17. Sezgin E, Chekeni F, Lee J, Keim S. Clinical Accuracy of Large Language Models and

Google Search Responses to Postpartum Depression Questions: Cross-Sectional Study. *Journal of medical Internet research.* 2023;25:e49240.

18. Berrezueta-Guzman S, Kandil M, Martin-Ruiz ML, de la Cruz IP, Krusche S, Ieee Computer SOC. Exploring the Efficacy of Robotic Assistants with ChatGPT and Claude in Enhancing ADHD Therapy: Innovating Treatment Paradigms. Proceedings Paper presented at 20th International Conference on Intelligent Environments, IE 2024; 2024, 2024.
19. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *NPJ Digit Med.* 2022;5(1):194.
20. Hodson N, Williamson S. Can Large Language Models Replace Therapists? Evaluating Performance at Simple Cognitive Behavioral Therapy Tasks. *Jmir ai.* 2024;3:e52500.
21. Bartal A, Jagodnik KM, Chan SJ, Dekel S. AI and narrative embeddings detect PTSD following childbirth via birth stories. *Scientific reports.* 2024;14(1):8336.
22. Wang Y, Yu Y, Liu Y, Ma Y, Pang PC. Predicting Patients' Satisfaction With Mental Health Drug Treatment Using Their Reviews: Unified Interchangeable Model Fusion Approach. *JMIR Ment Health.* 2023;10:e49894.
23. Wang X, Liu K, Wang C. Knowledge-enhanced Pre-training large language model for depression diagnosis and treatment. Paper presented at: 2023 IEEE 9th International Conference on Cloud Computing and Intelligent Systems (CCIS); 12-13 Aug. 2023, 2023.
24. Zygadlo A. A Therapeutic Dialogue Agent for Polish Language. Paper presented at: 2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW); 28 Sept.-1 Oct. 2021, 2021.
25. Huang S, Fu F, Yang K, Zhang K, Yang F. Empowerment of Large Language Models in Psychological Counseling through Prompt Engineering. Paper presented at: 2024 IEEE 4th International Conference on Software Engineering and Artificial Intelligence (SEAI); 21-23 June 2024, 2024.
26. Gargari OK, Fatehi F, Mohammadi I, Firouzabadi SR, Shafiee A, Habibi G. Diagnostic accuracy of large language models in psychiatry. *Asian J Psychiatr.* 2024;100:104168.
27. Tanana MJ, Soma CS, Kuo PB, et al. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behavior research methods.* 2021;53(5):2069-2082.
28. Gargari OK, Habibi G, Nilchian N, Farzan AS. Comparative analysis of large language models in psychiatry and mental health: A focus on GPT, AYA, and Nemotron-3-8B - 8B. *Asian Journal of Psychiatry.* 2024;99.
29. Lossio-Ventura JA, Weger R, Lee AY, et al. A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT) Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data. *JMIR Ment Health.* 2024;11:e50150.
30. Jain B, Goyal G, Sharma M. Evaluating Emotional Detection & Classification Capabilities of GPT-2 & GPT-Neo Using Textual Data. Paper presented at: 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence); 18-19 Jan. 2024, 2024.
31. Goyal T, Rajeshbai DH, Gopalkrishna N, T M, M HR. Mobile Machine Learning Models for Emotion and Sarcasm Detection in Text: A Solution for Alexithymic Individuals. Paper presented at: 2024 3rd International Conference for Innovation in Technology (INOCON); 1-3 March 2024, 2024.

32. Stigall W, Al Hafiz Khan MA, Attota D, Nweke F, Pei Y. Large Language Models Performance Comparison of Emotion and Sentiment Classification. Proceedings of the 2024 ACM Southeast Conference on ZZZ; 2024.
33. Ghanadian H, Nejadgholi I, Osman HA. Socially Aware Synthetic Data Generation for Suicidal Ideation Detection Using Large Language Models. *IEEE Access.* 2024;12:14350-14363.
34. AygÜN i, Kaya M. Use of large language models for medical synthetic data generation in mental illness. Paper presented at: 7th IET Smart Cities Symposium (SCS 2023); 3-5 Dec. 2023, 2023.
35. Rathje S, Mirea DM, Sucholutsky I, Marjieh R, Robertson CE, Van Bavel JJ. GPT is an effective tool for multilingual psychological text analysis. *Proceedings of the National Academy of Sciences of the United States of America.* 2024;121(34):e2308950121.