e-ISSN: 2789-5661
2025: Vol5(1) pp 1-16. https://doi.org/10.38140/ijss-2025.vol5.1.06
Interdisciplinary Journal of Sociality Studies

A Scoping Review of Culturally Sensitive Large Language Models-based Cognitive Behavioural Therapy for Anxiety and Depression: Global Lessons for African Implementation

Kevin Igwe[1*]
Kevin Durrheim[2]

AFFILIATIONS
[1&2]Department of Psychology, University of Johannesburg, Johannesburg, South Africa.

CORRESPONDENCE
Email: kigwe@uj.ac.za*

EDITORIAL INFORMATION
Received: 22 November 2024
Revised: 20 March 2025
Accepted: 21 March 2025
Published: 23 May 2025

DOI: 10.38140/ijss-2025.vol5.1.06

Abstract: Anxiety and depression are significant global mental health challenges. In Africa, these conditions are critical social issues deeply connected to factors such as socio-economic disparities, cultural stigma, and limited healthcare resources. These factors create substantial barriers to effective care, highlighting the need for innovative approaches to mental health treatment. Large Language Model-based (LLM-based) Cognitive Behavioural Therapy (CBT) addresses this need by leveraging CBT's structured and effective interventions while allowing for innovative approaches to scale the intervention for these conditions. However, existing research predominantly explores LLM integration in Western contexts, with minimal focus on African cultural dynamics. This scoping review investigates the integration of culturally sensitive elements in LLM-based CBT interventions for anxiety and depression, focusing on addressing the unique considerations for African implementation. Scopus, Web of Science (WOS), EBSCO, and Google Scholar were searched to identify studies published between 2019 and 2024. The review examines global practices of integrating cultural elements into LLM-based CBT and specific considerations for implementing these interventions in Africa. Findings reveal key challenges, including limited culturally representative datasets, diverse norms, traditional beliefs, and ethical concerns. Collaboration with African researchers and communities is crucial for addressing these gaps and ensuring culturally appropriate solutions. LLM-based CBT can address Africa's mental health needs if culturally sensitive practices are prioritised. This review offers guidance for ethical, accessible, and effective interventions, combining global best practices with local insights.

Keywords: Anxiety, depression, cognitive behavioural therapy, large language model, mental health.

1. Introduction

Mental health disorders, particularly anxiety and depression, represent a significant global public health challenge, affecting over 280 million individuals worldwide (World Health Organization, 2023). In Africa, these conditions are of pressing concern, with an estimated prevalence of depression ranging between 5% and 10%, surpassing the global average of 3.8% (World Health Organization, 2023). This elevated burden is exacerbated by a severe shortage of mental health resources, with an 85% gap in service accessibility (Sodi et al., 2024). This situation necessitates innovative approaches to mental health treatment that can effectively bridge this gap while respecting and incorporating cultural contexts.

Among various treatment modalities for anxiety and depression (see Coplan et al., 2015), Cognitive Behavioural Therapy (CBT) is particularly well-suited for technological enhancement through Large Language Models (LLMs), as evidenced by Sham Sundhar et al. (2024). The effectiveness of CBT in treating anxiety and depression is well-documented across multiple contexts (Nozizwe, 2024; Twomey et al., 2015), including remote CBT (Ando et al., 2024). Hays (2009) noted, "Given CBT's emphasis on scientific analysis and quantifiable outcomes, it is not surprising that CBT has also

How to cite this article:
Igwe, K., & Durrheim, K. (2025). A scoping review of culturally sensitive large language models-based cognitive behavioural therapy for anxiety and depression: Global lessons for African implementation. *Interdisciplinary Journal of Sociality Studies*, *5*(1), a06. https://doi.org/10.38140/ijss-2025.vol5.1.06

become the most widely researched evidence-based psychotherapy" (p. 354). Its structured, step-by-step approach, with clearly defined goals and outcomes at each stage, renders it especially amenable to LLM integration. The systematic nature of CBT, which encompasses specific phases such as assessment, cognitive restructuring, and behavioural activation, provides clear points for technological intervention while maintaining therapeutic integrity.

Recent advances in LLMs, such as GPT-4 and Gemini, have demonstrated promising potential in enhancing the delivery of CBT (Obasa, 2024). The structured nature of CBT aligns well with LLMs' capabilities to process and generate context-aware responses, making it an ideal candidate for technological augmentation (Sham Sundhar et al., 2024). This notion was further reinforced by Stade et al. (2024), who proposed various methods for integrating LLMs into behavioural healthcare. For instance, Stade et al. (2024) highlighted that a collaborative LLM can be utilised to deliver structured psychotherapeutic interventions. This alignment is particularly significant because: (i) CBT adheres to a systematic, evidence-based protocol that LLMs can effectively model; (ii) the therapy's structured steps allow for clear integration points for technological support; (iii) CBT's focus on identifiable thought patterns and behaviours corresponds well with LLM processing capabilities; and (iv) the therapy's measurable outcomes facilitate effective evaluation of LLM integration (S. Lee et al., 2024; Sham Sundhar et al., 2024).

Nevertheless, implementing LLM-based CBT interventions presents various challenges and concerns. These issues include ethical, privacy, and security considerations, which also pertain to LLMs and other AI tools in mental health practices (Baguma et al., 2023; De Choudhury et al., 2023). This review centres on the challenges of integrating cultural context, particularly in settings where indigenous knowledge systems play a vital role in understanding and treating mental health conditions. For example, LLM frameworks are trained using internet data, which may not adequately represent diverse knowledge systems and healing practices (Baguma et al., 2023). This limitation raises important questions regarding how to effectively incorporate indigenous knowledge and healing practices into LLM-based CBT to facilitate culturally sensitive interventions.

1.1 Related reviews and knowledge gap

Previous reviews have focused on a wide-ranging evaluation of LLMs in mental health care, exploring diverse applications, benefits, and challenges on a global scale (Guo et al., 2024; Hua et al., 2024). For example, Guo et al. (2024) conducted a broad review of AI applications in mental healthcare, concentrating on general therapeutic applications without specific attention to CBT or cultural integration. Their key focus areas included LLM applications in early screening and digital interventions. The review concludes that while LLMs are effective in detecting mental health conditions and providing tailored interventions, ethical implications, biases, and the "black box" nature of LLMs require further attention. Hua et al. (2024) explored the technical aspects of LLMs in mental health support but did not address the integration of indigenous knowledge. While Phiri and Munoriyarwa (2023) examined AI applications in African mental health contexts, their review concentrated on general mental health technologies rather than specific therapeutic modalities. More specifically related to CBT, Jiang et al. (2024) provided an overview of AI integration into CBT practices; however, their review focused primarily on technological implementation without addressing cultural adaptation or context. Ahmed et al. (2023) examined conventional (e.g., rule-based) chatbot features, not cultural integration in LLM-based CBT for anxiety and depression.

The importance of culturally responsive CBT has been a topic of discussion for over a decade (Hays, 2009; Hinton & Patel, 2017). Hays (2009) suggested ten ways to integrate diverse cultural considerations into CBT. These include emphasising (i) respectful behaviours that resonate with the client's cultural norms, (ii) identifying and incorporating culturally related strengths and support, such as personal pride, spiritual beliefs, and community connections, and (iii) understanding the client's experiences within systems of privilege and oppression. For example, the greeting "Hi" may be considered disrespectful in some African and Asian countries but not in Western countries (Hays, 2009). The discussion on culturally responsive CBT has continued among proponents of CBT (Huey Jr et al., 2023; Jalal et al., 2020), especially in Africa (Nozizwe, 2024). According to Nozizwe (2024), culturally adapted CBT interventions have been shown to improve anxiety symptoms and hold promise for scaling CBT interventions. Consequently, researchers are increasingly discussing the integration of cultural elements into LLMs for mental health practice (Aleem et al., 2024), including LLM-based CBT to produce responses that are domain-specific (Liu et al., 2023) and culturally resonant with the client.

Although LLM-based CBT holds promise for delivering and scaling culturally adapted CBT interventions, their effectiveness, and that of LLMs in general, depends on how well they can integrate domain-specific knowledge (Ling et al., 2023). Generally, there are three primary approaches for knowledge integration in LLMs: finetuning, prompting, and Retrieval-Augmented Generation (RAG) (Patil & Gudivada, 2024). Each method offers distinct advantages and faces unique challenges, particularly when dealing with data constraints. Finetuning adapts pre-trained models to specific domains or tasks through additional training on specialised datasets (Anisuzzaman et al., 2024, 2025). However, finetuning faces significant challenges when comprehensive training data are unavailable. For example, data scarcity can lead to overfitting, where the LLM overly emphasises a specific context (Raiaan et al., 2024). Domain-specific terminology and concepts may be inadequately captured, making quality control crucial with small datasets. Prompting (Shah et al., 2024) serves as an alternative to overcome data limitations. It requires no training data and allows real-time adaptation; however, it is less reliable than finetuning for complex tasks such as CBT, which require a personalised approach. Furthermore, results can be inconsistent across different prompts, as performance heavily depends on prompt engineering expertise (Zaghir et al., 2024). Retrieval-augmented generation (RAG) offers a middle ground by dynamically incorporating external knowledge, thus reducing hallucination. Hallucination is the convincing presentation of false information as truth. RAG allows for regular knowledge updates without retraining and works well with limited domain-specific training data. However, the technical and resource requirements for implementing RAG can be challenging for low-income countries.

Despite discussions on culturally responsive or adapted CBT over the past decade (Hays, 2009; Hinton & Patel, 2017; Huey Jr et al., 2023; Nozizwe, 2024) and various avenues for integrating domain-specific knowledge into LLMs, culturally responsive or adapted LLM-based CBT remains scarce and fragmented, necessitating synthesis. This scoping review addresses this critical gap in the existing literature by:

- Focusing specifically on LLM integration within CBT frameworks
- Synthesise literature on cultural adaptation approaches in LLM-based CBT delivery
- Analysing global best practices for potential African implementation

1.2 Purpose and research questions

This scoping review aims to systematically map and analyse the existing evidence concerning the integration of culturally sensitive contexts in LLM-based cognitive-behavioural therapy (CBT) for anxiety and depression. Our investigation addresses four key research questions: *How are LLMs being integrated with cultural elements in CBT interventions globally? What methods are employed to ensure cultural sensitivity in LLM-based CBT for anxiety and depression? What specific considerations must be addressed when implementing culturally sensitive LLM-based CBT in Africa? What infrastructure and support systems are needed to facilitate the successful implementation of LLM-based CBT in African settings?* To answer these questions, we pursue three primary objectives:

- Map the current evidence on LLM-based CBT for anxiety and depression globally.
- Identify and analyse successful strategies for cultural integration in LLM-based CBT interventions.

- Synthesise recommendations for implementing culturally appropriate LLM-based CBT in Africa.

Through this scoping review, we examined global experiences and their implications for implementation in Africa. Our aim is to contribute to the development of culturally appropriate and effective LLM-based CBT interventions to address the significant mental health challenges in Africa.

2. Methodology

This scoping review adhered to the minimum requirements outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018). The process is illustrated in Figure 1 and described in detail in the following subsections.

A comprehensive literature search was conducted on July 21, 2024, across the following databases: Scopus, Web of Science (WOS), EBSCO, and Google Scholar (for grey literature). The search strategy utilised Boolean operators to combine terms related to both the target mental health conditions and large language models (LLMs) in the context of cognitive behavioural therapy (CBT). The search terms included combinations of the following keywords to refine the search for relevant studies: ("depressed" OR "depression" OR "anxiety") AND ("large language model*" OR "LLMs" OR "GPT*" OR "ChatGPT" OR "BERT" OR "Transformer" OR "LaMDA" OR "PaLM" OR "Claude" OR "Gemini" OR "BLOOM" OR "LLaMA") AND ("Cognitive Behavioural Therapy" OR "CBT"). Although this review focuses on cultural integration, we employed broader search terms that did not include "Culture" and "Africa" for two reasons: (i) some studies did not explicitly mention cultural context and (ii) to avoid restricting the results to Africa or omitting papers that mentioned a specific country in Africa without explicitly mentioning Africa. The process is illustrated in Figure 1 and described in detail in the following subsections.

2.1 Inclusion and exclusion criteria

The inclusion criteria for the studies are as follows: (i) they must be published in peer-reviewed journals, conference proceedings, or preprint platforms between 2019 and July 2024, and (ii) they must investigate the use of LLM-based CBT interventions specifically targeting anxiety and/or depression. We selected literature from 2019 onwards because Bidirectional Encoder Representations from Transformers (BERT), the foundational language representation model for LLMs, was introduced in 2018 (Devlin, 2018) and gained popularity in 2019 (Kenton & Toutanova, 2019). A broad spectrum of study designs was considered, including randomised controlled trials, observational studies, and qualitative research. Studies were excluded if they were review articles, did not primarily focus on anxiety or depression, or examined interventions unrelated to LLMs or CBT. Additionally, studies addressing mental health disorders outside of anxiety and depression were also excluded.

2.2 Study selection

The initial database search yielded a total of 106 studies, distributed as follows: 4 from WOS, 15 from Scopus, 57 from EBSCO (which includes databases listed in Table 1), and 30 from Google Scholar (from an initial search hit of 119). As shown in Figure 1, duplicates (n = 32) were removed from EBSCO, and an additional (n = 5) duplicates were identified and excluded after all identified studies were combined in the EndNote reference management software. The studies were then exported to Rayyan (Ouzzani et al., 2016), where the first author, with the assistance of a librarian, conducted the screening of titles and abstracts according to the established inclusion and exclusion criteria. The screening was independently validated by the second author. Full-text articles (n = 62) were retrieved for studies that met the criteria or where eligibility was unclear. Disagreements between the reviewers were resolved through discussion and consensus. Ultimately, 31 studies were included in the review.

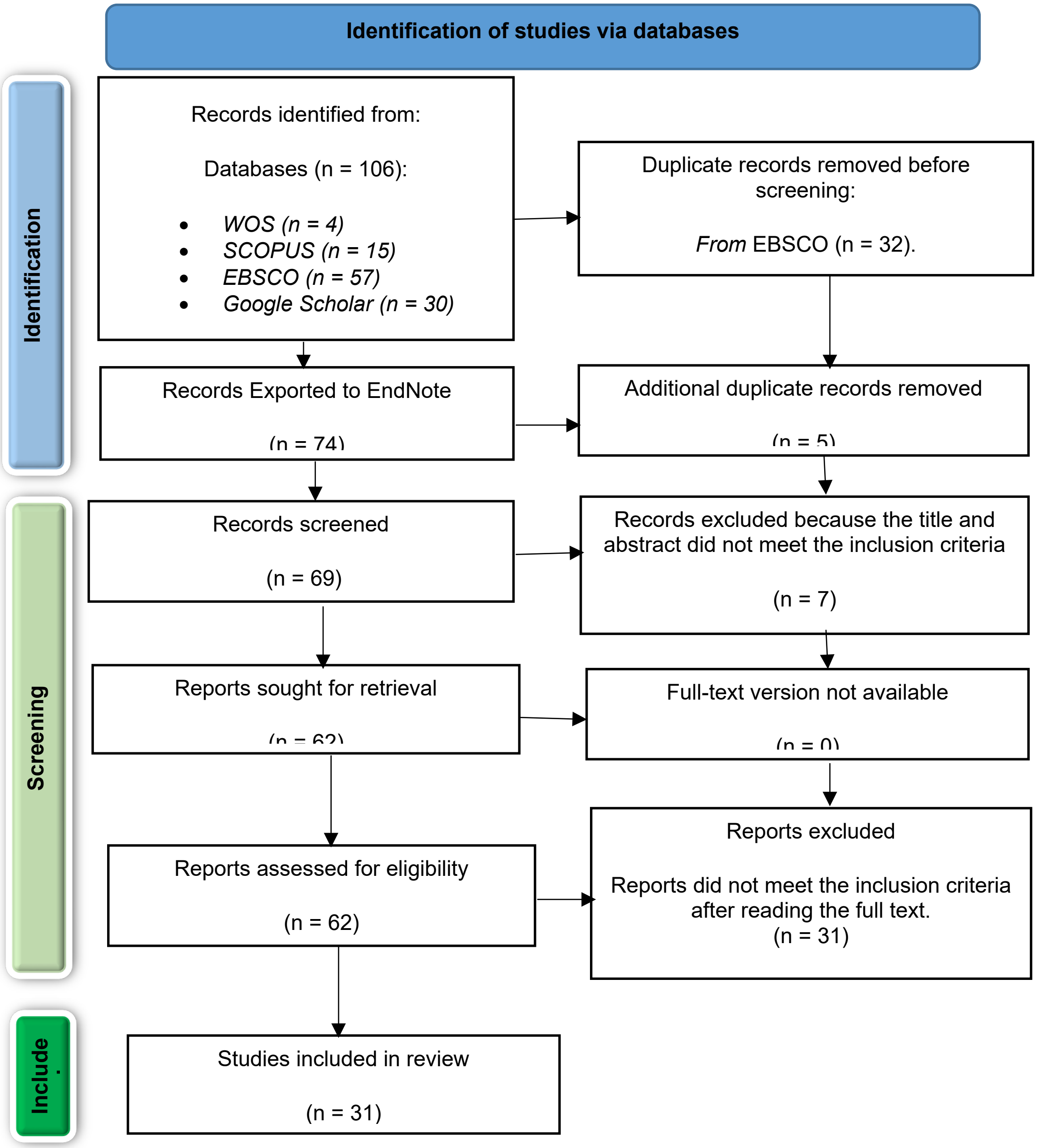

Figure 1: *PRISMA flow diagram*

Table 1: *Databases hosted by EBSCO, where records were found (with duplicates)*

Databases	Number of Records
MEDLINE	n = 14
APA PsycInfo	n = 13
Academic Search Ultimate	n = 12

E-Journals	n = 4
Family & Society Studies Worldwide	n = 3
Inspec	n = 3
Health Source: Nursing/Academic Edition	n = 2
APA PsycArticles	n = 1
CINAHL with Full Text	n = 1
MasterFILE Premier	n = 1
OpenDissertations	n = 1
Sociology Source Ultimate	n =1
Women's Studies International	n =1
	n =57

2.3 Data extraction, synthesis and analysis

Data were extracted from each included study based on the following key areas: Study Characteristics (author(s), year of publication, country/region, and study design), Study Target (mental health conditions addressed, such as anxiety and depression), and Integration of Culturally Sensitive LLMs-based CBT (how LLMs were integrated into CBT interventions to enhance cultural sensitivity). Due to the heterogeneity of study designs, interventions, and reported outcomes, a narrative synthesis approach (Snilstveit et al., 2012) was adopted to summarise the findings. The synthesis focused on identifying key methods, application stages, and challenges associated with the use of culturally sensitive LLMs in CBT interventions for anxiety and depression.

As this study involved the review of existing literature, formal ethical approval was not required. The review was conducted in accordance with ethical research practices, ensuring the accurate representation of original study findings and appropriate citation of all sources.

3. Presentation of Results

This scoping review provides an understanding of how countries and geographical regions are involved in studies integrating LLMs into CBT for anxiety and depression. It addresses key research questions, including how LLMs are integrated or used to incorporate indigenous knowledge, and highlights the stages in the CBT process where LLMs are applied.

3.1 Region, country and study type

The studies reviewed were conducted in a diverse range of countries, reflecting a global interest in exploring the role of LLMs in CBT. We identified the country of study based on the authors' affiliations. According to the available literature, the United States (USA) led with the highest number of studies (9), as shown in Figure 2, followed by China with 7. India and South Korea each contributed four studies, demonstrating significant research engagement in these regions.

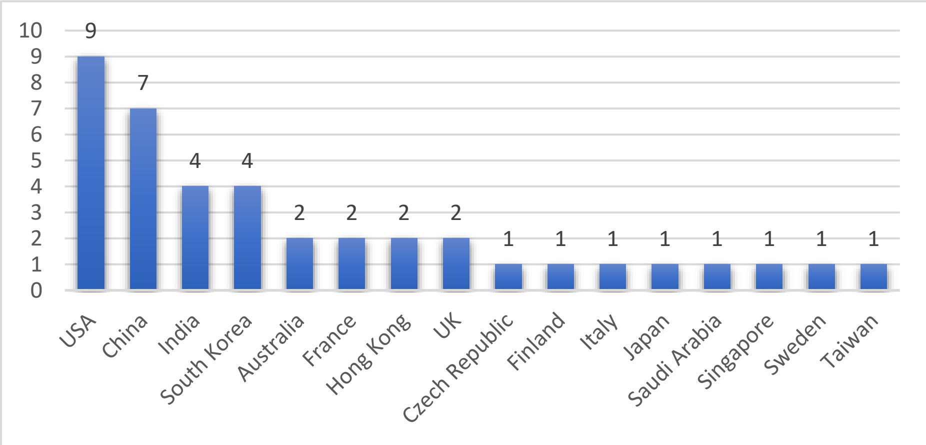

Figure 2: Countries where studies were conducted

Australia, France, the United Kingdom (UK), and Hong Kong each accounted for two studies, highlighting a growing interest in these countries. Other nations, such as Finland, Sweden, Saudi Arabia, Italy, Japan, Taiwan, the Czech Republic, and Singapore, each contributed one study, signalling emerging research efforts in LLM-based CBT interventions. One study was categorised under "Unknown," indicating that the country affiliation of the authors or the study's location could not be identified.

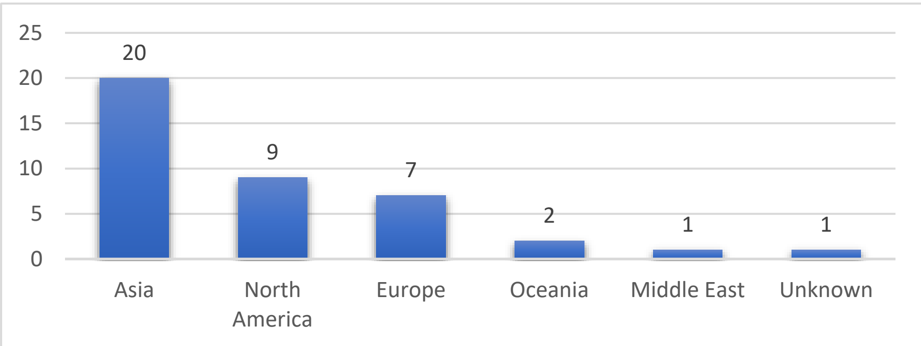

Figure 3: Regions where studies were conducted

Figure 2 shows that while the USA and China are at the forefront of research in this domain, there is also a notable presence from countries across Asia, Europe, and other regions (see Figure 3). However, some regions, particularly Africa, were not represented in the reviewed studies, indicating a gap in the research landscape. This highlights opportunities for expanding the scope of research, fostering global collaboration, and ensuring that LLM-based CBT interventions are explored more broadly, especially in underrepresented areas.

3.2 Strategies for cultural integration in LLM-based CBT interventions

3.2.1 Key Methods of integrating cultural content in LLM-based CBT for Anxiety and Depression

The methods of integrating specific context or cultural content into LLMs for CBT have been approached in several ways, as shown in Table 2. These methods, which aim to enhance the therapeutic effectiveness of LLMs, include the use of curated datasets that reflect different cultural perspectives (Na, 2024; Schiff, 2024); fine-tuning models based on user-specific contexts and/or cultural identities (Abubakar et al., 2024; Adhikary et al., 2024; Agrawal & Gupta, 2024; Izumi et al., 2024; S. Lee et al., 2024; Na, 2024; Nie et al., 2024; Zhang et al., 2024); and incorporating feedback mechanisms that allow for continuous adjustment of responses (Abubakar et al., 2024). Additionally, some research (Gabriel et al., 2024; Izumi et al., 2024; Na, 2024) has focused on training LLMs to recognise specific cognitive distortions and adapt therapeutic techniques like Socratic questioning to align with cultural norms. We discuss the details of each method in the following subsections. However, it is noteworthy that two or more of these methods can be combined to improve the effectiveness of the intervention, as seen in Na (2024).

Table 2: *Summarised methods implemented in various papers for the use/integration of LLMs in CBT*

Key methods used for the effectiveness of LLM-Based CBT	A brief explanation of the category	Example of Studies
Inclusion of Cultural Sensitivity Training Data	Curated datasets that encompass a wide range of cultural contexts	(Na, 2024); (Schiff, 2024)
Contextual Fine-Tuning	Using data specific to a targeted therapeutic context	(Abubakar et al., 2024); (Adhikary et al., 2024); (Agrawal & Gupta, 2024); (Izumi et al., 2024); (S. Lee et al., 2024); (Na, 2024); (Nie et al., 2024); (Zhang et al., 2024)
Feedback and adjustment mechanisms	Allow users to report responses they find culturally insensitive	(Abubakar et al., 2024); External: (Chiu et al., 2024); (Kian et al., 2024); (Y. K. Lee et al., 2024); (Zhang et al., 2024)
Handling specific cultural or contextual cognitive distortions	Questioning techniques to reflect individual beliefs	(Gabriel et al., 2024); (Na, 2024); (Izumi et al., 2024)

- ***Inclusion of Cultural Sensitivity Training Data***: Research in this category trained LLMs on extensive, curated datasets that encompassed a wide range of cultural contexts, including mental health perspectives and cognitive distortions unique to specific cultures. This method involved generating a dataset that is semantically equivalent to a Western-dominated training dataset, aiming to reduce bias that could arise from such datasets used to train the foundation model. For instance, Na (2024) addressed the challenges of cultural sensitivity and data quality in mental health interventions by utilising the PsyQA dataset (Sun et al., 2021), which is specifically designed for Chinese mental health question-answering tasks. Na (2024) employed prompts grounded in CBT principles to ensure that the dataset follows structured, professional therapeutic strategies. The author instructed ChatGPT to generate responses incorporating CBT components for each question and description in the PsyQA dataset.
- ***Contextual Fine-Tuning:*** This involves fine-tuning a large language model (LLM) using a dataset crafted by an expert, such as a counselling psychologist, to guide the LLM's responses in a targeted context. This method enables LLMs to dynamically adjust their responses based on user input while maintaining a specific context. For example, Schiff (2024) fine-tuned Meta’s LLaMA-3 8B model using scenarios generated by the AI assistant Claude. The goal was to address specific cognitive distortions, such as all-or-nothing thinking and overgeneralisation. Similarly, Agrawal and Gupta (2024) fine-tuned the GPT, LLaMA, and Gemini models using few-shot prompting based on criteria specified in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), to diagnose depression and deliver CBT-based therapeutic interventions. Their approach ensured empathetic, contextually relevant interactions. Both studies demonstrated how fine-tuning can improve the models’ ability to detect, diagnose, and treat various cognitive and emotional disorders using CBT methods.

- ***Feedback and adjustment mechanisms:*** These mechanisms allow users to report responses they find culturally insensitive. Learning algorithms are implemented that adjust future responses based on this feedback. For example, Abubakar et al. (2024) applied a three-stage method for developing a mental health therapy chatbot using Reinforcement Learning from Human Feedback (RLHF) to optimise chatbot responses. The process begins with supervised fine-tuning of an LLaMa model to generate appropriate responses, followed by training a reward model to evaluate response quality. The final stage employs reinforcement learning techniques to optimise the model while maintaining consistency with its initial training. The model was trained using a dataset of counsellor-client messages from the SNAP Counselling Conversation Analysis dataset (Althoff et al., 2016), covering various mental health issues, including depression and anxiety. Thus, Abubakar et al. (2024) combined both automated learning and human feedback to create a chatbot capable of providing contextually appropriate therapeutic responses. The method emphasises maintaining high-quality interactions while allowing for continuous improvement based on human feedback. Others collect feedback from users via questionnaires or surveys to improve LLMs for future use.
- ***Socratic question generation and adaptation***. This method involves adapting the LLM's questioning techniques to reflect individual beliefs. For example, Izumi et al. (2024) examined the integration of Socratic questioning techniques with LLMs in CBT dialogue scenarios. The researchers implemented three distinct approaches: traditional Socratic questioning, pure LLM-generated responses (using either OsakaED or GPT-4), and a hybrid combining both methods. The Socratic dialogue was designed to help clients objectively examine and reconsider their automatic thoughts. The traditional Socratic questions followed a structured progression, prompting clients to evaluate the evidence for and against their automatic thoughts. This approach allows them to consider alternative perspectives and assess the worst, best, and most realistic outcomes.

3.3 Categorisation of methods of usage vs integration of LLM for CBT

Understanding the distinction between LLM usage and integration in CBT is crucial for maximising its effectiveness. While usage typically involves employing LLMs as support tools for tasks like session summaries and question generation, integration involves the incorporation of LLMs into the therapeutic process itself. Integration positions LLMs as active participants in the CBT journey, contributing to real-time cognitive distortion identification, reframing exercises, and guided discovery processes. Thus, integration enables dynamic therapeutic adjustments based on patient responses. Although usage and integration are similar concepts, the key difference between the two lies in the depth of involvement. Integration requires modification of the LLM, while LLMs may be used without modifying the pre-trained model. For example, Jiang et al. (2024) used GPT-4 to summarise posts and identify cognitive distortions, while Xiao et al. (2024) utilised the "HealMe" model to assist in cognitive reframing. Abubakar et al. (2024) integrated RLHF into a chatbot to optimise real-time therapy sessions. Similarly, Kian et al. (2024) combined an LLM with a socially assistive robot to guide patients through CBT exercises, providing real-time interaction and feedback.

3.4 Stages of CBT and LLM usage or integration

Cognitive Behavioural Therapy (CBT) typically follows a structured framework comprising several key stages: Assessment, Identification of Cognitive Distortions, Cognitive Restructuring, Behavioural Activation, and Maintenance and Relapse Prevention. Each stage targets specific therapeutic objectives, and recent studies have explored how Large Language Models (LLMs) can be integrated into these stages to enhance therapeutic processes.

During the Assessment stage, therapists aim to develop an initial understanding of a patient's emotional and mental state. Studies such as Adhikary et al. (2024) have demonstrated the use of LLMs—specifically Mental Llama and Mistral—for summarising therapy sessions. These models support therapists by rapidly synthesising key information from session transcripts, thereby assisting in the early identification of cognitive distortions. In the Identification of Cognitive Distortions stage, LLMs play a crucial role by detecting unhelpful or irrational thought patterns embedded in clients' narratives. Research by Jiang et al. (2024) and Schiff (2024) employed models like GPT-4 and CBTLlama to extract cognitive distortions from textual data, providing a valuable diagnostic tool to inform subsequent interventions.

The Cognitive Restructuring stage involves helping clients to reframe and challenge distorted thinking. Xiao et al. (2024) and Izumi et al. (2024) illustrated how LLMs could effectively guide clients through this process by delivering empathetic and logic-driven responses. These models also utilise Socratic questioning techniques to prompt clients to critically examine and re-evaluate maladaptive thoughts. In the Behavioural Activation phase, clients are encouraged to engage in positive, goal-directed activities to counter symptoms of anxiety and depression. Kian et al. (2024) integrated LLMs into a Socially Assistive Robot to facilitate CBT exercises. This approach successfully promoted behavioural engagement and yielded measurable improvements in psychological distress. Finally, in the Maintenance and Relapse Prevention stage, LLMs offer tools for sustained therapeutic support. Nie et al. (2024) developed CaiTI, a Conversational AI Therapist, which employs LLMs to monitor clients' daily functioning and deliver personalised interventions. This ongoing support plays a critical role in maintaining therapeutic gains and preventing relapse post-treatment.

4. Discussion

One critical gap identified in the literature is the underrepresentation of African literature in LLM-based CBT. Current implementations predominantly focus on Western and Chinese contexts, with no integration of culturally specific African contexts into LLM-based CBT. However, this review presents an opportunity to learn from global practices and address this gap. The review highlighted four primary methods for incorporating cultural context into LLM-based CBT.: (i) curating culturally contextualised datasets from existing datasets to achieve semantic equivalence but syntactic variation; (ii) utilising expert-generated data grounded in indigenous knowledge and contextual applications of CBT within specific settings; (iii) employing adaptive techniques such as RLHF to enable LLMs to learn and adapt to users' contexts and belief systems; and (iv) prompting LLMs to specialise in particular aspects of CBT, such as Socratic questioning or specific stages within the therapeutic process. We present some challenges and suggestions to facilitate African implementation.

4.1 Recommendations for implementing culturally appropriate LLM-based CBT

Implementing large language model-based cognitive behavioural therapy (LLM-based CBT) in the African context presents several culturally specific challenges that demand nuanced, context-aware solutions. The African continent is marked by strong traditional and spiritual beliefs about mental health, collective social structures that shape therapy engagement, limited technological infrastructure in many areas, diverse linguistic and cultural realities, and socio-economic conditions that influence how therapy is received and practised. These factors necessitate a multi-integrative approach to LLM implementation—particularly important given the scarcity of culturally appropriate training datasets.

A primary challenge is the limited availability of culturally sensitive training data. As Baguma et al. (2023) note, African languages and cultural expressions are predominantly oral, leaving them underrepresented in digital formats and, consequently, in the datasets used to train LLMs. To address this, African researchers must focus on contextual fine-tuning and the incorporation of feedback mechanisms that can embed traditional and spiritual belief systems into therapeutic content. Collective social structures also play a central role in therapy outcomes. Research by Jameel et al. (2022) and Kunorubwe (2023) shows that individuals from collectivist cultures often value family involvement in therapy decisions, requiring LLMs to be culturally attuned to these dynamics. To support this, there is a need for curated datasets that reflect African communal structures, alongside a call for bibliometric mapping to foster collaboration among African scholars working at the intersection of mental health and AI.

Another major hurdle is the problem of hallucinations—instances where LLMs, such as GPT-4, generate factually incorrect or misleading information. Jiang et al. (2024) emphasise the risks associated with hallucinations in therapeutic contexts, where factual accuracy and emotional reliability are paramount. This underscores the critical need for continual model refinement and oversight. Despite advances in fine-tuning LLMs for therapeutic use, human validation remains indispensable. Schiff (2024) argues that responses generated by LLMs must be reviewed by qualified therapists to ensure they adhere to professional clinical standards and maintain therapeutic integrity. Without this validation, there is a risk of divergence from evidence-based CBT practices.

One strategic pathway forward is to develop an African-informed framework for LLM-based CBT grounded in local cultural norms and values. While a standardised framework is yet to be developed, the concept of culturally grounded and culturally adapted interventions offers a valuable foundation. Culturally grounded approaches align interventions with a client's worldview and belief systems (Anakwenze, 2022; Caloudas et al., 2024), while culturally adapted approaches involve active collaboration with community leaders to ensure interventions are socially resonant and contextually valid (Anakwenze, 2022; Spanhel et al., 2021). These approaches can be operationalised in LLMs through combined methods such as prompting, fine-tuning, and iterative validation involving local communities.

Ethical concerns must also be addressed to ensure the safe and just application of LLM-based CBT. Studies such as Gabriel et al. (2024) reveal that LLMs may exhibit uneven empathy responses across demographic groups, raising questions about fairness and bias. This is particularly dangerous in high-stakes situations, such as suicidal ideation, where safety and trust are paramount. Manvi et al. (2024) warn that without representative training data, LLMs risk reinforcing existing stereotypes and disseminating culturally inaccurate or harmful information. These issues point to the broader ethical imperative of promoting fairness, respect, and equity in AI-based mental health interventions.

Thus, the African implementation of LLM-based CBT requires more than just technological adaptation; it demands cultural legitimacy, ethical reflexivity, and collaborative innovation. Solutions lie in expanding culturally sensitive data sources, reducing model hallucinations, embedding human oversight, and rooting interventions in community-based cultural validation. Doing so will not only enhance therapeutic relevance but also help counter the embedded inequalities within global AI systems.

5. Conclusion

LLM-based CBT holds promise for the effectiveness and support of CBT within the African context due to LLMs' ability to understand, process, and generate large volumes of text. Many regions have begun exploring this potential. Despite the promise and fast-evolving implementation in other areas, adoption within African contexts remains scarce. We identify dataset challenges, diverse African cultures and norms, traditional and spiritual beliefs, and ethical concerns, among others, as key factors that must be addressed to promote implementation in Africa. We recommend bibliometric analysis as future work to identify African researchers in this niche area, which would facilitate collaboration among these researchers and is necessary for competing globally.

6. Declarations

Author contributions: Conceptualisation (K.I. & K.D.); literature review (K.I.); methodology (K.I. & K.D.); software (N/A.); validation (K.D.); formal analysis (K.I.); investigation (K.I. & K.D.); data

curation (N/A); drafting and preparation (K.I. & K.D.); review and editing (K.I. & K.D.); supervision (K.D.); project administration (K.I.); funding acquisition (N/A). All authors have read and approved the published version of the article.

Funding: This research receives no external funding.

Acknowledgement: The authors declare no acknowledgement.

Conflict of interest: The authors have no conflicts of interest to disclose.

Data availability statement: This review is based entirely on publicly available data and information sourced from peer-reviewed articles, reports, and other academic publications cited in the manuscript. No new primary data was generated or analysed during this study. Readers may refer to the cited sources for detailed information.

References

Abubakar, A. M., Gupta, D., & Parida, S. (2024). A Reinforcement Learning Approach for Intelligent Conversational Chatbot For Enhancing Mental Health Therapy. *Procedia Computer Science*, *235*, 916-925. https://doi.org/10.1016/j.procs.2024.04.087

Adhikary, P. K., Srivastava, A., Kumar, S., Singh, S. M., Manuja, P., Gopinath, J. K., Krishnan, V., Gupta, S. K., Deb, K. S., & Chakraborty, T. (2024). Exploring the efficacy of large language models in summarising mental health counselling sessions: Benchmark study. *JMIR Mental Health*, *11*, e57306. https://doi.org/10.2196/57306

Agrawal, A., & Gupta, N. (2024). Illuminate: Depression diagnosis, explanation, and proactive therapy using prompt engineering. *Journal of Student Research*, *13*(2). https://doi.org/10.47611/jsrhs.v13i2.6718

Ahmed, A., Hassan, A., Aziz, S., Abd-Alrazaq, A. A., Ali, N., Alzubaidi, M., Al-Thani, D., Elhusein, B., Siddig, M. A., & Ahmed, M. (2023). Chatbot features for anxiety and depression: A scoping review. *Health Informatics Journal*, *29*(1), 1-17. https://doi.org/10.1177/14604582221146719

Aleem, M., Zahoor, I., & Naseem, M. (2024, November). Towards culturally adaptive large language models in mental health: Using ChatGPT as a case study. In *Companion publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing* (pp. 240–247). ACM Digital Library.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counselling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463–476. https://doi.org/10.1162/tacl_a_00111

Anakwenze, O. (2022). The cultural sensitivity continuum of mental health interventions in Sub-Saharan Africa: A systematic review. *Social Science & Medicine*, *306*, 115124. https://doi.org/10.1016/j.socscimed.2022.115124

Ando, M., Kao, Y.-C., Lee, Y.-C., Tai, S.-A., Mendez, S. R., Sasaki, K., Tang, W., & Papatheodorou, S. (2024). Remote cognitive behavioural therapy for older adults with anxiety symptoms: A systematic review and meta-analysis. *Journal of Telemedicine and Telecare*, *30*(9), 1376–1385. https://doi.org/10.1177/1357633X231151

Anisuzzaman, D., Malins, J. G., Friedman, P. A., & Attia, Z. I. (2024). Fine-tuning LLMs for specialised use cases. *Mayo Clinic Proceedings: Digital Health*. https://doi.org/10.1177/1357633X231151788

Anisuzzaman, D., Malins, J. G., Friedman, P. A., & Attia, Z. I. (2025). Fine-tuning large language models for specialised use cases. *Mayo Clinic Proceedings: Digital Health*, *3*(1), 1-13. https://doi.org/10.1016/j.mcpdig.2024.11.005

Baguma, R., Namuwaya, H., Nakatumba-Nabende, J., & Rashid, Q. M. (2023). Examining potential harms of large language models (LLMs) in Africa. *International Conference on Safe, Secure, Ethical, Responsible Technologies and Emerging Applications*, 3-19. https://doi.org/10.1007/978-3-031-56396-6_1

Caloudas, A. B., Frosio, K. E., Torous, J., Goss, C. W., Novins, D. K., Lindsay, J. A., & Shore, J. H. (2024). Mobile mental health applications for American Indian and Alaska Native communities:

Review and recommendations. *Journal of Technology in Behavioural Science*, *9*(3), 474–485. https://doi.org/10.1007/s41347-023-00348-9

Chiu, Y. Y., Sharma, A., Lin, I. W., & Althoff, T. (2024). A computational framework for behavioural assessment of LLM therapists. *arXiv preprint arXiv:2401.00820*. https://arxiv.org/abs/2401.00820

Coplan, J. D., Aaronson, C. J., Panthangi, V., & Kim, Y. (2015). Treating comorbid anxiety and depression: Psychosocial and pharmacological approaches. *World Journal of Psychiatry*, *5*(4), 366. https://doi.org/10.5498/wjp.v5.i4.366

De Choudhury, M., Pendse, S. R., & Kumar, N. (2023). Benefits and harms of large language models in digital mental health. *arXiv preprint arXiv:2311.14693*. https://arxiv.org/abs/2311.14693

Devlin, J. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Gabriel, S., Puri, I., Xu, X., Malgaroli, M., & Ghassemi, M. (2024). Can AI relate: Testing large language model response for mental health support. *arXiv preprint arXiv:2405.12021*. https://doi.org/10.48550/arXiv.2405.12021

Guo, Z., Lai, A., Thygesen, J. H., Farrington, J., Keen, T., & Li, K. (2024). Large language model for mental health: A systematic review. *arXiv preprint arXiv:2403.15401*. https://doi.org/10.48550/arXiv.2403.15401

Hays, P. A. (2009). Integrating evidence-based practice, cognitive–behaviour therapy, and multicultural therapy: Ten steps for culturally competent practice. *Professional Psychology: Research and Practice*, *40*(4), 354. https://doi.org/10.1037/a0016250

Hinton, D. E., & Patel, A. (2017). Cultural adaptations of cognitive behavioural therapy. *Psychiatric Clinics*, *40*(4), 701-714. https://doi.org/10.1016/j.psc.2017.08.006

Hua, Y., Liu, F., Yang, K., Li, Z., Sheu, Y.-h., Zhou, P., Moran, L. V., Ananiadou, S., & Beam, A. (2024). Large language models in mental health care: A scoping review. *arXiv preprint arXiv:2401.02984*. https://doi.org/10.48550/arXiv.2401.02984

Huey Jr, S. J., Park, A. L., Galán, C. A., & Wang, C. X. (2023). Culturally responsive cognitive behavioural therapy for ethnically diverse populations. *Annual Review of Clinical Psychology*, *19*(1), 51–78. https://doi.org/10.1146/annurev-clinpsy-080921-072750

Izumi, K., Tanaka, H., Shidara, K., Adachi, H., Kanayama, D., Kudo, T., & Nakamura, S. (2024). Response generation for cognitive behavioural therapy with large language models: Comparative study with Socratic questioning. *arXiv preprint arXiv:2401.15966*. https://doi.org/10.48550/arXiv.2401.15966

Jalal, B., Kruger, Q., & Hinton, D. E. (2020). Culturally adapted CBT (CA-CBT) for traumatised indigenous South Africans (Sepedi): A randomised pilot trial comparing CA-CBT to applied muscle relaxation. *Intervention Journal of Mental Health and Psychosocial Support in Conflict Affected Areas*, *18*(1), 61–65. https://doi.org/10.4103/INTV.INTV_68_18

Jameel, S., Munivenkatappa, M., Arumugham, S. S., & Thennarasu, K. (2022). Cultural adaptation of cognitive behaviour therapy for depression: A qualitative study exploring views of patients and practitioners from India. *The Cognitive Behaviour Therapist*, *15*, e16. https://doi.org/10.1017/S1754470X22000137

Jiang, M., Yu, Y. J., Zhao, Q., Li, J., Song, C., Qi, H., Zhai, W., Luo, D., Wang, X., & Fu, G. (2024). AI-enhanced cognitive behavioural therapy: Deep learning and large language models for extracting cognitive pathways from social media texts. *arXiv preprint arXiv:2404.11449*. https://doi.org/10.48550/arXiv.2404.11449

Kenton, J. D. M. W. C., & Toutanova, L. K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the NAACL-HLT* (Vol. 1, No. 2).

Kian, M. J., Zong, M., Fischer, K., Singh, A., Velentza, A.-M., Sang, P., Upadhyay, S., Gupta, A., Faruki, M. A., Browning, W., Arnold, S. M. R., Krishnamachari, B., & Mataric, M. J. (2024). Can an LLM-powered socially assistive robot effectively and safely deliver cognitive behavioural therapy? A study with university students. *arXiv preprint arXiv:2402.17937*. https://doi.org/10.48550/arXiv.2402.17937

Kunorubwe, T. (2023). Cultural adaptations of group CBT for depressed clients from diverse backgrounds: A systematic review. *The Cognitive Behaviour Therapist*, *16*, e35. https://doi.org/10.1017/S1754470X23000302

Lee, S., Kim, S., Kim, M., Kang, D., Yang, D., Kim, H., Kang, M., Jung, D., Kim, M. H., & Lee, S. (2024). Cactus: Towards psychological counselling conversations using cognitive behavioural theory. *arXiv preprint arXiv:2407.03103*. https://doi.org/10.48550/arXiv.2407.03103

Lee, Y. K., Suh, J., Zhan, H., Li, J. J., & Ong, D. C. (2024). Large language models produce responses perceived to be empathic. *arXiv preprint arXiv:2403.18148*. https://doi.org/10.48550/arXiv.2403.18148

Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., & Zhang, X. (2023). Domain specialization as the key to making large language models disruptive: A comprehensive survey. *arXiv preprint arXiv:2305.18703*. https://arxiv.org/abs/2305.18703

Liu, J. M., Li, D., Cao, H., Ren, T., Liao, Z., & Wu, J. (2023). Chatcounselor: A large language model for mental health support. *arXiv preprint arXiv:2309.15461*. https://doi.org/10.48550/arXiv.2309.15461

Manvi, R., Khanna, S., Burke, M., Lobell, D., & Ermon, S. (2024). Large language models are geographically biased. *arXiv preprint arXiv:2402.02680*. https://doi.org/10.48550/arXiv.2402.02680

Na, H. (2024). CBT-LLM: A Chinese large language model for cognitive behavioural therapy-based mental health question answering. *arXiv preprint arXiv:2403.16008*. https://doi.org/10.48550/arXiv.2403.16008

Nie, J., Shao, H., Fan, Y., Shao, Q., You, H., Preindl, M., & Jiang, X. (2024). LLM-based conversational AI therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. *arXiv preprint arXiv:2403.10779*. https://doi.org/10.48550/arXiv.2403.10779

Nozizwe, N. (2024). The effectiveness of cognitive-behavioural therapy for treating anxiety disorders in low-resource settings in South Africa. *International Journal of Psychology*, *9*(1), 47–59. https://doi.org/10.47604/ijp.2361

Obasa, A. E. (2024). Large language models through the lens of Ubuntu for health research in sub-Saharan Africa. *South African Journal of Science*, *120*(5/6).

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, *5*, 1-10. https://doi.org/10.1186/s13643-016-0384-4

Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, *14*(5), 2074. https://doi.org/10.3390/app14052074

Phiri, M., & Munoriyarwa, A. (2023). Health chatbots in Africa: Scoping review. *Journal of Medical Internet Research*, *25*, e35573. https://doi.org/10.2196/35573

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., ... & Azam, S. (2024). A review on large language models: Architectures, applications, taxonomies, open issues and challenges. *IEEE access*, *12*, 26839-26874. https://doi.org/10.1109/ACCESS.2024.3365742

Schiff, D. (2024). CBTLlama: Fine tuning large language models for identifying thought distortions. file:///C:/Users/omoda/Downloads/CBT_LLM.pdf

Shah, K., Xu, A. Y., Sharma, Y., Daher, M., McDonald, C., Diebo, B. G., & Daniels, A. H. (2024). Large language model prompting techniques for advancement in clinical medicine. *Journal of Clinical Medicine*, *13*(17), 5101. https://doi.org/10.3390/jcm13175101

Sham Sundhar, R., Shivavardhini, T., Daphine Desona Clemency, C., & Roobini, M. (2024). Lecter: A large language model chatbot for cognitive behavioural therapy. In M. L. Owoc, F. E. Varghese Sicily, K. Rajaram, & P. Balasundaram (Eds.), *Computational intelligence in data science. ICCIDS 2024* (Vol. 717, pp. 41–53). Springer. https://doi.org/10.1007/978-3-031-69982-5_4

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness*, *4*(3), 409-429. https://doi.org/10.1080/19439342.2012.710641

Sodi, T., Abas, M., Abdulaziz, M., Amos, A., Burgess, R. A., Hanlon, C., Kakunze, A., Kpobi, L., Lund, C., & Mwangi, K. J. (2024). A research agenda for mental health in sub-Saharan Africa. *Nature Medicine*, *30*(3), 616-617. https://doi.org/10.1038/s41591-023-02779-6

Spanhel, K., Balci, S., Feldhahn, F., Bengel, J., Baumeister, H., & Sander, L. B. (2021). Cultural adaptation of internet- and mobile-based interventions for mental disorders: A systematic review. *NPJ Digital Medicine*, *4*(1), 128. https://doi.org/10.1038/s41746-021-00498-1

Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., DeRubeis, R. J., Willer, R., & Eichstaedt, J. C. (2024). Large language models could change the future of behavioural healthcare: A proposal for responsible development and evaluation. *NPJ Mental Health Research*, *3*(1), 12. https://doi.org/10.1038/s44184-024-00056-z

Sun, H., Lin, Z., Zheng, C., Liu, S., & Huang, M. (2021). Psyqa: A Chinese dataset for generating long counselling text for mental health support. *arXiv preprint arXiv:2106.01702*. https://doi.org/10.48550/arXiv.2106.01702

Tonmoy, S., Zaman, S., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A comprehensive survey of hallucination mitigation techniques in large language models. *arXiv preprint arXiv:2401.01313*. https://doi.org/10.48550/arXiv.2401.01313

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., & Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Twomey, C., O'Reilly, G., & Byrne, M. (2015). Effectiveness of cognitive behavioural therapy for anxiety and depression in primary care: A meta-analysis. *Family Practice*, *32*(1), 3–15. https://doi.org/10.1093/fampra/cmu060

World Health Organisation. (2023). *Depressive disorder (depression)*. World Health Organisation.

Xiao, M., Xie, Q., Kuang, Z., Liu, Z., Yang, K., Peng, M., Han, W., & Huang, J. (2024). HealMe: Harnessing cognitive reframing in large language models for psychotherapy. *arXiv preprint arXiv:2403.05574*. https://doi.org/10.48550/arXiv.2403.05574

Zaghir, J., Naguib, M., Bjelogrlic, M., Névéol, A., Tannier, X., & Lovis, C. (2024). Prompt engineering paradigms for medical applications: Scoping review. *Journal of Medical Internet Research*, *26*, e60501. https://doi.org/10.2196/60501

Zhang, H., Qiao, Z., Wang, H., Duan, B., & Yin, J. (2024). VCounselor: A psychological intervention chat agent based on a knowledge-enhanced large language model. *arXiv preprint arXiv:2403.13553*.