

A Project Report on
COMPANIONMIND – A FRIENDLY AI COMPANION
FOR MENTAL WELL-BEING USING LLAMA

Submitted in partial fulfilment of the requirement for the award of the Degree

Bachelor of Technology

in

Computer Science and Engineering

Submitted By

MULUKUTLA POOJITHA SAI SRI LALITHA	21031A0534
BANDI KASI VENKATA KARTHIKEYA	21031A0506
BODDU SATYA SRI	22035A0502
ALLADI KOMALIKA	21031A0501
RAYUDU SRAVANI	21031A0545

Under the esteemed guidance of

Mr. CH. RAKESH

Assistant Professor of CSE Department



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY COLLEGE OF ENGINEERING NARASARAOPET
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
NARASARAOPET-522601, PALNADU, ANDHRA PRADESH

2024-2025

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY COLLEGE OF ENGINEERING NARASARAOPET
JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
NARASARAOPET-522601, PALNADU, ANDHRA PRADESH**

2024-2025



C E R T I F I C A T E

This is to certify that the dissertation entitled "**CompanionMind – A Friendly AI Companion For Mental Well-Being Using LLAMA**" that is being submitted by **MULUKUTLA POOJITHA SAI SRI LALITHA** (21031A0534), **BANDI KASI VENKATA KARTHIKEYA** (21031A0506), **BODDU SATYA SRI** (22035A0502), **ALLADI KOMALIKA** (21031A0501), **RAYUDU SRAVANI** (21031A0545), in partial fulfilment for the award of Bachelor of Technology in Computer Science and Engineering to the University College of Engineering Narasaraopet, Jawaharlal Nehru Technological Kakinada is a record of Bonafide work carried out by them under my guidance and supervision.

The results embedded in this dissertation have not been submitted to any other university/institute for the award of any degree/diploma

PROJECT SUPERVISOR

Mr. CH. RAKESH
Assistant Professor(c)
Department of CSE
UCEN-JNTUK

HEAD OF THE DEPARTMENT

Dr. G. MADHAVI
Assistant Professor & HOD (i/c)
Department of CSE
UCEN-JNTUK

EXTERNAL EXAMINER

DECLARATION

We hereby declare that the work described in the project report entitled "**CompanionMind – A Friendly AI Companion For Mental well-being Using LLAMA**" which is submitted by us in partial fulfilment of the requirements for the award of the degree Bachelor of Technology in the department of **COMPUTER SCIENCE AND ENGINEERING** to the college, University College of Engineering Narasaraopet, Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh is a record of original and independent research work done by us during the academic year 2024-2025 under the supervision of **Mr. CH. RAKESH**, Assistant Professor(c), CSE Department. The work is original and has not been submitted for the award of any Degree or Diploma or Associateship or Fellowship or other similar title to this or any other university.

Name of The Student	Roll No.	Signature
MULUKUTLA POOJITHA SAI SRI LALITHA	21031A0534	
BANDI KASI VENKATA KARTHIKEYA	21031A0506	
BODDU SATYA SRI	22035A0502	
ALLADI KOMALIKA	21031A0501	
RAYUDU SRAVANI	21031A0545	

PLACE:

DATE:

ACKNOWLEDGEMENT

It is needed with a great sense of pleasure and immense sense of gratitude that we acknowledge the help of these individuals. We express our sincere thanks and gratitude to many people who helped, supported, encouraged, and challenged us throughout our project and academic program. We take privilege to express our heartfelt gratitude to our Project Supervisor **Mr. CH. RAKESH**, Assistant Professor (c) of CSE Department and our Project Coordinator **Mrs. B. JYOTHI**, Assistant Professor (c) of CSE Department and our Head of the Department **Dr. G. MADHAVI**, Assistant Professor & HOD(i/c) Department of CSE, for their valuable suggestions and constant motivation that greatly helped us in successful completion of the project.

We express our sincere thanks to our beloved **Dr. Y. S. KISHORE BABU**, Vice-Principal, UCEN JNTUK and **Prof. CH. SRINIVASA RAO**, Principal, UCEN JNTUK for their support and constant motivation in successful completion of the project. We are thankful to all the faculty members for extending their kind cooperation and assistance. Finally, we are extremely thankful to our parents and friends for their constant help and moral support.

TEAM MEMBERS

MULUKUTLA POOJITHA SAI SRI LALITHA	21031A0534
BANDI KASI VENKATA KARTHIKEYA	21031A0506
BODDU SATYA SRI	22035A0502
ALLADI KOMALIKA	21031A0501
RAYUDU SRAVANI	21031A0545

TABLE OF CONTENTS

S.NO	CHAPTER NAME	PG.NO
	ABSTRACT	iii
	LIST OF FIGURES	iv
	LIST OF FIGURES	vi
	LIST OF ABBREVIATIONS	vii
1	INTRODUCTION	1
	1.1 Introduction	2
	1.2 Significance of the Project	4
	1.3 Summary	5
2	LITERATURE SURVEY	6
	2.1 Literature Survey	7
	2.2 Motivation	11
	2.3 Problem Statement	11
	2.4 Summary	11
3	SYSTEM ANALYSIS	12
	3.1 Existing System	13
	3.2 Proposed System	16
	3.3 Functional Requirements	22
	3.4 Non-Functional Requirements	23
	3.5 Hardware Requirements	24
	3.6 Software Requirements	25
	3.7 Summary	25
4	SYSTEM DESIGN	26
	4.1 System Architecture	27
	4.2 Modules	28
	4.3 UML Diagrams used in design	30
	4.4 Summary	40
5	SYSTEM IMPLEMENTATION	41
	5.1 Libraries	42

	5.2 Sample Project Code	45
6	SYSTEM TESTING	49
	6.1 Introduction	50
	6.2 Purpose Of Testing	50
	6.3 Types Of Testing	51
	6.4 Testing Methods	53
	6.5 Test cases	54
	6.6 Summary	74
7	SYSTEM RESULTS	75
8	CONCULSION AND FUTURE WORK	84
9	REFERENCES	87
10	BASE PAPER	89

ABSTRACT

The digital age has transformed mental health support by incorporating advanced AI-driven solutions, CompanionMind – A Friendly AI Companion for Mental Well-Being using Llama is designed to make mental health resources more accessible, efficient, and user-friendly by leveraging Large Language Models (LLMs) like Llama3.2. It provides immediate and personalized mental health assistance through empathetic conversations, enabling users to express their emotions, discuss challenges, and seek guidance. By understanding complex emotional contexts, the AI assistant offers real-time responses, practical coping strategies, and mental health exercises such as stress relief techniques, mindfulness meditation, and breathing exercises to help users improve their well-being. Additionally, depression assessment tests powered by machine learning algorithms are integrated to provide accurate evaluations—Support Vector Classification (SVC) is used for students and working professionals, while Gradient Boosting is applied as a common model for all users. These tests predict whether a user has depression or not and generate detailed reports, with an accuracy of 82% for Student Depression Test, 99% for Working Professional Depression Test, 74% for Common People Depression Test. The system also provides personalized interventions such as relaxation techniques, guided meditation, and calming music to support emotional wellbeing. Furthermore, CompanionMind ensures continuous support for individuals who may not have immediate access to therapy due to geographical, financial, or time constraints. By offering 24/7 accessibility, it bridges the gap in mental health services, providing comfort, encouragement, and structured well-being plans to those in need. This innovative solution empowers users to take control of their mental health, fostering self-awareness, resilience, and long-term emotional stability, while making mental health support more inclusive and readily available to everyone.

LIST OF FIGURES

S.NO	FIGURE.NO	DESCRIPTION	PAGE.NO
1	Figure 4.1.1	System Architecture	26
2	Figure 4.3.1	Usecase Diagram of CompanionMind	30
3	Figure 4.3.2	Class Diagram of CompanionMind	33
4	Figure 4.3.3	Sequence Diagram of CompanionMind	38
5	Figure 6.4.1	Black Box Testing	53
6	Figure 6.4.2	White Box Testing	53
7	Figure 6.5.1	Invalid User Authentication	54
8	Figure 6.5.2	Valid User Authentication	55
9	Figure 6.5.3	Student Depression Test Flow	58
10	Figure 6.5.4	Working Professional Test Flow	59
10	Figure 6.5.5	Common People Depression Test Flow	61
12	Figure 6.5.6.1	Identifying Support Needs	63
13	Figure 6.5.6.2	Identifying Support Needs	63
14	Figure 6.5.7.1	Recognizing Positive Well-being	64
15	Figure 6.5.7.2	Identifying Positive Well-being	65
16	Figure 6.5.8.1	Generating Personalized Reports with Well-being Recommendation	66
17	Figure 6.5.8.2	Generating Personalized Reports with Well-being Recommendation	66
18	Figure 6.5.9.1	Generating Personalized Reports with General tips	67
19	Figure 6.5.9.2	Generating Personalized Reports with General tips	68
20	Figure 6.5.10.1	Chatbot Providing helpful Responses	69
21	Figure 6.5.10.2	Chatbot Providing helpful Responses	69
23	Figure 6.5.11	Chatbot Offering Positive Affirmations	70
24	Figure 6.5.12.1	Chatbot Delivering Guided Relaxation	71
25	Figure 6.5.12.2	Chatbot Delivering Guided Relaxation	72

26	Figure 6.5.13	Ensuring Chatbot Response Only to mental health – Relate Queries	73
27	Figure 6.5.14.1	Click in Logout Button	74
28	Figure 6.5.14.2	Redirected to website main page	74
29	Figure 7.1.1	CompanionMind Page	77
30	Figure 7.1.2	Displaying Key Features – Assistance AI, Depression, powered chatbot	77
31	Figure 7.1.3	Displaying Key Features – Mindfulness Meditation, Breathing Use	78
32	Figure 7.1.4	Sig in page of companionMind	78
33	Figure 7.1.5	Sign up page of CompanionMind	79
34	Figure 7.1.6	Home page of companionMind-After signin	79
35	Figure 7.1.7	Activies of CompanionMind	80
36	Figure 7.1.8	Meditation timer started along with soothing background music	80
37	Figure 7.1.9	Meditation timer reset	81
38	Figure 7.1.10	Breathing Exercise for Stress relief Along with Instructions	81
39	Figure 7.1.11	Depression Test for Students, working Professionals, common people	82
40	Figure 7.1.12	Downloading Reports of Depression Test	82
41	Figure 7.1.13	Blogs to read regarding various concepts mental health	83
42	Figure 7.1.14	AI Powered Chatbot for Users	83

LIST OF TABLES

S.NO	Table Name	Description	PG.NO
1	Table 6.5.1	Test case 1	55
2	Table 6.5.2	Test case 2	56
3	Table 6.5.3	Test case 3	57
4	Table 6.5.4	Test case 4	59
5	Table 6.5.5	Test case 5	60
6	Table 6.5.6	Test case 6	62
7	Table 6.5.7	Test case 7	64
8	Table 6.5.8	Test case 8	66
9	Table 6.5.9	Test case 9	68
10	Table 6.5.10	Test case 10	70
11	Table 6.5.11	Test case 11	72
12	Table 6.5.12	Test case 12	73
13	Table 6.5.13	Test case 13	74
14	Table 6.5.14	Test case 14	75

LIST OF ABBREVIATIONS

S.NO	ABBREVIATION	EXPANSION
1	LLM	Large Language Models
2	ML	Machine Learning
3	SVC	Support Vector Classifier
4	GB	Gradient Boosting
5	PDF	Portable Document Format
6	UML	Unified Modeling Language
7	UAT	User Acceptance Testing
8	UI	User Interface
9	API	Application Programming Interface

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Mental health refers to a person's emotional, psychological, and social well-being. It affects how we think, feel, and act, and it plays a significant role in how we handle stress, relate to others, and make choices. Just as physical health is essential to overall well-being, mental health is a crucial aspect of a person's life, influencing every part of their daily functioning. Mental health is not just the absence of mental disorders, but also the presence of positive psychological functioning.

Several factors contribute to mental health, including biological, environmental, and lifestyle elements. Genetics, brain chemistry, and family history can all influence a person's mental state. Likewise, external factors such as childhood experiences, trauma, social support, and life events can have a lasting impact. The modern world, with its high pace and demanding nature, has introduced new challenges to maintaining mental health, such as social media pressures, work-related stress, and economic uncertainty. These factors can lead to mental health conditions such as anxiety, depression, or even more severe disorders.

Taking care of mental health involves both preventive and therapeutic measures. For many, maintaining mental health requires regular self-care routines, including physical exercise, balanced nutrition, social connection, and mindfulness practices. In addition, seeking professional help when needed is essential for managing mental health conditions. Therapies like counselling, cognitive-behavioural therapy (CBT), and medication can be effective in managing and treating various mental health issues. It is important to break the stigma surrounding mental health so that individuals feel empowered to seek help without fear of judgment.

Furthermore, mental health is closely connected to physical health. Mental disorders, if left untreated, can manifest in physical symptoms like fatigue, headaches, or gastrointestinal issues. Likewise, chronic physical conditions may lead to mental health challenges. The mind and body are interlinked, making an integrated approach to well-being essential for overall health. Awareness, education, and accessible mental health resources are key to supporting individuals in maintaining a balanced and healthy life.

Importance of Mental Health:

In today's fast-paced and often stressful world, the importance of mental health has never been more evident. The pressures of modern life—ranging from work demands and financial instability to the constant connectivity of social media—have made it increasingly difficult for individuals to manage their emotional and psychological well-being. Mental health issues such as anxiety, depression, and burnout are on the rise, affecting people across all age groups and walks of life. As society becomes more aware of these challenges, the need to prioritize mental health has become a critical aspect of both personal well-being and public health. Maintaining good mental health is not only necessary for managing stress and improving life satisfaction but also plays a vital role in building resilience to cope with life's challenges.

Moreover, the COVID-19 pandemic further highlighted the importance of mental health, as many people faced isolation, uncertainty, and trauma. The increased focus on mental health has led to more open conversations about the stigma surrounding mental illnesses and a growing recognition that mental health is just as important as physical health. With more resources and support systems being developed, individuals are now encouraged to seek help without fear of judgment. Fostering good mental health can enhance overall quality of life, improve productivity, strengthen relationships, and help individuals lead more fulfilling lives. In a world where mental health challenges are increasingly common, it is essential that mental well-being remains a top priority for individuals, communities, and healthcare systems.

Large Language Models (LLMs):

LLMs, such as OpenAI's GPT (Generative Pre-trained Transformer) series, have demonstrated exceptional proficiency in understanding and generating human-like text. These models are pretrained on vast amounts of text data, enabling them to learn intricate language patterns and semantics. Through fine-tuning on specific tasks, LLMs can be tailored to excel in various NLP applications, including question answering.

Using large language models (LLMs) in mental health offers exciting possibilities for improving access to care, enhancing therapeutic interventions, and providing support to those dealing with mental health issues. These advanced AI models, such as GPT-4, have the potential to assist in a variety of ways, from offering immediate emotional support to augmenting traditional therapeutic practices.

One of the most promising applications of LLMs is in providing mental health support through chatbots or virtual assistants. These AI-driven tools can offer users a safe, anonymous space to express their feelings and concerns. While they are not a substitute for professional therapy, they can help bridge the gap between individuals and mental health services, particularly for those who may not have easy access to a therapist due to geographical, financial, or social barriers. LLMs can engage in empathetic conversations, helping users to process emotions, practice coping strategies, and even offer suggestions for managing stress, anxiety, or depression. These virtual assistants can also provide real-time feedback, helping users identify and challenge negative thought patterns, thus supporting therapeutic techniques like cognitive-behavioural therapy (CBT).

1.2 SIGNIFICANCE OF THE PROJECT:

Using large language models (LLMs) in mental health plays a crucial role in addressing the challenges associated with mental health care, such as stigma, accessibility, and affordability. One of the most significant barriers to seeking help for mental health issues is the stigma that still surrounds them, often preventing individuals from reaching out for support. By integrating LLMs into mental health solutions, we can provide users with a safe, anonymous, and non-judgmental environment to explore their emotions and seek guidance. The project not only aids in removing this stigma but also helps overcome other barriers, such as financial constraints, societal expectations, and geographic limitations. With the widespread availability of smartphones and the internet, AI-driven mental health tools powered by LLMs can be accessed by anyone, anywhere, at any time, democratizing access to mental health support.

This project will begin by assessing users' depression through a comprehensive and personalized questionnaire, enabling the system to gauge their emotional state, identify areas of concern, and provide tailored feedback. It also includes relaxation techniques like guided meditation with soothing background music, breathing exercises. Additionally, users can engage in empathetic conversations with a chatbot powered by LLMs. The chatbot can offer comforting dialogue, emotional validation, and practical coping strategies, mimicking the supportive role of a therapist or counsellor. While not a replacement for professional care, this interaction provides users with immediate support and fosters emotional well-being. Furthermore, the AI can continually adapt to users' evolving needs, offering real-time, personalized advice that supports their mental health journey.

1.3 SUMMARY:

In this chapter we have discussed about the development of a platform that mainly focuses on an individual Mental Health using LLMs. We have also provided a quick view of how this system is being built.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

There are various standard papers on using LLMs in Mental Health applications. The methodology used, merits and demerits of those papers are:

[1] Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation

This paper provides a roadmap for the ambitious yet responsible application of clinical LLMs in psychotherapy. First, a technical overview of clinical LLMs is presented. Second, the stages of integration of LLMs into psychotherapy are discussed while highlighting parallels to the development of autonomous vehicle technology. Third, potential applications of LLMs in clinical care, training, and research are discussed, highlighting areas of risk given the complex nature of psychotherapy. Fourth, recommendations for the responsible development and evaluation of clinical LLMs are provided, which include centering clinical science, involving robust interdisciplinary collaboration, and attending to issues like assessment, risk detection, transparency, and bias. Lastly, a vision is outlined for how LLMs might enable a new generation of studies of evidence-based interventions at scale, and how these studies may challenge assumptions about psychotherapy.

[2] The Opportunities and Risks of Large Language Models in Mental Health

In this paper, we summarize the extant literature on efforts to use LLMs to provide mental health education, assessment, and intervention and highlight key opportunities for positive impact in each area. We then highlight risks associated with LLMs' application to mental health and encourage the adoption of strategies to mitigate these risks. The urgent need for mental health support must be balanced with responsible development, testing, and deployment of mental health LLMs. It is especially critical to ensure that mental health LLMs are fine-tuned for mental health, enhance mental health equity, and adhere to ethical standards and that people, including those with lived experience with mental health concerns, are involved in all stages from development through deployment. Prioritizing these efforts will minimize potential harms to mental health and maximize the likelihood that LLMs will positively impact mental health globally.

[3] Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data

Advances in large language models (LLMs) have empowered a variety of applications. However, there is still a significant gap in research when it comes to understanding and enhancing the capabilities of LLMs in the field of mental health. In this work, we present a comprehensive evaluation of multiple LLMs on various mental health prediction tasks via online text data, including Alpaca, Alpaca-LoRA, FLAN-T5, GPT-3.5, and GPT-4. We conduct a broad range of experiments, covering zero-shot prompting, few-shot prompting, and instruction fine-tuning. The results indicate a promising yet limited performance of LLMs with zero-shot and few-shot prompt designs for mental health tasks. More importantly, our experiments show that instruction finetuning can significantly boost the performance of LLMs for all tasks simultaneously. Our best-finetuned models, Mental-Alpaca and Mental-FLAN-T5, outperform the best prompt design of GPT-3.5 (25 and 15 times bigger) by 10.9% on balanced accuracy and the best of GPT-4 (250 and 150 times bigger) by 4.8%. They further perform on par with the state-of-the-art task-specific language model. We also conduct an exploratory case study on LLMs' capability on mental health reasoning tasks, illustrating the promising capability of certain models such as GPT-4. We summarize our findings into a set of action guidelines for potential methods to enhance LLMs' capability for mental health tasks. Meanwhile, we also emphasize the important limitations before achieving deployability in real-world mental health settings, such as known racial and gender bias. We highlight the important ethical risks accompanying this line of research.

[4] Large Language Models for Mental Health Applications: Systematic Review

Large language models (LLMs) are advanced artificial neural networks trained on extensive datasets to accurately understand and generate natural language. While they have received much attention and demonstrated potential in digital health, their application in mental health, particularly in clinical settings, has generated considerable debate. This systematic review aims to critically assess the use of LLMs in mental health, specifically focusing on their applicability and efficacy in early screening, digital interventions, and clinical settings. By systematically collating and assessing the evidence from current studies, our work analyzes models, methodologies, data sources, and outcomes, thereby highlighting the potential of LLMs in mental health, the challenges they present, and the prospects for their clinical use.

[5] Large Language Models in Mental Health Care: a Scoping Review

The integration of large language models (LLMs) in mental health care is an emerging field. There is a need to systematically review the application outcomes and delineate the advantages and limitations in clinical settings. This review aims to provide a comprehensive overview of the use of LLMs in mental health care, assessing their efficacy, challenges, and potential for future applications. A systematic search was conducted across multiple databases including PubMed, Web of Science, Google Scholar, arXiv, medRxiv, and PsyArXiv in November 2023. All forms of original research, peer-reviewed or not, published or disseminated between October 1, 2019, and December 2, 2023, are included without language restrictions if they used LLMs developed after T5 and directly addressed research questions in mental health care settings. From an initial pool of 313 articles, 34 met the inclusion criteria based on their relevance to LLM application in mental health care and the robustness of reported outcomes. Diverse applications of LLMs in mental health care are identified, including diagnosis, therapy, patient engagement enhancement, etc. Key challenges include data availability and reliability, nuanced handling of mental states, and effective evaluation methods. Despite successes in accuracy and accessibility improvement, gaps in clinical applicability and ethical considerations were evident, pointing to the need for robust data, standardized evaluations, and interdisciplinary collaboration. LLMs hold substantial promise for enhancing mental health care. For their full potential to be realized, emphasis must be placed on developing robust datasets, development and evaluation frameworks, ethical guidelines, and interdisciplinary collaborations to address current limitations.

[6] Opportunities and risks of large language models in psychiatry

The integration of large language models (LLMs) into mental healthcare and research heralds a potentially transformative shift, one offering enhanced access to care, efficient data collection, and innovative therapeutic tools. This paper reviews the development, function, and burgeoning use of LLMs in psychiatry, highlighting their potential to enhance mental healthcare through improved diagnostic accuracy, personalized care, and streamlined administrative processes. It is also acknowledged that LLMs introduce challenges related to computational demands, potential for misinterpretation, and ethical concerns, necessitating the development of pragmatic frameworks to ensure their safe deployment. We explore both the promise of LLMs in enriching psychiatric care and research through examples such as predictive analytics and therapy chatbots and risks including labor substitution, privacy concerns, and the necessity for responsible AI practices. We conclude by advocating for processes to develop responsible guardrails, including red-teaming, multi-stakeholder-

oriented safety, and ethical guidelines/frameworks, to mitigate risks and harness the full potential of LLMs for advancing mental health.

[7] PsycoLLM: Enhancing LLM for Psychological Understanding and Evaluation

Mental health has attracted substantial attention in recent years and LLM can be an effective technology for alleviating this problem owing to its capability in text understanding and dialogue. However, existing research in this domain often suffers from limitations, such as training on datasets lacking crucial prior knowledge and evidence, and the absence of comprehensive evaluation methods. In this paper, we propose a specialized psychological large language model (LLM), named PsycoLLM, trained on a proposed high-quality psychological dataset, including single-turn QA, multi-turn dialogues and knowledge-based QA. Specifically, we construct multi-turn dialogues through a three-step pipeline comprising generation, evidence judgment, and refinement. We augment this process with real-world psychological case backgrounds extracted from online platforms, enhancing the relevance and applicability of the generated data. Additionally, to compare the performance of PsycoLLM with other LLMs, we develop a comprehensive psychological benchmark based on authoritative psychological counseling examinations in China, which includes assessments of professional ethics, theoretical proficiency, and case analysis. The experimental results on the benchmark illustrates the effectiveness of PsycoLLM, which demonstrates superior performance compared to other LLMs

2.2 MOTIVATION

The motivation behind choosing this project stems from the growing need for accessible and personalized mental health support. Many individuals struggling with depression lacks immediate access to professional help due to financial constraints, social stigma, or limited availability of mental health professionals. By integrating assessment tools, personalized recommendations, and an empathetic chatbot, this platform aims to bridge that gap, providing users with a safe space for self-reflection and support. Additionally, incorporating mindfulness meditation and breathing exercises promotes holistic well-being, helping individuals manage stress and anxiety effectively. This project is driven by the vision of leveraging technology to make mental health care more inclusive, proactive, and easily accessible to those in need.

2.3 PROBLEM STATEMENT

The goal is to create a comprehensive platform that provides a clear assessment of an individual's depression, offers personalized recommendations to address any concerns, and features a chatbot for empathetic conversations and support mindfulness meditation, breathing exercises.

2.2 SUMMARY

This chapter reviews existing works on using LLMs for mental health, highlighting the advantages and limitations of current systems. While several approaches to creating mental health platforms with LLMs exist, each has limitations in fully meeting the diverse needs and features desired by individuals.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISITING SYSTEMS

There are some systems which perform similar functions as our system, though they didn't match in every aspect of our system functionalities.

3.1.1.1 Components of the Existing Mental Health System

3.1.1.1 Hospitals & Clinics

- **Public Hospitals:** Government-run hospitals offer psychiatric services, typically with a focus on inpatient care for severe mental health conditions like schizophrenia, bipolar disorder, or major depressive disorder. These hospitals may also provide medication management and outpatient services.
- **Private Clinics:** Private mental health clinics are often staffed by psychiatrists, psychologists, and therapists. These facilities typically cater to individuals seeking personalized care, often with shorter wait times, though they come with a higher cost.

3.1.1.2 Psychiatrists & Psychologists

- **Psychiatrists:** These medical professionals diagnose and treat mental health disorders with medication and sometimes therapy. They often work in hospitals, private practices, or clinics. As medical doctors, psychiatrists can prescribe medication, manage complex mental health cases, and provide treatment for mental illnesses like schizophrenia, major depression, and anxiety disorders.
- **Psychologists:** Psychologists focus on understanding the mind through assessments, testing, and psychotherapy (talk therapy). They may treat conditions like anxiety, depression, PTSD, and more, but unlike psychiatrists, they cannot prescribe medication. Psychologists may specialize in clinical psychology, counseling, or school psychology.

3.1.1.3 Community Health Centers

- **Government-run Centers:** These centers are typically available in urban and semi-urban regions and provide mental health services as part of public health initiatives. They offer counseling, mental health education, support groups, and may even offer free or subsidized services.
- **NGO-run Centers:** Non-governmental organizations (NGOs) often fill gaps in the mental health infrastructure, especially in underserved areas. These centers may provide essential mental health services, advocacy, and community support at low or no cost.

3.1.1.4 Helplines & Online Therapy

- **24/7 Mental Health Helplines:** National and regional helplines like the National Suicide Prevention Lifeline or local mental health crisis lines provide immediate, anonymous support to individuals in emotional distress. These services are vital in moments of crisis, offering immediate interventions and guiding individuals to local resources.
- **Virtual Therapy Platforms:** Online therapy platforms, such as BetterHelp or Talkspace, offer virtual counseling services, making mental health support more accessible, particularly for those who are unable to attend in-person therapy. These platforms often connect users with licensed therapists via chat, video, or phone calls.

3.1.1.5 Rehabilitation & Support Programs

- **Addiction Recovery Facilities:** These programs are designed to help individuals struggling with substance use disorders. They provide detoxification, counseling, group therapy, and long-term recovery strategies. Rehabilitation can also extend to people recovering from trauma, such as domestic abuse or war-related trauma.
- **Suicide Prevention Programs:** Many hospitals and non-profits run programs specifically aimed at preventing suicide, including crisis intervention teams, hotlines, and community education to identify warning signs and provide support.
- **Post-Trauma Care:** Programs focused on trauma recovery, such as Cognitive Behavioral Therapy (CBT), Eye Movement Desensitization and Reprocessing (EMDR), and trauma-informed care, support people recovering from physical or emotional trauma.

3.1.1 Drawbacks in the Existing System

3.1.2.1 Limited Accessibility

- **Rural & Remote Areas:** One of the biggest challenges in mental health care is the geographic disparity. Rural areas often have very few mental health professionals, forcing residents to travel long distances to seek care. In many remote regions, access to hospitals or community centers that offer mental health services is limited or non-existent.
- **Lack of Infrastructure:** Even in urban settings, some low-income neighborhoods suffer from a lack of mental health infrastructure, such as community centers or affordable clinics. This makes it difficult for people to access treatment on a consistent basis.

3.1.2.2 Stigma & Awareness Issues

- **Cultural Barriers:** In many cultures, mental illness is seen as a taboo topic, and individuals may feel embarrassed or ashamed to seek treatment. This cultural stigma can prevent individuals from acknowledging their mental health struggles, leading to untreated conditions.
- **Lack of Awareness:** There is still limited awareness about mental health conditions, particularly in developing countries or among marginalized populations. Without proper education and resources, many individuals may not even recognize the symptoms of mental health issues, leading to delayed treatment and worsening conditions.

3.1.2.3 High Costs

- **Expensive Treatment:** Private therapy, psychiatric care, and medication can be prohibitively expensive for many individuals, especially in countries with no universal healthcare. Even if mental health services are available through insurance, many people lack sufficient coverage or have high deductibles.
- **Limited Insurance Coverage:** Mental health is still often underfunded and undervalued in many insurance plans. Coverage for therapy and counseling might be limited, with many plans offering only a few sessions per year or excluding certain treatments altogether.

3.1.2.4 Shortage of Mental Health Professionals

- **Undertrained Workforce:** While the number of mental health professionals is growing, there is still a critical shortage in many areas. This is particularly true in low-income or rural regions, where the demand for mental health services exceeds the available supply of qualified professionals.
- **Burnout and Retention Issues:** Mental health professionals are often overburdened due to high caseloads, leading to burnout and a shortage of skilled workers. Additionally, professionals working in public institutions may face low pay or poor working conditions, contributing to higher turnover rates.

3.2 PROPOSED SYSTEM

The goal of **CompanionMind** is to create a comprehensive platform that integrates advanced language model technology to provide individuals with a clear and insightful assessment of their depression. This platform not only evaluates users' emotional and psychological states but also

offers personalized recommendations and tools to address any concerns they may have. Through this system, users can access a unique, user-friendly mental health experience that combines evaluation, support, and actionable advice, all driven by AI-powered language models.

The system begins by providing an in-depth assessment of an individual's depression. Using Machine Learning algorithms, the system asks targeted questions and interprets responses to create a comprehensive understanding of the user's depression. This evaluation helps to identify depression levels, allowing users to better understand their depression levels.

Once the assessment is complete, the system provides personalized recommendations tailored to the individual's unique needs. These recommendations can include self-care strategies, mindfulness exercises, and practical advice for managing mental health. By leveraging evidence-based approaches, the system ensures that users receive relevant and effective solutions that are designed to promote mental well-being.

In addition to assessments and recommendations, the platform features an empathetic chatbot powered by an LLM that engages users in real-time, compassionate conversations. The chatbot offers a safe, non-judgmental space for individuals to express their feelings, share their concerns, and receive emotional support. The language model is designed to understand the emotional tone of the conversation, providing thoughtful, empathetic responses that make users feel heard and supported.

The beauty of **CompanionMind** lies in its accessibility and responsiveness. Users can interact with the platform at any time, making it a valuable tool for those who need immediate support or guidance. Whether it's someone experiencing a moment of anxiety or someone looking to build healthier mental habits, the platform provides real-time feedback and ongoing support, empowering users to take control of their mental health journey.

3.2.1 Key Components and Functionalities:

- **Comprehensive Mental Health Assessment:** The platform begins by guiding users through a detailed mental health evaluation. Using Machine Learning, the system analyses responses to a series of carefully curated questions. This assessment not only identifies emotional states like anxiety, stress, or depression but also helps individuals gain a clearer understanding of their mental health landscape.

- **Personalized Recommendations** After the assessment, the platform provides tailored suggestions aimed at addressing the specific mental health needs of the user. These can range from self-care practices, mindfulness exercises, stress-relief techniques, to expert-backed advice on managing emotional well-being. By offering personalized content, the platform ensures that each individual receives practical and evidence-based support for their unique mental health needs.

- **Empathetic, AI-powered Chatbot** One of the standout features of the platform is its AI-driven, empathetic chatbot. Available 24/7, the chatbot offers a safe space for users to share their thoughts, concerns, and emotional experiences. Whether users are dealing with a specific mental health challenge or simply need someone to talk to, the chatbot listens attentively and provides thoughtful, compassionate responses. The LLM is trained to understand emotional tones, ensuring that interactions are both comforting and validating.

- **Additional Features for Mental Well-Being:**
 - **Meditation & Mindfulness:** The platform provides **guided meditation sessions** designed to reduce stress, improve focus, and promote emotional well-being.
 - **Breathing Exercises:** A variety of **breathing techniques** are available to help users manage anxiety and stress.

3.2.2 Advantages of Proposed System:

- **Holistic Mental Health Support:**

The platform offers a comprehensive approach to mental well-being by combining assessments, personalized recommendations, real-time emotional support, and practical exercises. This helps address mental health from multiple angles, supporting users in managing both short-term challenges and long-term well-being.

- **Personalized Care and Attention:**

By assessing the user's mental health state and providing tailored recommendations, the platform ensures that each individual receives the support and advice that best fits their unique needs. This individualized approach increases the effectiveness of the interventions and boosts user engagement.

➤ **Convenience and Accessibility:**

The platform's features, such as the 24/7 empathetic chatbot and on-demand mindfulness and breathing exercises, make it easy for users to access mental health support whenever they need it. Users can engage in self-care practices, talk to the AI chatbot, or complete a mental health assessment at any time, without having to wait for an appointment or scheduled therapy session.

➤ **Empowerment Through Self-Management:**

The system provides users with tools like meditation, mindfulness exercises, and breathing techniques to manage their own mental health. These proactive practices empower individuals to take control of their emotional well-being, building long-term coping strategies and resilience.

➤ **Emotional Support Without Stigma:**

Many people hesitate to seek mental health support due to fear of judgment. The platform's AI-powered chatbot offers a private, non-judgmental space where users can openly discuss their feelings without the fear of stigma, fostering a sense of relief and validation.

➤ **Immediate Relief and Comfort:**

In moments of distress, the chatbot can offer instant emotional support, helping to reduce feelings of isolation or anxiety. The breathing exercises and mindfulness practices also provide immediate relief, helping users to calm down and refocus during stressful times.

➤ **Reduces Barriers to Mental Health Care:**

The platform offers an affordable and accessible alternative to traditional therapy and mental health services, reducing the financial, time, and social barriers many individuals face in seeking professional help. It provides a stepping stone for those who may be hesitant to seek therapy but still need support.

➤ **Comprehensive Mental Health Education:**

Users not only benefit from personalized support but also gain valuable knowledge about their own mental health, such as learning about specific emotional states and effective ways to address them. This education promotes greater self-awareness and helps users make informed decisions about their well-being.

3.2.3 Overcoming Drawbacks of Existing System:

The existing mental health systems have certain limitations that hinder accessibility, affordability, and the overall effectiveness of care. The **CompanionMind** platform has been designed to address these limitations through advanced AI technologies, ensuring a more comprehensive, accessible, and personalized mental health experience. Here's how the proposed system overcomes the challenges in traditional mental health care:

➤ Personalized and Scalable Mental Health Assessments:

- Existing System Drawbacks:

- **Lack of Tailored Assessments:** Many traditional mental health systems offer generalized assessments or rely solely on clinician-based evaluations. These approaches often miss nuances in individual mental health needs and emotional states.

- **Limited Reach:** In-person assessments are limited by geographical constraints and the availability of trained professionals, leading to delays and potentially inaccurate assessments.

- How the Proposed System Overcomes This:

- **Advanced AI-Driven Assessments:** The platform uses **Machine Learning (ML)** techniques to create comprehensive, personalized mental health assessments. By analysing user responses, the system accurately identifies specific emotional states (such as stress, anxiety, or depression) and provides personalized feedback.

- **Scalable and Accessible:** The system is accessible globally, allowing individuals to take the assessment at any time and from any location, overcoming the limitations of traditional in-person assessments.

➤ Real-Time, Compassionate Conversations:

- Existing System Drawbacks:

- **Limited Real-Time Emotional Support:** Many traditional systems lack instant access to emotional support, and users may have to wait for an appointment or scheduled session to receive care, leading to delays in critical moments.

- **Stigma and Anxiety:** Some users may feel uncomfortable sharing sensitive information with human therapists due to stigma, fear of judgment, or societal pressures.

- How the Proposed System Overcomes This:

- **Instant, Empathetic AI Conversations:** The platform offers a **24/7 AI chatbot** that understands and responds to emotional tones with empathy and compassion. It provides a private, non-judgmental space where users can share their feelings, concerns, and struggles, which reduces feelings of isolation and fear of stigma.
- **Real-Time Emotional Support:** In times of emotional distress, users can access immediate support through interactive conversations with the chatbot, offering comfort and practical advice in real time.

➤ **Scalable and Accessible for All Users:**

- **Existing System Drawbacks:**
- **Limited Reach and Scalability:** Many traditional systems cannot efficiently scale to serve a large population, particularly in underserved or rural areas, due to the shortage of professionals and infrastructure limitations.
- **Inaccessibility due to Cost and Time Constraints:** Traditional therapy can be cost-prohibitive for many individuals, and the time commitment of regular in-person therapy sessions makes it hard for some to access consistent care.
- **How the Proposed System Overcomes This:**
- **Scalable to Global Users:** The platform is built to handle large volumes of users, providing consistent, high-quality support to anyone, anywhere. It ensures scalability, which means that whether there are 100 or 10,000 users, the system can provide personalized care without limitations.
- **Affordable and Accessible:** The LLM platform provides mental health support without the significant costs associated with private therapy or hospital visits. Users can access assessments, chatbot support, educational content, and self-care tools at little to no cost, making mental health care more affordable and widely accessible.

SOFTWARE REQUIREMENT SPECIFICATIONS

A Software Requirements Specifications (SRS), also known as a software requirements document, serves as a comprehensive description of the intended functionality, behaviour, and constraints of a software system. It provides a blueprint for software development, outlining what need to be built and how it should behave.

3.3 FUNCTIONAL REQUIREMENTS:

These requirements serve as a foundation for the design, development, and testing of the software. When documenting functional requirements, it is essential to provide comprehensive and unambiguous descriptions of the desired system behaviour. Each requirement should specify a particular action or feature that the software must support. Additionally, functional requirements should be measurable and verifiable, meaning that they should be testable to ensure compliance during the software development lifecycle.

1. User Management

- Users should be able to **sign up, log in, and log out** securely.
- The system should allow users to **update their profile** (age, gender, mental health history, etc.).

2. Depression Assessment

- Users should be able to take a **depression assessment**.
- The system should analyze the responses and provide a **detailed assessment report**.

3. Personalized Recommendations

- Based on the assessment, the system should generate **personalized recommendations**.

4. Chatbot for Emotional Support

- The chatbot should provide **empathetic and human-like conversations**.
- Users should be able to **ask mental health-related questions** and receive supportive responses.

5. Mindfulness & Breathing Exercises

- The platform should include **guided mindfulness meditation sessions**.
- Users should be able to access **breathing exercises** for stress relief.

3.4 NON – FUNCTIONAL REQUIREMENTS:

Non-functional requirements are the essential guidelines that dictate how a software system should perform and behave, focusing on aspects beyond specific functionalities. Unlike functional requirements that define what the system should do, non-functional requirements specify how well the system should do it. These requirements are essential for ensuring that the software meets the desired standards of performance, reliability, security, usability, and other key aspects.

- **Usability** – The system should have an intuitive and user-friendly interface, ensuring ease of navigation for individuals of all age groups and backgrounds.

- **Scalability** – The system should be able to handle multiple users simultaneously without performance degradation, ensuring seamless support as the user base grows.
- **Performance** – Responses from the chatbot and assessment modules should be fast and efficient, with minimal latency to enhance user experience.
- **Reliability** – The system should have high uptime (99.9%) and ensure that users can access services without frequent crashes or downtimes.
- **Security & Privacy** – User data should be stored securely, ensuring confidentiality and compliance.
- **Maintainability** – The codebase should be well-documented to allow easy updates, debugging, and feature enhancements over time.
- **Ethical AI & Bias Mitigation** – The chatbot and recommendation system should be designed to provide unbiased and ethical mental health guidance, avoiding harmful or misleading advice.

3.5 HARDWARE REQUIREMENTS:

Hardware requirements refer to the specific hardware components and specifications necessary to run a software application or system effectively. These requirements are outlined to ensure that the software functions optimally and reliably on the intended hardware environment. Hardware requirements typically include details such as processor type and speed, memory, storage capacity, display resolution, graphics card, network connectivity, and any additional peripherals or devices needed for proper operation.

- **Processor:** Intel Xeon / AMD EPYC (16+ cores)
- **RAM:** Minimum **32GB** (Recommended **64GB+** for handling AI models)
- **Storage:** SSD (Minimum **1TB**, recommended **2TB+** for data storage and model files)
- Users will access the platform via web or mobile applications, requiring:
 1. **Smartphones/Tablets**
 2. **PC/Laptops**
 3. **Web Browsers:** Chrome, Firefox, Edge, Safari (latest versions)

3.6 SOFTWARE REQUIREMENTS:

Software requirements involve defining software resource requirements that need to be installed on a computer to provide optimal functioning of an application. These requirements are usually not included in the software installation package and required to be installed separately before the software is installed.

- Operating System: Windows 10/11
- IDE: Visual Studio Code
- Software: Ollama – To run LLMs Locally on system
- Programming Language: Python
- Libraries: Streamlit, Ollama, Pickle, fpdf, numpy, pandas

3.7 SUMMARY:

In this chapter we discussed about the system requirements like what hardware is required and what are the software requirements one must have to know to use our model along with functional and non-functional requirements.

CHAPTER 4

SYSTEM DESIGN

SYSTEM DESIGN

System design is an essential step in the software development life cycle, which involves designing the system architecture, modules, components, and interfaces to satisfy the specified requirements. The system design in a project is critical to ensure that the system meets the specified requirements, is scalable, and is maintainable in the long run.

4.1 SYSTEM ARCHITECHTURE

A system architecture is the conceptual model that defines the structure, behaviour, and more views of a system. It shows the connections between the various components of the system and indicates what functions are to be performed.

The system architecture of the LLM-based Mental Health platform is designed to provide personalized mental health support through AI-powered recommendations, meditation guidance, and a chatbot.

1. **User Authentication:** Users can sign up or log in to access the platform. Upon login, they are directed to the home page, where they can explore various mental health resources.
2. **Mental Health Assessment:** The system employs machine learning models such as **SVC Algorithm** and **Gradient Boosting** to analyze user inputs and determine their mental health status. Based on the results, personalized recommendations are provided to help users prevent or overcome depression.
3. **Core Features:** The home page offers multiple services, including **meditation exercises**, a **depression test**, and **breathing exercises** with interactive timers and soothing background music for relaxation.
4. **AI-Powered Chatbot:** A chatbot powered by **Llama 3.2** is integrated to process user queries. It converts user input into tokens and determines whether the query is related to mental health. If relevant, it generates an appropriate response; otherwise, it provides predefined responses for general queries.
5. **Blog Section:** Users can explore various blogs related to mental well-being, self-care techniques, and coping strategies.

CompanionMind – A friendly AI companion for mental well-being using Llama

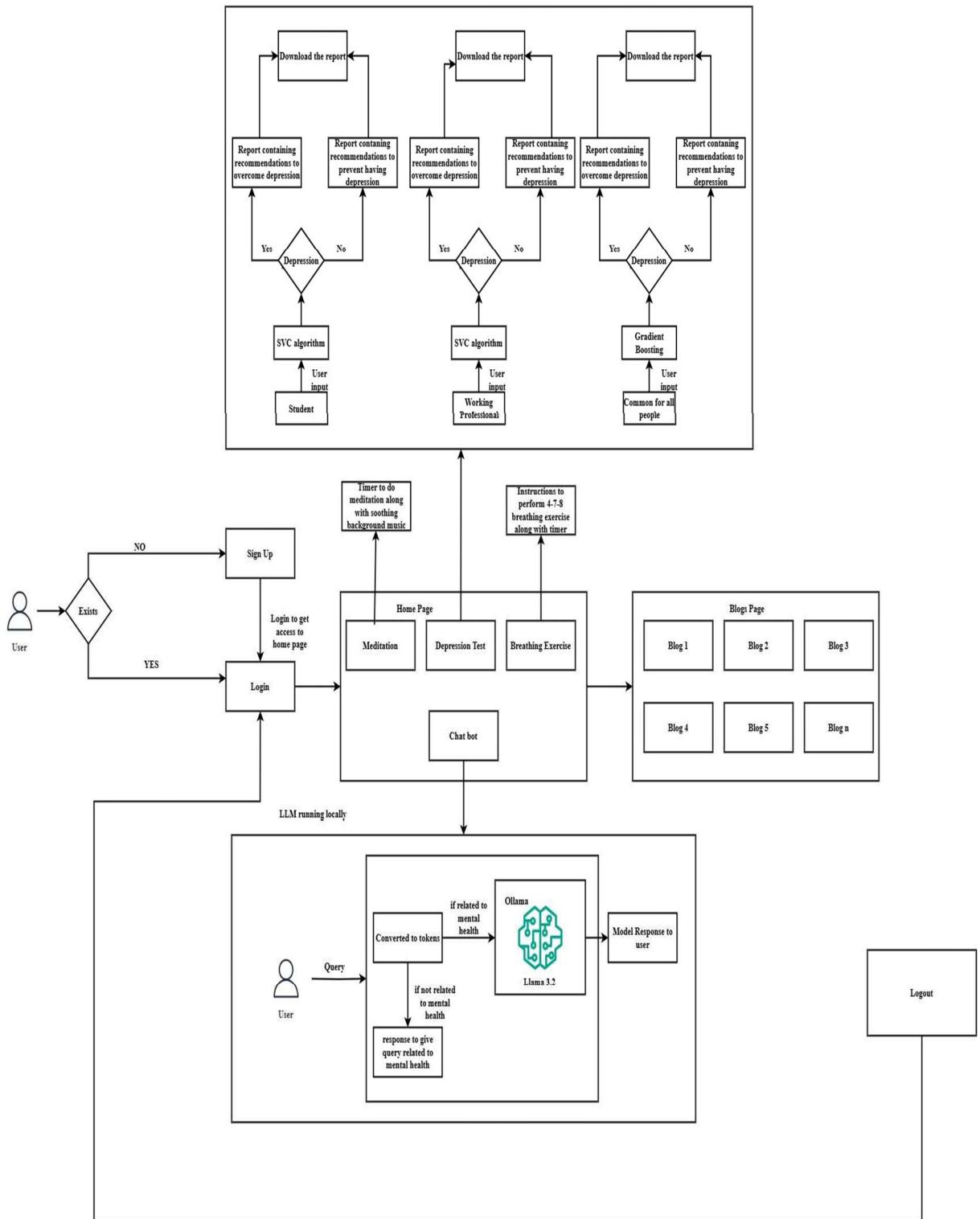


Fig. 4.1.1 System Architecture

4.2 MODULES:

1. User Management Module

Functionality:

- User **sign-up** and **login** functionality.
- User authentication and session management.

Importance:

- Ensures **secure access** to the platform.
- Provides **personalized experience** based on user history.

2. Depression Assessment Module

Functionality:

- Provides a **depression test**.
- Uses **SVC algorithm (for students and professionals)** and **Gradient Boosting algorithm (for general users)** to analyze responses.
- Generates **reports** with recommendations to prevent or overcome depression.
- Allows users to **download their reports**.

Importance:

- Offers **scientific and data-driven assessment** of mental health.
- Helps users understand their **mental health condition**.
- Provides **personalized recommendations** for mental well-being.

3. Chatbot Module

Functionality:

- Processes user queries using **Llama 3.2 LLM**.
- Identifies whether the query is **mental health-related** or not.
- Generates **empathetic and supportive responses**.
- Provides **mental health guidance** or general wellness tips.

Importance:

- Acts as a **24/7 virtual mental health assistant**.
- Offers **immediate emotional support**.
- Reduces **stigma** by allowing anonymous conversations.

4. Meditation Module

Functionality:

- Provides **guided meditation sessions** with background music.
- Includes a **timer** for meditation practice.

Importance:

- Helps users achieve **mental relaxation**.
- Reduces **stress and anxiety** through mindfulness.
- Encourages **regular meditation habits**.

5. Breathing Exercise Module

Functionality:

- Provides **4-7-8 breathing exercises** for relaxation.
- Includes **timed instructions** for correct practice.

Importance:

- Helps users **manage anxiety and panic attacks**.
- Improves **focus, sleep quality, and emotional regulation**.
- Provides **immediate stress relief techniques**.

6. Blog Module

Functionality:

- Displays **informative blogs** on mental health and well-being.
- Categorizes blogs for **easy navigation**.

Importance:

- Provides **educational resources** on mental health.
- Increases **awareness and self-help** opportunities.
- Encourages **community engagement**.

4.3 UML DIAGRAMS USED IN DESIGN

UML is a standardized modelling language consisting of an integrated set of diagrams, developed to help system and software developers for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for business modelling and other non-software systems. The UML is a very important part of developing object-oriented software and the software development process.

5.3.1 USECASE DIAGRAM:

Use-case diagrams model the behaviour of a system and help to capture the requirements of the system. Use case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. **Actors**

- **User:** The primary actor who interacts with the system by signing up, signing in, taking depression tests, and using self-care features.
- **Chatbot:** An assisting actor responsible for processing user queries and fetching relevant responses.
- **System:** The backend system that manages data storage, retrieval, and depression prediction.

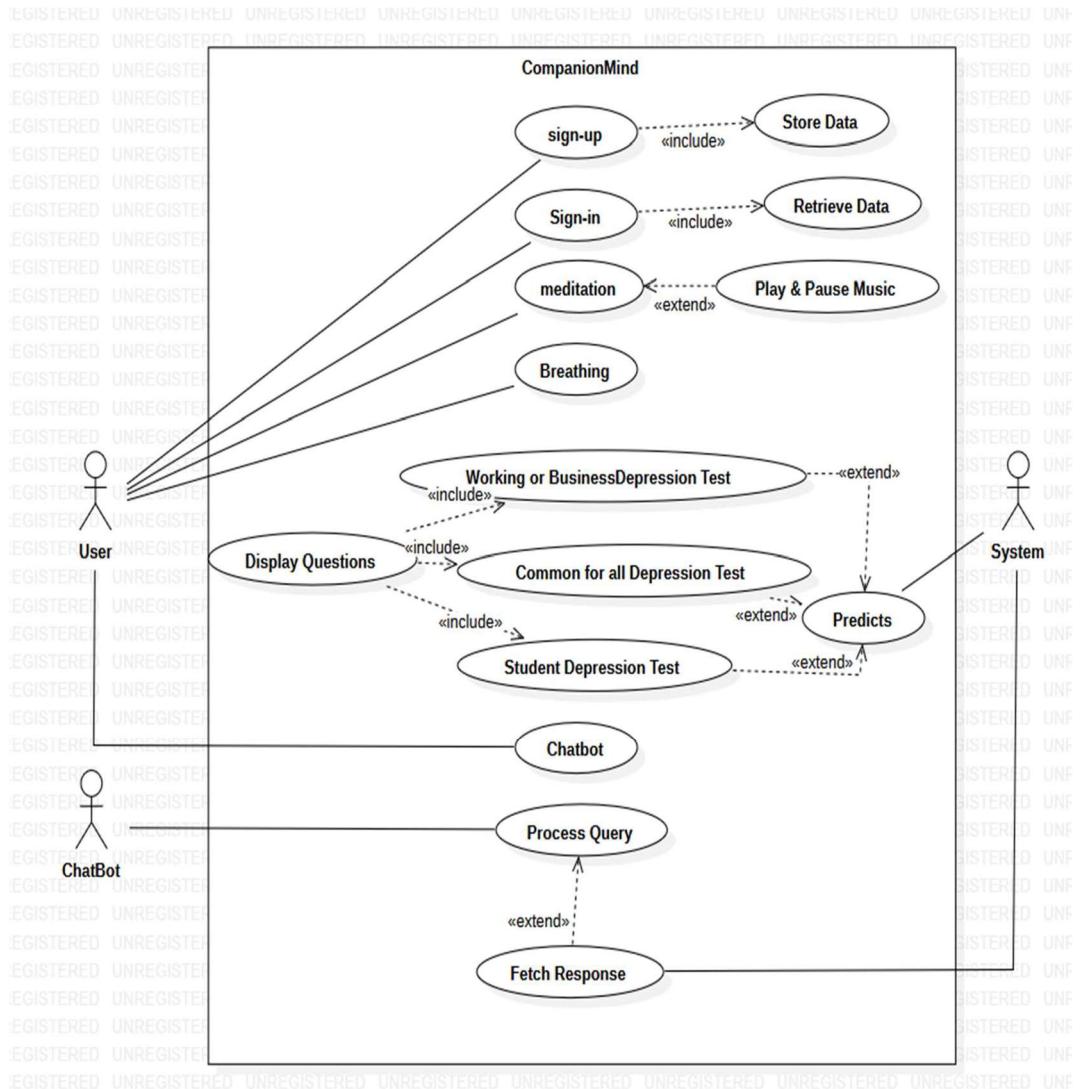


Fig. 4.3.1 Usecase Diagram of CompanionMind

Use Cases and Their Relationships

Core Functionalities

1. Sign-up & Sign-in:

- Users can register (sign-up) and log in (sign-in) to access the platform.
- These actions include storing and retrieving user data.

2. Meditation & Breathing Exercises:

- Users can engage in meditation and breathing exercises.
- Meditation extends to an additional feature that allows playing and pausing music for a better experience.

Depression Tests

3. Display Questions:

- Users interact with the system by answering depression assessment questions.
- This includes questions that are common for all depression tests.

4. Depression Test Types:

- **Student Depression Test:** Focuses on assessing depression levels among students.
- **Working or Business Depression Test:** Evaluates depression levels for professionals in a workplace setting.
- Both test cases extend to a **Prediction** module, which analyzes responses and determines the depression level.

Chatbot Functionality

5. Chatbot Query Processing:

- Users can interact with a chatbot for assistance.
- The chatbot processes queries and extends its functionality to fetch responses from the system.

5.3.2 CLASS DIAGRAM:

Class diagrams are fundamental to the object modelling process and model the static structure of a system. Depending on the complexity of a system, you can use a single class diagram to model an entire system, or you can use several class diagrams to model the components of a system. Class diagrams are the blueprints of your system or subsystem. You can use class diagrams to model the objects that make up the system, to display the relationships between the objects, and to describe what those objects do and the services that they provide. Class diagrams are useful in many stages of system design. Class diagrams depict the static structure of the system, focusing on classes, attributes, methods, and their relationships.

Classes and Their Attributes/Methods

1. User Class

- **Attributes:**

- username
- email
- password
- confirm password

- **Methods:**

- signup()
- login()
- storedata()
- retrievedata()

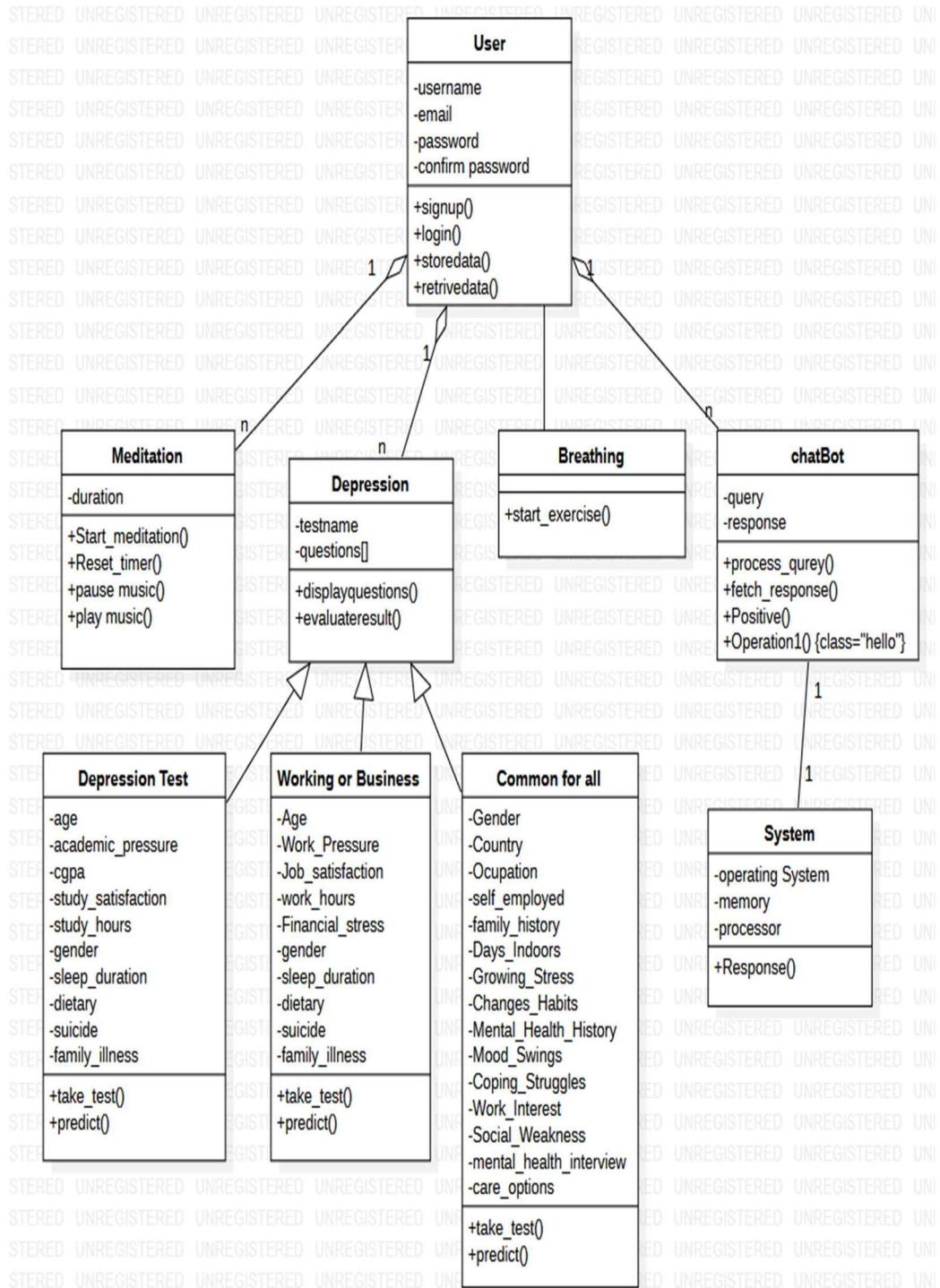


Fig. 4.3.2 Class Diagram of CompanionMind

2. Depression Class

- **Attributes:**
 - testname

- questions[]

- **Methods:**

- displayquestions()
- evaluateresult()

3. Depression Test (Specialized Tests)

- **Attributes:**

- age
- academic_pressure
- cgpa
- study_satisfaction
- study_hours
- gender
- sleep_duration
- dietary
- suicide
- family_illness

- **Methods:**

- take_test()
- predict()

4. Working or Business Depression Test

- **Attributes:**

- age
- work_pressure
- job_satisfaction

- work_hours
- financial_stress
- gender
- sleep_duration
- dietary
- suicide
- family_illness

- **Methods:**

- take_test()
- predict()

5. Common for All Depression Tests

- **Attributes:**

- gender
- country
- occupation
- self_employed
- family_history
- days_indoors
- growing_stress
- changes_habits
- mental_health_history
- mood_swings
- coping_struggles
- work_interest

- social_weakness
- mental_health_interview
- care_options

- **Methods:**

- take_test()
- predict()

6. Meditation Class

- **Attributes:**

- duration

- **Methods:**

- Start_meditation()
- Reset_timer()
- pause_music()
- play_music()

7. Breathing Class

- **Methods:**

- start_exercise()

8. ChatBot Class

- **Attributes:**

- query
- response

- **Methods:**

- process_query()
- fetch_response()

- Positive()
- Operation10() {class="hello"}

9. System Class

- **Attributes:**

- operating system
- memory
- processor

- **Methods:**

- Response()

3. Relationships in the Diagram

- The **User** class is associated with **Depression, Meditation, Breathing, and ChatBot** functionalities.
- **Depression** extends into **various test types (Student, Working, Common Tests)** to assess different user categories.
- The **ChatBot** interacts with the **System** to process and fetch responses.
- The **Meditation and Breathing** modules help users with self-care exercises

5.3.3 SEQUENCE DIAGRAM:

A sequence diagram consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction. Sequence diagrams illustrate the interactions between objects or components over time.

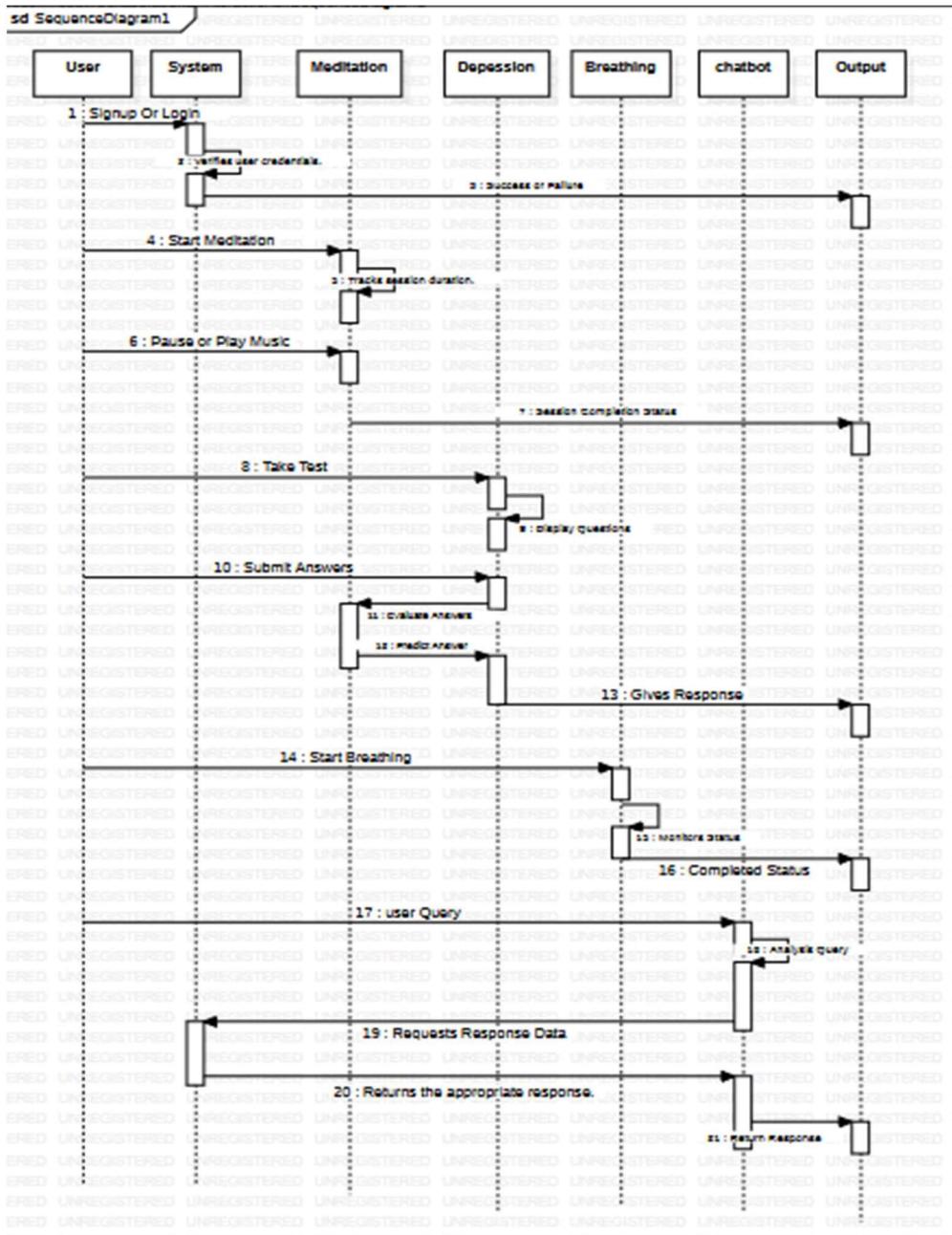


Fig. 4.3.3 Sequence Diagram of CompanionMind

Components Involved

The sequence diagram consists of the following key components:

- **User**: The individual interacting with the system.
- **System**: The backend system handling authentication and processes.
- **Meditation**: The module responsible for meditation and relaxation exercises.

- **Depression:** The module handling depression assessments.
- **Breathing:** The module managing breathing exercises.
- **Chatbot:** The AI assistant that processes queries and provides responses.
- **Output:** The final response or result displayed to the user.

Sequence of Actions

User Authentication

1. User initiates signup or login.
2. System verifies credentials.
3. System returns success or failure status.

Meditation Module

4. User starts meditation.
5. System tracks meditation duration.
6. User pauses or plays music.
7. System updates meditation status.

Depression Test

8. User initiates the depression test.
9. System displays relevant questions.
10. User submits answers.
11. System evaluates and processes the answers.
12. System provides a response based on evaluation.
13. System returns the result to the user.

Breathing Module

14. User starts breathing exercises.
15. System monitors the session.
16. System marks the session as completed.

Chatbot Interaction

17. User inputs a query.
18. Chatbot processes the query.
19. Chatbot requests response data.
20. System returns the appropriate response.
21. Chatbot displays the response to the user.

4.4 SUMMARY:

This chapter describes the System Architecture, UML diagrams like Use Case diagram, Class diagram, Sequence diagram which depicts the working of system both structurally and behaviourally. These diagrams help to understand the system and the interactions that take place in the system.

CHAPTER-5

SYSTEM IMPLEMENTATION

SYSTEM IMPLEMENTATION

System implementation is the process of putting the designed system into action or making it functional. It involves developing the system in a way that meets the requirements and objectives specified during the system analysis and design phase. It involves writing the code for the system based on the design specifications. In this chapter all the required libraries are explained briefly.

5.1 LIBRARIES

Libraries like streamlit, ollama, pickle, fpdf, numpy, pandas used for designing various parts of the system.

- **Streamlit:**

Streamlit is an open-source app framework in python language. It helps us create beautiful web apps for data science and machine learning in a little time. It is compatible with major python libraries such as scikit-learn, keras, PyTorch, latex, numpy, pandas, matplotlib, etc.

Application in CompanionMind:

Streamlit is used to build the complete UI for the entire system, providing a seamless and interactive user experience from start to finish. It handles user authentication, enabling a smooth login/signup process, and powers the dynamic form for depression assessment, allowing real-time input processing and instant personalized recommendations. The chatbot, integrated using Streamlit, facilitates empathetic conversations with users, offering AI-driven support. Additionally, Streamlit enhances mindfulness meditation sessions by incorporating background music and guided timers, creating a calming atmosphere. The breathing exercise module is designed with interactive instructions and animations to help users practice effectively. A dedicated blog section allows users to explore mental health resources, while real-time data visualization helps track their progress over time. With its simplicity, flexibility, and seamless Python integration, Streamlit ensures an intuitive and engaging mental health support platform, making advanced AI-powered features easily accessible to users.

- **Numpy:**

NumPy is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently.

Application in CompanionMind:

Data Preprocessing: NumPy is used for loading, preprocessing, and manipulating depression data sets for various category of people like Students, Working Professionals and common for all people before feeding them into machine learning models.

- **Pandas:**

Pandas is a powerful library for data manipulation and analysis. It provides data structures like DataFrame, which are ideal for handling structured data and time-series data.

Application in CompanionMind:

Data Handling: Pandas DataFrame is used to organize and manage depression data sets, facilitating data cleaning, exploration, and transformation.

- **Pickle:**

Pickle library in Python is used for serializing and deserializing objects, allowing data to be saved to a file and loaded later. It converts Python objects into a byte stream (pickling) and reconstructs them back into their original form (unpickling). This is useful for saving machine learning models, caching data, and transferring objects between programs.

Application in CompanionMind:

The pickle library is used for efficiently handling various components, including **Machine Learning models** for depression test, **user sign up and login, and report generation**. We use pickle to serialize and store trained machine learning models for depression assessment, allowing quick loading and inference without retraining. It is also utilized for saving user authentication data and session states, ensuring seamless logins and personalized experiences. Additionally, pickle helps in storing and retrieving dynamically generated reports, making it easier to provide users with personalized mental health recommendations. By leveraging pickle, we enhance the

system's efficiency, reduce computation time, and enable smooth data persistence across different functionalities.

- **Fpdf:**

Fpdf library in Python is used to generate PDF documents programmatically. It allows for creating custom reports, invoices, and certificates by adding text, images, tables, and formatting styles.

Application in CompanionMind:

Report Generation: Fpdf is used to generate personalized depression assessment reports, which users can download and share. It provides flexibility in structuring reports, embedding recommendations, and ensuring a professional layout, making it an essential tool for delivering structured and printable documents.

- **Ollama:**

Ollama library in Python is used for running and interacting with **Large Language Models (LLMs)** locally or via APIs. It enables seamless integration of AI-powered chatbots, text generation, and NLP tasks within applications.

Application in CompanionMind:

Chatbot: Ollama is utilized to power the empathetic chatbot, providing personalized support, answering user queries, and guiding them through mindfulness exercises. Its efficiency and ease of use make it a crucial component for delivering AI-driven mental health assistance.

5.2 SAMPLE SOURCE CODE :

```
tabs = st.tabs(["Home", "About", "Contact Us","blogs","chatBot"])

with tabs[0]:
    decor()

    image = Image.open('home.jpg')
    re_image = image.resize((1400,650))
    st.image(re_image)
```

```
with st.container():
```

```
    with st.container():
```

```
        st.markdown("<div style='font-size:32px; padding : 50px; text-align: center; font-weight: bold;'>Activities</div>",unsafe_allow_html=True)
```

```
        st.markdown("""
```

```
<style>
```

```
body {
```

```
    text-align: center;
```

```
    background: linear-gradient(to right, #ff7e5f, #feb47b); /* Gradient background */
```

```
    font-family: 'Poppins', sans-serif;
```

```
}
```

```
.stButton>button { text-align: center;
```

```
    background-color: linear-gradient(94.5deg, #F7A70D 0%, #FACA6E 73.52%, #F7A70D 106.59%);
```

```
    color: black;
```

```
    border-radius: 10%; font-size: 12px;
```

```
    transition: 0.3s ease;
```

```
}
```

```
.stButton>button:hover { text-align:
```

```
    center;
```

```
    background-color: #e0f7fa; /* Light cyan color */ }
```

```
.stForm {
```

```
    text-align: center; display: flex;
```

```
border-color : black;  
left-padding : 20px;  
flex-direction: row;  
align-items: center;  
}  
  
div[data-testid="stForm"] {  
background-color: #e0f7fa; /* Light cyan color */  
padding: 20px;  
margin: 10px;  
border-radius: 10px;  
box-shadow: 0px 10px 20px rgba(0, 0, 0, 0.2);  
transition: box-shadow 0.3s ease, transform 0.3s ease;  
transform: translateY(-10px);  
border: 2px solid black !important; /* Change 'red' to any color */  
border-radius: 40px; /* Optional: Round the corners */  
box-shadow: 0px 10px 20px rgba(0, 0, 0, 0.1);  
transform: translateY(-10px);  
}  
  
div[data-testid="stForm"]:hover {  
background-color: #f0f0f5;  
border: 2px solid black !important; /* Change 'red' to any color */  
border-radius: 40px; /* Optional: Round the corners */  
box-shadow: 50px 40px 60px rgba(225, 225, 225, 225);  
transform: translateY(-1px);  
}  
}
```

</style>

```
= Image.open('save.jpg')

re_imag"""",unsafe_allow_html=True)

col1,col2,col3 = st.columns(3)

with col1:

    with st.form('meditate'):

        st.markdown("<div style='font-size:24px; text-align: center; font-weight:bold;'>Meditation</div>",unsafe_allow_html=True)

        image e1 = image.resize((700,450))

        st.image(re_image1)

        st.markdown("<div style='font-size:18px; text-align: center; font-weight:bold; padding:15px; '>In today's fast-paced world, mental stability is often challenged by stress, anxiety, and constant distractions. Meditation is a simple yet powerful practice that helps bring clarity, emotional balance, and inner peace. </div>",unsafe_allow_html=True)

        submit = st.form_submit_button('Meditation Exercise') if

        submit:

st.session_state.page = "meditate" st.rerun()

with col2:

    with st.form('depress'):

        st.markdown("<div style='font-size:24px; text-align: center; font-weight:bold;'>Depression</div>",unsafe_allow_html=True)

        image = Image.open('depress.jpg') re_image1 =

        image.resize((700,450)) st.image(re_image1)

        st.markdown("<div style='font-size:18px; text-align: center; font-weight:bold;'>Depression can silently impact your thoughts, emotions, and daily life. But acknowledging it is the first step toward healing. A simple self-assessment can help you understand your mental health and take action toward a balanced, healthier mind.</div>",unsafe_allow_html=True)

        submit = st.form_submit_button('Depression Test') if

        submit:
```

```
st.session_state.page = "depress"
```

```
st.rerun()
```

with col3:

```
with st.form('Breath'):
```

```
    st.markdown("<div style='font-size:24px; text-align: center; font-weight:bold;'>Breathing</div>",unsafe_allow_html=True)
```

```
    image = Image.open('478.jpg')
```

```
    re_image1 = image.resize((700,450))
```

```
    st.image(re_image1)
```

```
    st.markdown("<div style='font-size:18px; text-align: center;font-weight:bold;'>In the fast-paced world we live in, stress, anxiety, and mental fatigue have become common challenges. Breathing exercises serve as a natural and powerful tool to regain mental stability, emotional balance, and overall well-being.</div>",unsafe_allow_html=True)
```

```
    submit = st.form_submit_button('Breathing Exercise')
```

if submit:

```
    st.session_state.page = "breath"
```

```
    st.rerun()
```

with tabs[1]:

```
    st.subheader("updated Soon")
```

with tabs[2]:

```
    st.subheader("updated Soon")
```

with tabs[3]:

```
    st.header("Blogs")
```

CHAPTER 6

SYSTEM TESTING

SYSTEM TESTING:

System testing is the process of testing our software or application regarding its performance, technical requirements, any technical glitches, and whether it is meeting the user requirements effectively and efficiently. Software testing not only focuses on finding faults but also helps in improving efficiency, accuracy, and usability.

6.1 INTRODUCTION:

Software bugs and glitches are inevitable in any software application. The bugs can be architecture, functions, the code itself, etc. The more complex the application the more bugs it will have. Unlike hardware, software that does not suffer from corrosion, wear-and-tear generally will not change until upgrades. So, once the software is shipped the design defects or bugs will be buried in and remain until activation.

Software testing can be divided into two steps:

Verification: It checks whether the software correctly implements a specific function

Validation: It checks whether the software build is reaching the customer's requirements.

6.2 PURPOSE OF TESTING:

- **Identifies defects early:** Developing complex applications can leave room for errors. Software testing identifies any issues and defects in the code so they can be fixed before the delivery of the software product.
- **Improves product quality:** Software testing helps the product pass quality assurance and meet the criteria and specifications defined by the user.
- **Increase customer trust and satisfaction:** Testing a product throughout its development lifecycle builds customer trust and satisfaction, as it provides visibility into the product's strong and weak points. By the time customers receive the product, it has been tried and tested multiple times and delivers on quality.
- **Detects security vulnerabilities:** Insecure application code can leave vulnerabilities that attackers can exploit. Since most applications are online today, they can be a leading vector for cyberattack and should be tested thoroughly during various stages of application development.
- **Helps with scalability:** Scalability test helps in increasing workload such as traffic, data volume, and transaction counts. It can also identify the point where an application might stop functioning and the reasons behind it, which may include meeting or exceeding a certain threshold, such as the total number of concurrent app users.

- **Saves money:** After launching the software it becomes difficult to find the bugs and errors and software patching is generally more expensive than testing during the development stages.

6.3 TYPES OF TESTING

- **UNIT TESTING:**

Unit testing is a method of testing individual units or components of a software application. It is typically done by developers and is used to ensure that the individual units of the software are working as intended. Unit testing is usually automated and is designed for specific parts of the code, such as a particular function or method. Unit testing is done at the lowest level of the software development process, where individual units of code are tested in isolation.

- **INTEGRATION TESTING:**

Integration testing mainly focuses on how different units of an application interact with each other. It is used to resolve any issues that arise when different components of the software application are combined. Integration testing is typically done after unit testing and before functional testing.

- **FUNCTIONAL TESTING**

The purpose of Functional tests is to test each function of the software application, by providing appropriate input, and verifying the output against the Functional requirements. Functional testing mainly involves black box testing and it is not concerned with the source code of the application. This testing checks the User Interface, APIs, Database, Security, Client/Server communication, and other functionality of the Application Under Test. The testing can be done either manually or using automation.

- **BETA TESTING:**

Beta Testing is performed by “real users” of the software application in “real environment” and it can be considered as a form of external user acceptance testing. It is the final test before shipping a product to the customers. Direct feedback from customers is a major advantage of Beta Testing. This testing helps to test products in customer’s environment. Beta version of the software is released to a limited number of end-users of the product to obtain feedback on the product quality. Beta testing reduces product failure risks and provides increased quality of the product through customer validation.

- **ALPHA TESTING:**

Alpha Testing is a type of acceptance testing; performed to identify all possible issues and bugs before releasing the final product to the end users. Alpha testing is carried out by the testers who are internal employees of the organization. The main goal is to identify the tasks that a typical user

might perform and test them. This testing comes under validation testing. It is a type of acceptance testing that is done before the product is released to customers. It is typically done by QA people.

- **SYSTEM TESTING:**

System testing is a level of software testing that evaluates a complete and integrated system to ensure that it meets specified requirements. It is a black-box testing technique where the system is tested as a whole, without knowledge of the internal workings of the system. The main objective of system testing is to identify and resolve any defects or issues that may exist in the system before it is deployed in the production environment. System testing can include functional, non-functional, regression, security, and performance testing to ensure the system meets all necessary requirements and performs as expected in various scenarios.

- **ACCEPTANCE TESTING**

Acceptance testing is a type of testing that determines whether a software product meets the customer's requirements and expectations. It is also known as user acceptance testing (UAT), customer acceptance testing (CAT), or end-user testing. The purpose of acceptance testing is to ensure that the software product meets the functional, usability, performance, and compatibility requirements as specified by the customer.

6.4 TESTING METHODS

1. BLACK BOX TESTING

Black box testing involves testing a system with no prior knowledge of its internal workings. A tester provides input and observes the output generated by the system under test. This makes it possible to identify how the system responds to expected and unexpected user actions, its response time, usability issues, and reliability issues. Black box testing is a powerful testing technique because it exercises a system end-to-end. Just like end-users “don't care” how a system is coded or architected, and expect to receive an appropriate response to their requests, a tester can simulate user activity and see if the system delivers on its promises. Along the way, a black box test evaluates all relevant subsystems, including UI/UX, web server or application server, database, dependencies, and integrated systems.

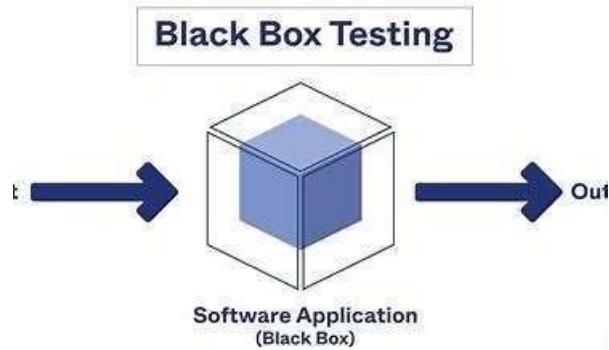


Fig 6.4.1 Black Box Testing

2. WHITE BOX TESTING

White box testing is an approach that allows testers to inspect and verify the inner workings of a software system—its code, infrastructure, and integrations with external systems. White box testing is an essential part of automated build processes in a modern Continuous Integration/Continuous Delivery (CI/CD) development pipeline. White box testing is often referenced in the context of Static Application Security Testing (SAST), an approach that checks source code or binaries automatically and provides feedback on bugs and possible vulnerabilities. White box testing provides inputs and examines outputs, considering the inner workings of the code.

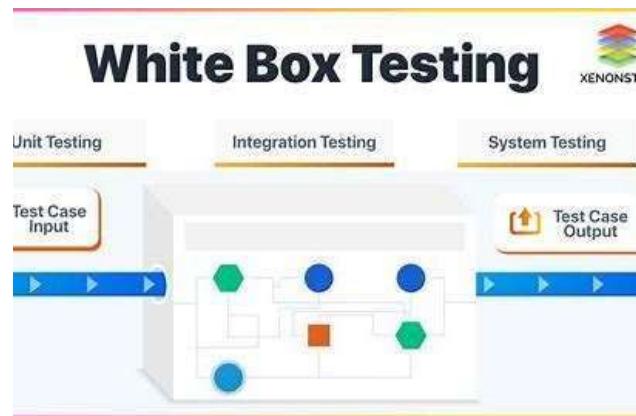


Fig 6.4.2 White Box Testing

6.5 TEST CASES

Test cases are a format of software testing to check if a particular application of the software is working or not. A test case consists of a certain set of conditions that need to be checked to test an application or software. If all the conditions are satisfied then the test case is passed if not the case is failed. The process of writing test cases can also help reveal errors or defects within the system.

Test case 1: Invalid User Authentication

In this scenario, we are trying to verify whether the user is a valid user or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
1	Invalid User Login	Error message should be displayed	Error message is displayed	Pass

Table 6.5.1 Test Case 1

Output:

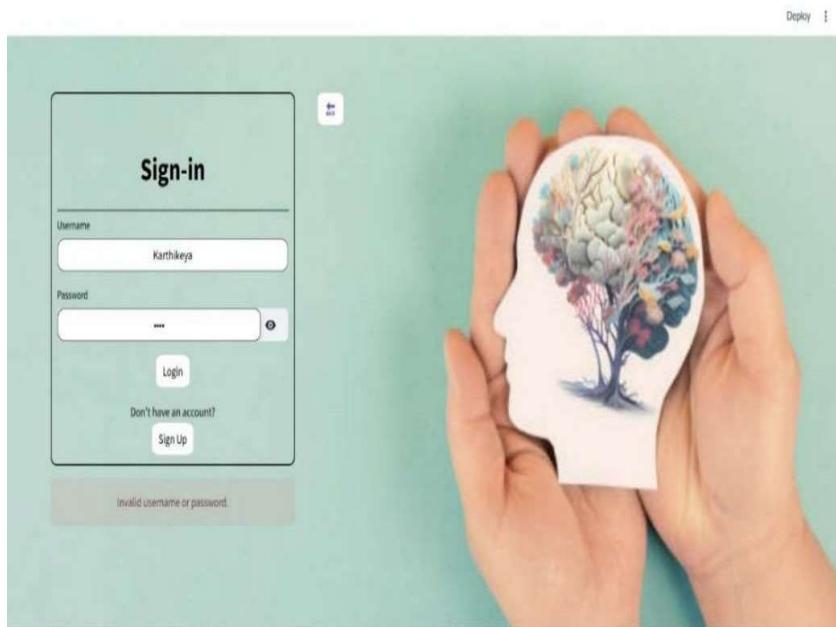
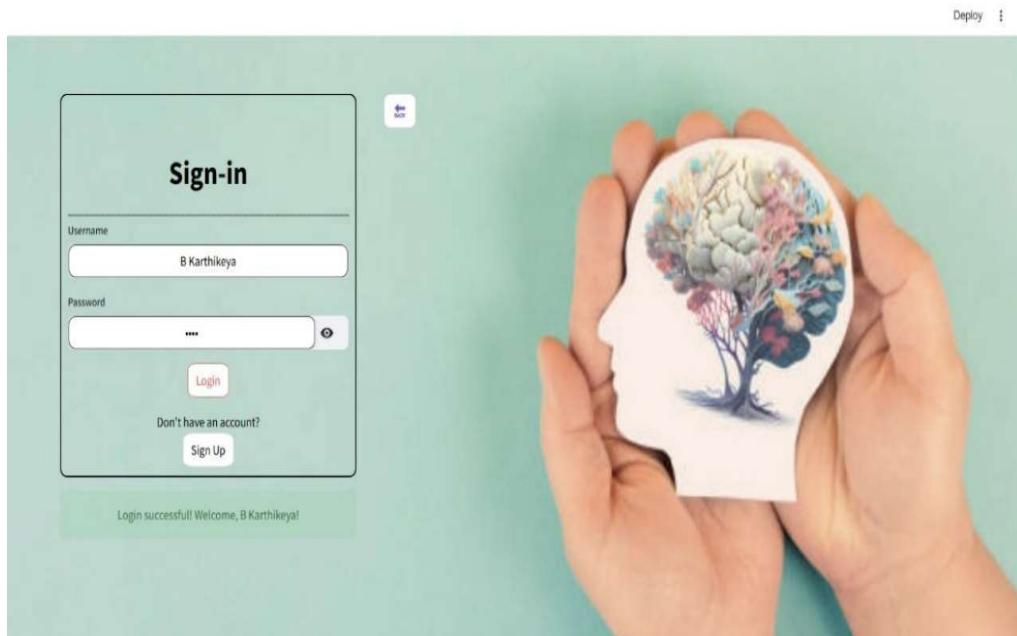


Fig. 6.5.1 Invalid User Authentication

Test case 2: Valid User Authentication

In this scenario, we are trying to verify whether the user is a valid user or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
2	Valid User Login	Allow access to home page	Access to home page is given	Pass

Table 6.5.2 Test Case 2**Output:****Fig. 6.5.2 Valid User Authentication**

Test case 3: Student Depression Test Flow

In this scenario, we are trying to verify whether the Student depression test flow is working correctly or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
3	Student takes depression test	Questions related to academic stress should be displayed	Questions are displayed correctly	Pass

Table 6.5.3 Test Case 3

Output:

The screenshot shows a user interface for a 'Student Depression Test'. At the top right, there are 'Deploy' and three-dot menu icons. Below the header, there are several input fields:

- Patient Name: A large text input field.
- Mobile Number: An input field containing '100000000' with minus and plus buttons for adjustment.
- Landline Number: A large text input field.
- self_employed: A dropdown menu showing 'nan'.
- select your country: A dropdown menu showing 'United States'.
- Occupation: A large text input field.
- Student Age: An input field containing '0' with minus and plus buttons for adjustment.
- Academic Pressure: A large text input field.

Deploy :

Academic Pressure
1

CGPA
0.00

Study Satisfaction (0 - Very Dissatisfied, 10 - Very Satisfied)
0

Studying in hours per day
0

Gender
Male

Sleep Duration per day
6

Dietary Habits
Healthy

Have you ever had suicidal thoughts? (Yes/No)
Yes

Family History of Mental Illness? (Yes/No)

Deploy :

0

Gender
Male

Sleep Duration per day
6

Dietary Habits
Healthy

Have you ever had suicidal thoughts? (Yes/No)
Yes

Family History of Mental Illness? (Yes/No)
Yes

Fig. 6.5.3 Student Depression Test Flow

Test case 4: Working Professional Test Flow:

In this scenario, we are trying to verify whether the Working Professional depression test flow is working correctly or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
4	Working Professional depression test	Questions related to work pressure should be displayed	Questions are displayed correctly	Pass

Table 6.5.4 Test Case 4**Output:**

The screenshot shows a user interface for a depression test. At the top right is a 'Deploy' button with three dots. Below it is the title 'Working/Business Depression Test' with a link icon. The form contains several input fields: 'Name' (text input), 'Mobile Number' (text input), 'Landline Number' (text input), 'select your country' (dropdown menu showing 'United States'), 'Occupation' (dropdown menu showing 'Employee/Business'), 'Employee/Business' (dropdown menu showing 'self-employed'), 'Employee Age' (number input with a range from 0 to 100), and 'Work Pressure' (number input with a range from 0.0 to 10.0). Each input field has a small downward arrow indicating it is a dropdown.

Job Satisfaction (0 - Very Dissatisfied, 5 - Very Satisfied)
0

Working in hours per day
0

Financial Stress
0

Employee Gender
Male

Sleep Duration per day of Employee
4

Dietary Habits of Employee
Healthy

Have you ever had suicidal thoughts or not? (Yes/No)
Yes

Family History of Mental Illness? (Yes/No)
Yes

Predicts

submit

Fig. 6.5.4 Working Professional Depression Test Flow

Test case 5: Common Depression Test Flow

In this scenario, we are trying to verify whether the common depression test flow is working correctly or not.

Test case No	Input	Expected Behavior	Observe Behavior	Status
5	Common test for all users	Generic mental health question should be displayed	Questions are displayed correctly	Pass

Table 6.5.5 Test Case 5**Output:**

Deploy ⋮

Common For All Depression Test

Name

Age
 - +

Mobile Number
 - +

Landline Number

Gender
 ▾

select your country
 ▾

Occupation
 ▾

self_employed
 ▾

Deploy ⋮

family_history
 ▾

Days_Outdoors
 ▾

Growing_Stress
 ▾

Changes_Habits
 ▾

Mental_Health_History
 ▾

Mood_Swings
 ▾

Coping_Struggles
 ▾

Work_Interest
 ▾

Social_Weakness

The screenshot displays a user interface for a depression test. At the top right, there is a 'Deploy' button with a three-dot menu icon. Below it, a series of dropdown menus are shown:

- Coping_Struggles: Medium
- No
- Work_Interest: No
- Social_Weakness: Yes
- mental_health_interview: No
- care_options: Not sure

At the bottom left, there are two buttons:

- Predicts
- submit2

Fig. 6.5.5 Common Depression Test Flow

Test case 6: Identifying Support Needs

In this scenario, we are trying to verify whether the user responses indicating emotional distress.

Test case No	Input	Expect ed Behavi or	Observ ed Behavi or	Status
6	User selects responses indicating emotional distress	Model suggests seeking support	Model correctly identifies support needs	Pass

Table 6.5.6 Test Case 6**Output:**

Deploy :

Student Depression Test

Patient Name
Sruthi

Mobile Number
9815648980 - + ⚡

Landline Number
8535645154784

self_employed
Yes

select your country
India

Occupation
student

Student Age
19 - +

Academic Pressure

Academic Pressure
5

CGPA
6.65 - +

Study Satisfaction (0 - Very Dissatisfied, 10 - Very Satisfied)
2 - + ⚡

Studying in hours per day
14 - +

Gender
Female

Sleep Duration per day
4

Dietary Habits
Unhealthy

Have you ever had suicidal thoughts? (Yes/No)
Yes

Family History of Mental Illness? (Yes/No)
No

Fig. 6.5.6.1 Identifying Support Needs

The screenshot shows a user profile at the top: "Female". Below it is a form titled "Identifying Support Needs". The form includes fields for "Sleep Duration per day" (set to 4), "Dietary Habits" (set to "Unhealthy"), "Have you ever had suicidal thoughts? (Yes/No)" (set to "Yes"), and "Family History of Mental Illness? (Yes/No)" (set to "Yes"). A red-bordered "Predict" button is visible. A pink message bar at the bottom states: "The model predicts a high likelihood of depression. Please seek professional help." A "submit" button is located at the bottom left.

Fig. 6.5.6.2 Identifying Support Needs

Test case 7: Recognizing Positive Well – being

In this scenario, we are trying to verify the user selecting responses indicating emotional stability.

Test case No	Input	Expected Behavior	Observed Behavior	Status
7	User selects response indicating emotion al stability	Model suggest s a positiv e well – being status	Model correctl y recogni zes well - being	Pass

Table 6.5.7 Test Case 7

Output:

The screenshot shows a form titled "Student Depression Test". It includes fields for Patient Name (Karthikeya), Mobile Number (9654156850), Landline Number (845621534589), self-employed status (No), country selection (India), occupation (Student), student age (10), and academic pressure. A "Deploy" button is visible in the top right corner.

Fig. 6.5.7.1 Recognizing Positive Well - being

The screenshot shows a form for identifying positive well-being. It includes fields for gender (Male), sleep duration (6 hours), dietary habits (Healthy), suicidal thoughts (No), family history of mental illness (No), and a "Predict" button. A message at the bottom states: "The model predicts no significant signs of depression." A "submit" button is located at the bottom left.

Fig. 6.5.7.2 Identifying Positive Well – being

Test case 8: Generating Personalized Reports with well – being Recommendations

In this scenario, we are trying to verify the user receives personalized reports with recommendations.

Test case No	Input	Expected Behavior	Observed Behavior	Status
8	User completes depression test and receives a result indicating need for additional support	Reports provides supportive recommend ations such as self – care activities etc.	Report correctly includes supportive recommend ations	Pass

Table 6.5.8 Test Case 8**Output:**
Fig 6.5.8.1: Generating Personalized Reports with Well-being Recommendation

Dietary Habits:	Unhealthy
History of Suicidal Thoughts:	Yes
Family History of Mental Illness:	Yes
Summary:	Sruthi may be experiencing heightened stress or other mental health concerns. Further evaluation is recommended.
Recommendations:	<ol style="list-style-type: none">1. Consider professional counseling for stress management.2. Engage in mindfulness and relaxation activities.3. Build a strong support network and seek help if needed.

Fig 6.5.8.2: Generating Personalized Reports with Well-being Recommendation

Test case 9: Generating Personalized Reports with general tips

In this scenario, we are trying to verify the user receives personalized reports with general tips.

Table 6.5.9 Test Case 9**Output :**

Test case No	Input	Expected Behavior	Observed Behavior	Status
9	User completes the depression test and receives a result indicating good mental health	Report provides general well-being such as mindfulness exercises, relaxation techniques, and healthy lifestyle suggestions	Report correctly includes well-being tips	Pass

CompanionMind
mental health report

Patient Information:			
Name: Karthikeya Age: 10 Gender: Male Country: India			
Mobile Number: 9654156850 Landline Number: 845621534589			
Occupation: Student Self-Employed: No			
Mental Health and Lifestyle Information			
Academmic Pressure Level:			
3			
cgpa):			
7.78			
Studying Hours per Day:			
10			
Study Satisfaction Level:			
8			
Sleep Duration per Day:			
6			

Fig 6.5.9.1: Generating Personalized Reports with General tips

Dietary Habits:

Healthy

History of Suicidal Thoughts:

No

Family History of Mental Illness:

No

Summary:

Karthikeya appears to have stable mental health. There are no major concerns detected based on the provided data.

Recommendations:

1. Maintain a balanced work-life routine.
2. Engage in social and physical activities to support well-being.
3. Continue monitoring stress levels and seek help if necessary.

Fig 6.5.9.2: Generating Personalized Reports with General tips

Test case 10: Chatbot Providing Helpful Responses

In this scenario, we are trying to verify whether the chatbot provides helpful responses or not.

Test case No	Input	Expected Behavior	Observe d Behavior	Status
10	User asks “How can I manage Stress better?”	Chatbot provides stress management tips	Chatbot gives an appropri ate response	Pass

Table 6.5.10 Test Case 10

Output:

The screenshot shows the CompanionMind AI chatbot interface. At the top, there are two small robot icons flanking the text "CompanionMind – Your AI Chat Companion". Below this is a sub-header "Your personal AI-powered support for mindfulness, well-being, and guidance." A "Deploy" button with a three-dot menu icon is in the top right corner. The main content area has a light gray background. It starts with a question "How can I manage Stress better?" followed by a yellow info icon. Below it is a list of four numbered tips for managing stress: 1. Identify your stressors, 2. Exercise regularly, 3. Practice relaxation techniques, and 4. Get enough sleep. A horizontal dashed line follows this. At the bottom is a search bar with the placeholder "How can I help you today?" and a right-pointing arrow icon.

Fig. 6.5.10.1 Chatbot Providing helpful Responses

This screenshot continues from Fig. 6.5.10.1. It shows a list of ten numbered tips for managing stress: 6. Stay connected with others, 7. Take breaks, 8. Prioritize self-care, 9. Set boundaries, and 10. Seek professional help. Below this is a section titled "Additionally, here are some quick stress-relief techniques you can try:" which lists six items: Deep breathing exercises, Progressive muscle relaxation, Mindfulness meditation / Visualization, Journaling, Listening to calming music, and Taking a short walk outside. A note at the bottom states: "Remember, managing stress is an ongoing process that requires patience, self-awareness, and practice. Experiment with different techniques to find what works best for you." The interface includes a "Deploy" button and a search bar at the bottom.

Fig. 6.5.10.2 Chatbot Providing helpful Responses

Test case 11: Chatbot Offering Encouraging Affirmations

In this scenario, we are trying to verify whether the chatbot provides encouraging affirmations or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
11	User clicks on “Give me a positive affirmation”	Chatbot generates an uplifting affirmation	Affirmation is generated correctly by chatbot	Pass

Table 6.5.11 Test Case 11

Output:

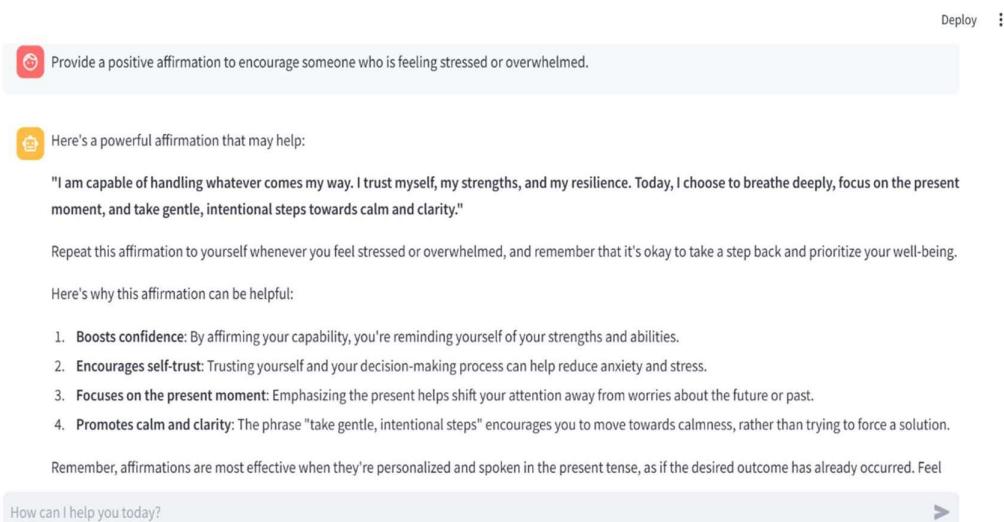


Fig. 6.5.11 Chatbot Offering Positive Affirmations

Test case 12: Chatbot Delivering Guided Relaxation Sessions

In this scenario, we are trying to verify whether the chatbot provides a guided meditation or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
12	User clicks “Give me a guided meditation”	Chatbot generates a short relaxation session	Meditation script is generated correctly by chatbot	Pass

Table 6.5.12 Test Case 12

Output:

The screenshot shows a user interaction with the CompanionMind AI companion. The user has typed "Provide a 3-minute guided meditation script to relax and reduce stress." The AI has responded with a message: "Here's a 3-minute guided meditation script to help you relax and reduce stress: "Breathing Space: A 3-Minute Meditation for Relaxation"" followed by "[Soothing music starts playing]". The AI then guides the user through a breathing exercise: "Guide (in a calm, gentle voice): Welcome to this moment of stillness. Find a comfortable seated or lying position, close your eyes, and take a deep breath in through your nose... and out through your mouth." It includes a pause: "[Pause for 30 seconds]" and continues with another guide: "Guide: Imagine yourself standing on the edge of a peaceful lake on a warm summer day. Feel the soft grass beneath your feet and the gentle breeze rustling your hair. With each breath, allow yourself to relax further into this serene environment." Finally, it includes another pause: "[Pause for 45 seconds]". At the bottom, there is a text input field with the placeholder "How can I help you today?" and a right-pointing arrow button.

Fig. 6.5.12.1 Chatbot Delivering Guided Relaxation Sessions

CompanionMind – A friendly AI companion for mental well-being using Llama

[Pause for 1 minute]

Guide: Imagine a warm, soothing light beginning to fill your body. Starting at the crown of your head, let this gentle radiance spread down through your face, neck, and shoulders, calming any muscles that may be holding tension. As it reaches your heart, allow yourself to feel a sense of peace and understanding.

[Pause for 30 seconds]

Guide: Now, visualize roots growing from the base of your spine, deep into the earth below. These roots are strong and sturdy, anchoring you firmly in the present moment. With each breath, feel these roots growing deeper, connecting you to the stability and calmness beneath you.

[Pause for 30 seconds]

Guide: As you continue to breathe deeply, repeat the following phrase to yourself: "I am safe, I am calm, I am relaxed." Allow these words to sink into your mind and body, filling you with a sense of inner peace.

[Soothing music continues to play for an additional 15 seconds before fading out]

Guide: When you're ready, take one final, deep breath in... and slowly open your eyes. Take a moment to notice how you feel before returning to your day with renewed calmness and clarity.

How can I help you today? >

Fig. 6.5.12.2 Chatbot Delivering Guided Relaxation Sessions

Test case 13: Ensuring Chatbot Responds Only to Mental Health-Related Queries

In this scenario, we are trying to verify whether the chatbot provides helpful responses or not.

Test case No	Input	Expected Behavior	Observed Behavior	Status
13	User asks a non – mental health related question (e.g., “Who won the last football match?”)	Chatbot politely declines and encourage s mental health related discussion	Chatbot restricts response to mental health topics	Pass

Table 6.5.13 Test Case 13

Output:

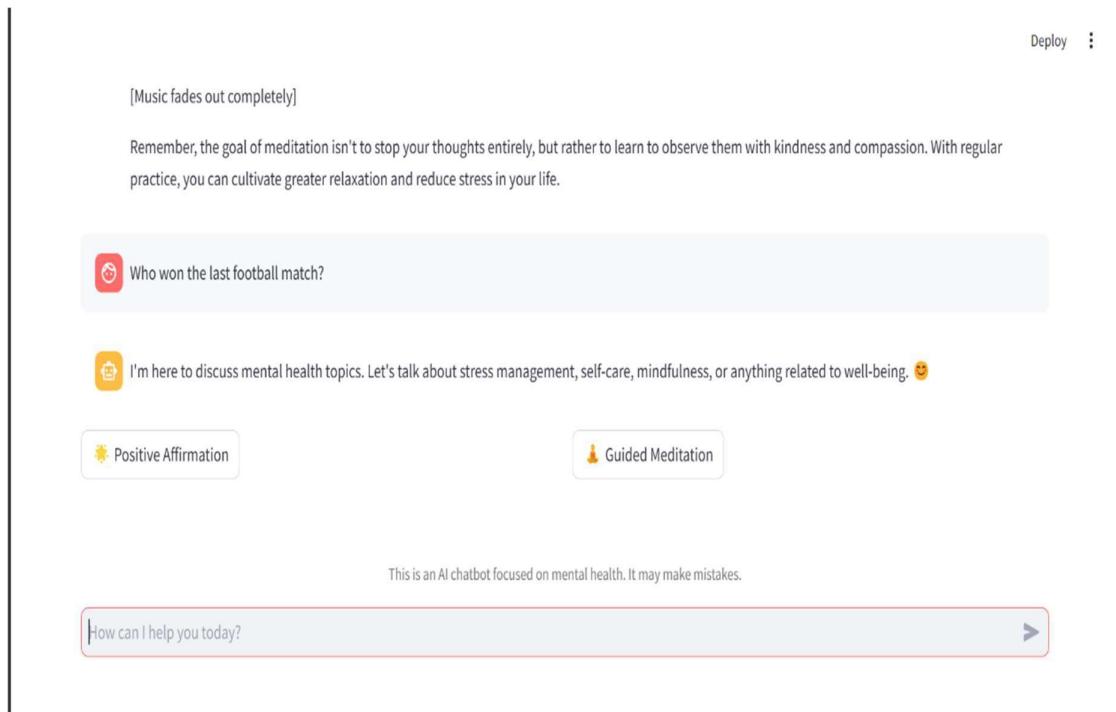


Fig. 6.5.13 Ensuring Chatbot Responds Only to Mental Health-Relate Queries

Test case 14: Logging out Smoothly

In this scenario, we are trying to verify whether the user can logout using logout button.

Test case No	Input	Expected Behavior	Observed Behavior	Status
14	User clicks Logout button	System logs out the user and redirect ts to website main page	User is success fully logged out	Pass

Table 6.5.14 Test Case 14

Output:

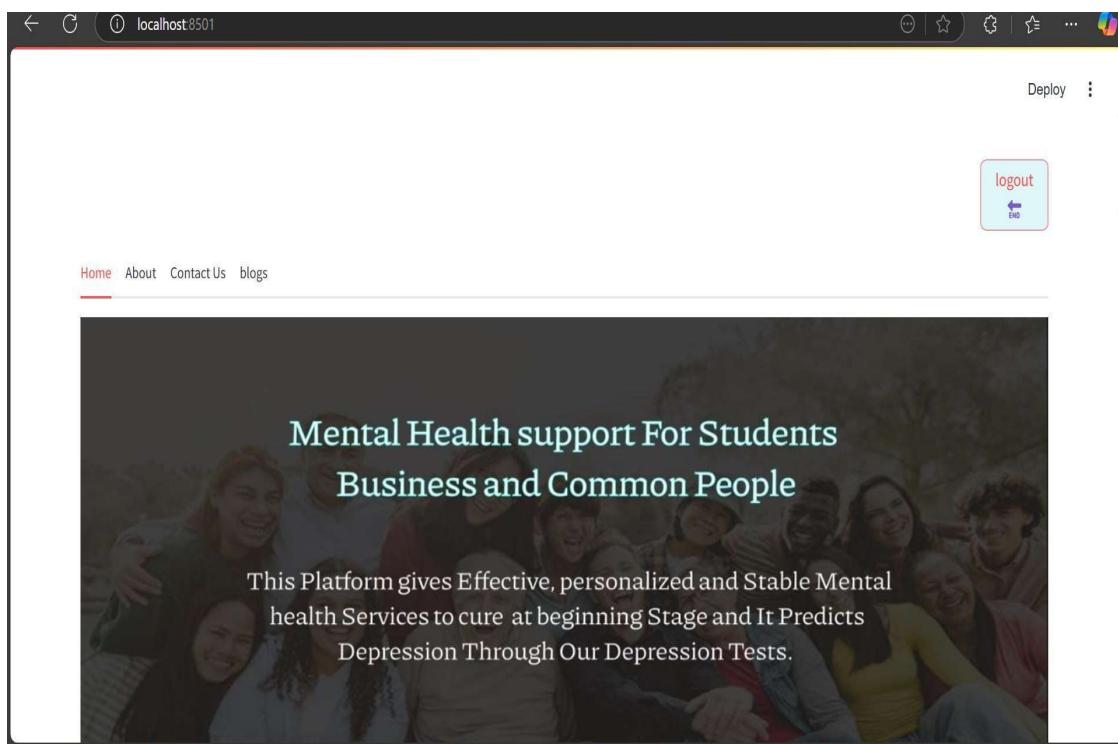


Fig. 6.5.14.1 Clicking on Logout Button

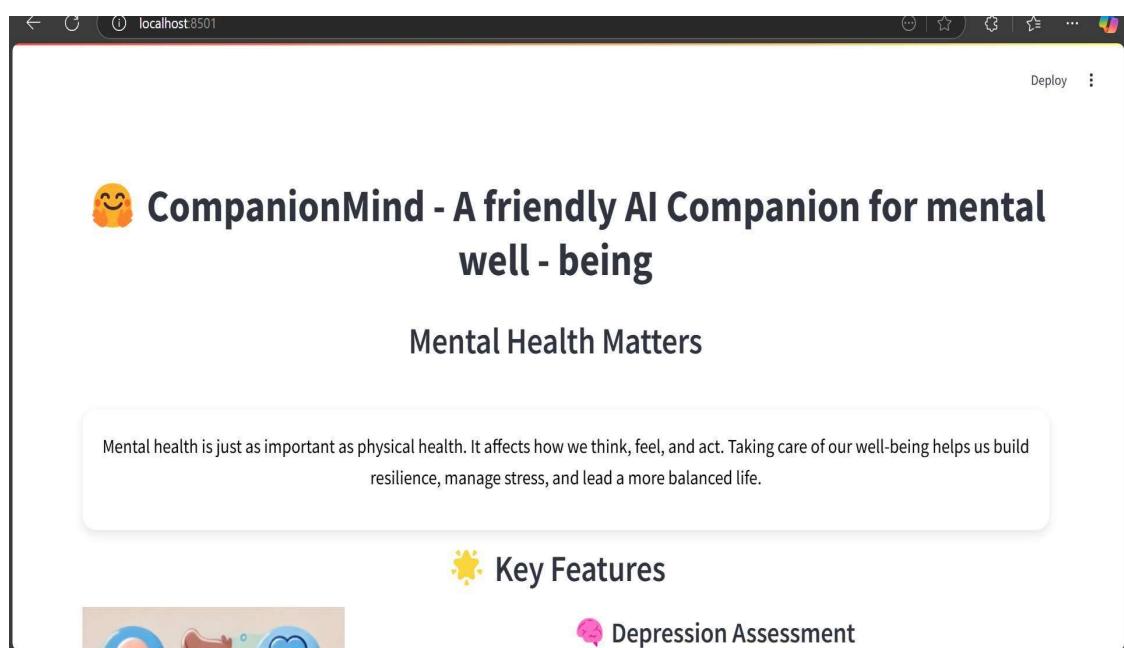


Fig. 6.5.14.2 Redirected to Website Main Page

6.6 SUMMARY:

In this chapter, we have outlined various test cases to ensure the smooth functionality of the system. These test cases cover different aspects, including user authentication, well-being assessments, chatbot interactions, report generation, and overall user experience. Each test case is designed to verify that the system provides accurate responses, personalized recommendations, and a seamless user journey.

+

CHAPTER 7

SYSTEM RESULTS

SYSTEM RESULTS

System results are the outcome of the whole process of software testing life cycle. The results thus produced, offer an insight into the deliverables of a software project, significant in representing the status of the project to the stakeholders. The analysis of system results is an important part of system evaluation, as it provides insight into the effectiveness and efficiency of the system and helps to identify areas for improvement.

7.1 User Interface:

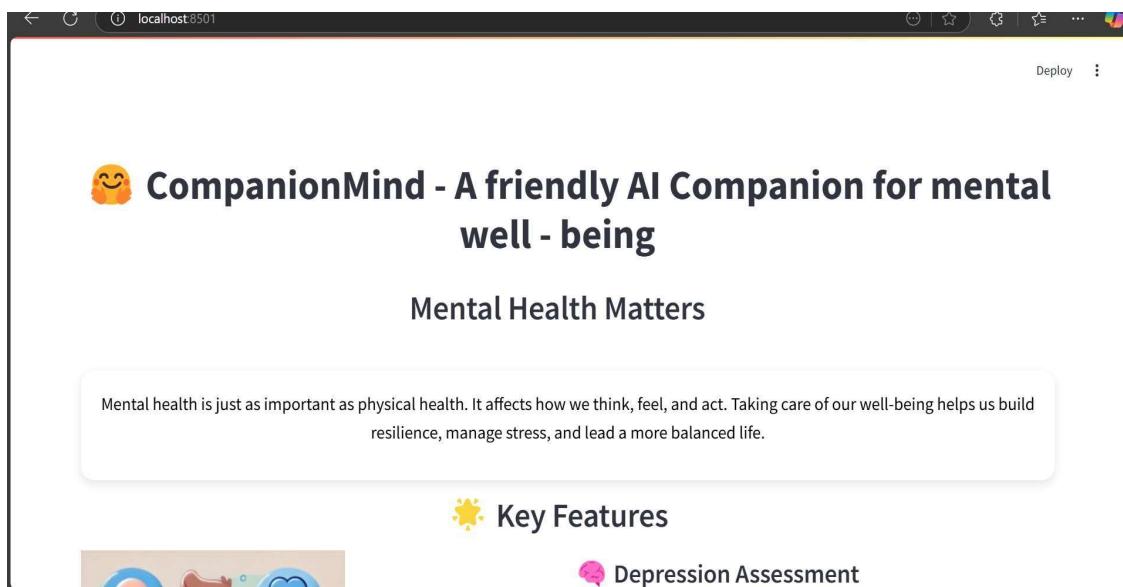


Fig. 7.1.1 CompanioinMind Main Page

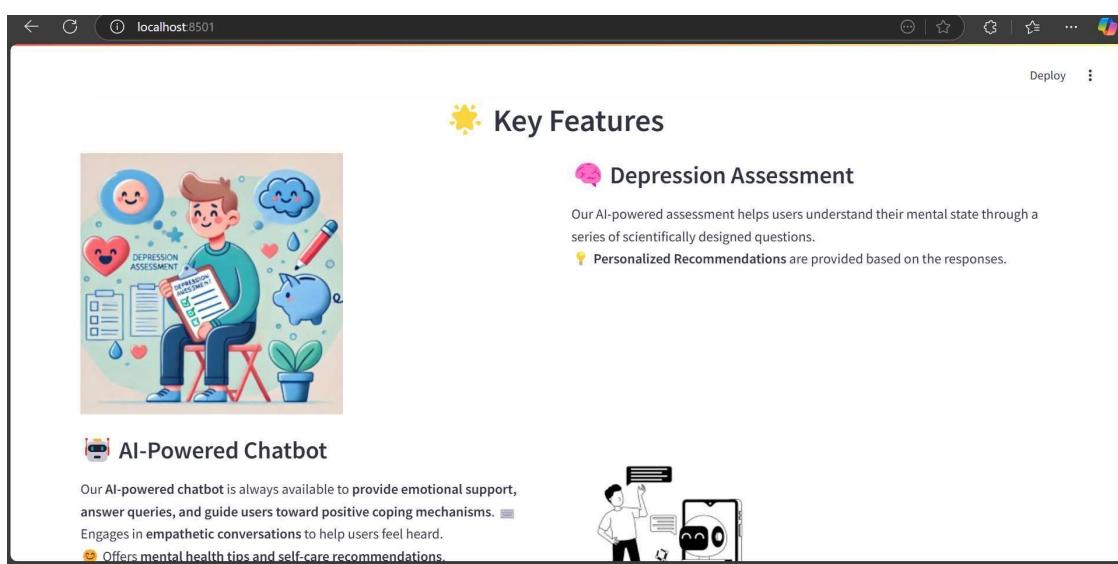


Fig. 7.1.2 Key Features – Depression Assessment, AI – Powered Chatbot

CompanionMind – A friendly AI companion for mental well-being using Llama

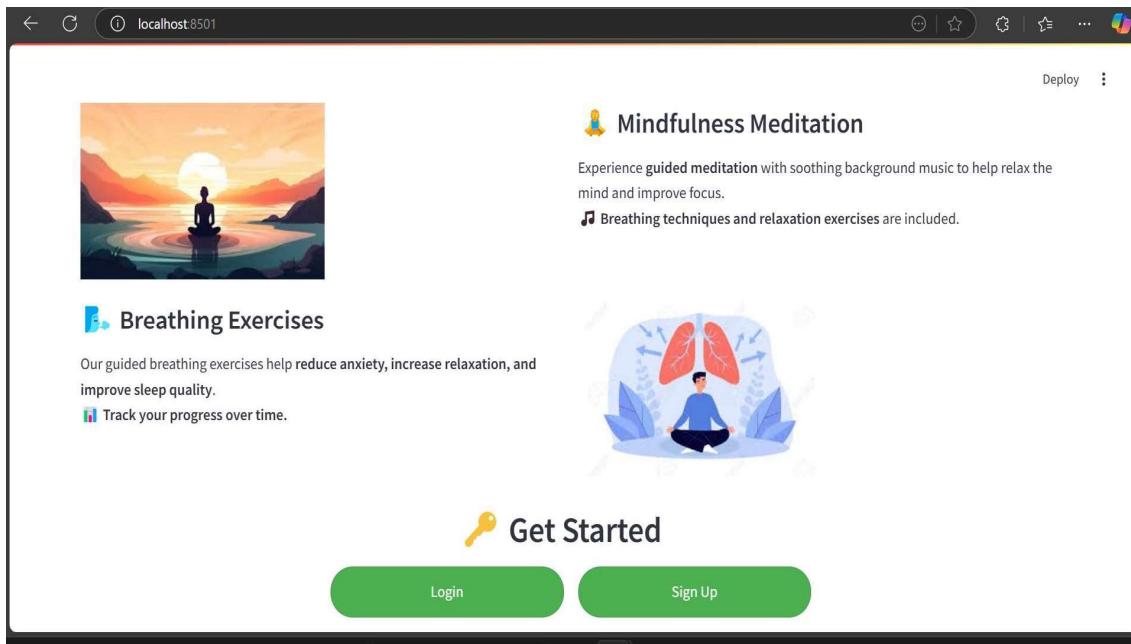


Fig. 7.1.3 Key Features – Mindfulness Meditation, Breathing Exercise

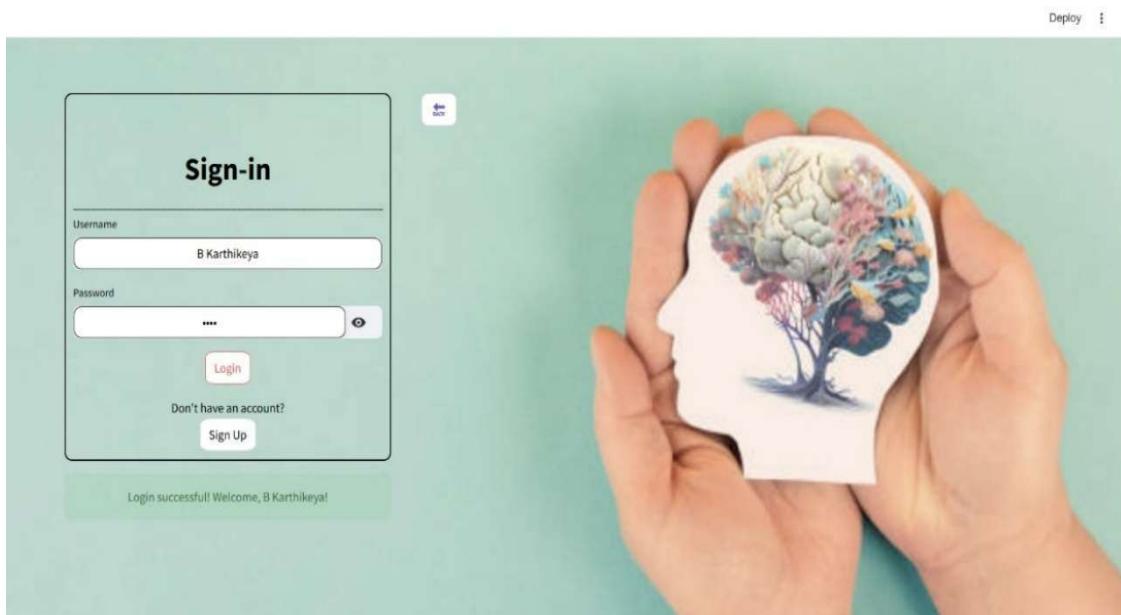


Fig. 7.1.4 Sign in Page of CompanionMind

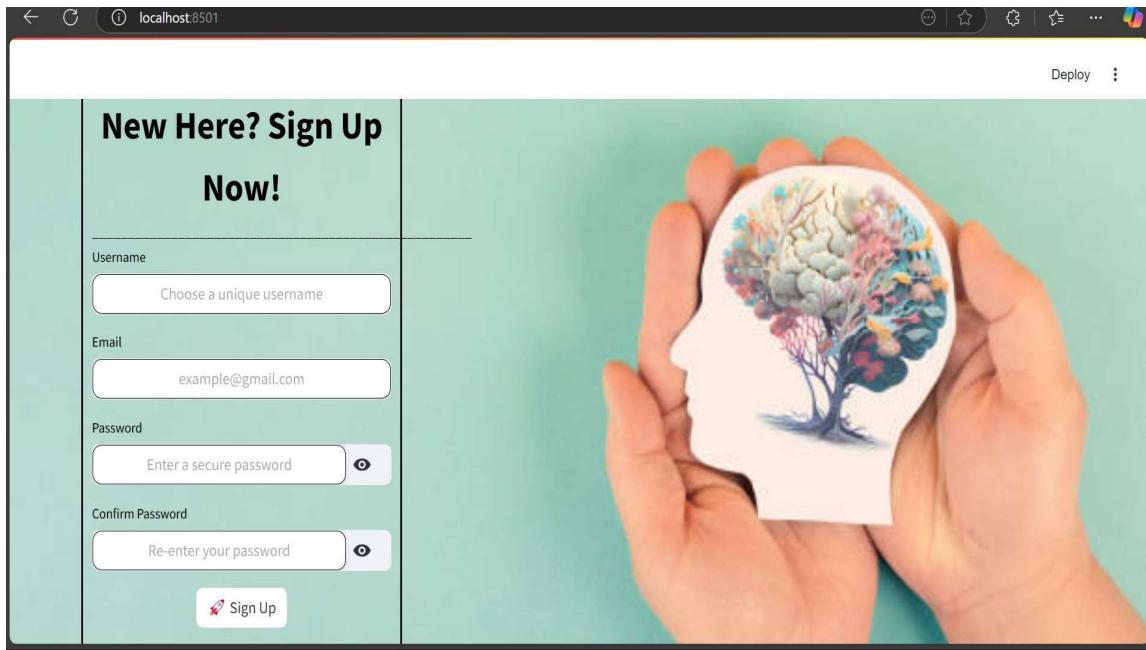


Fig. 7.1.5 Sign Up Page of CompanionMind

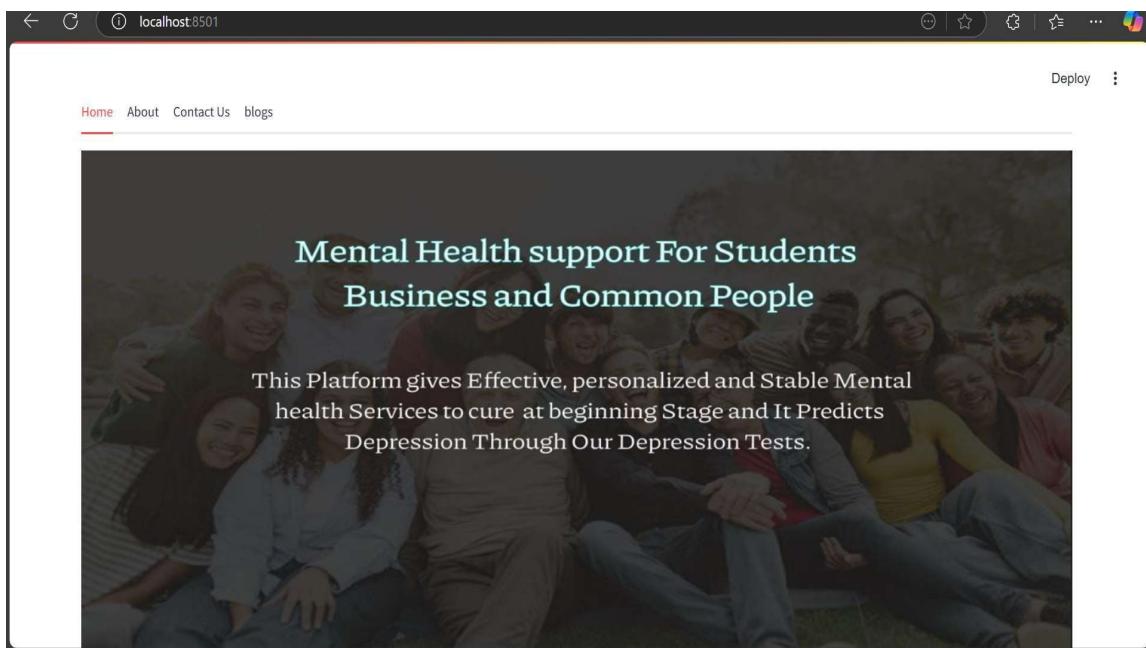


Fig. 7.1.6 Home Page of CompanionMind After Sign in

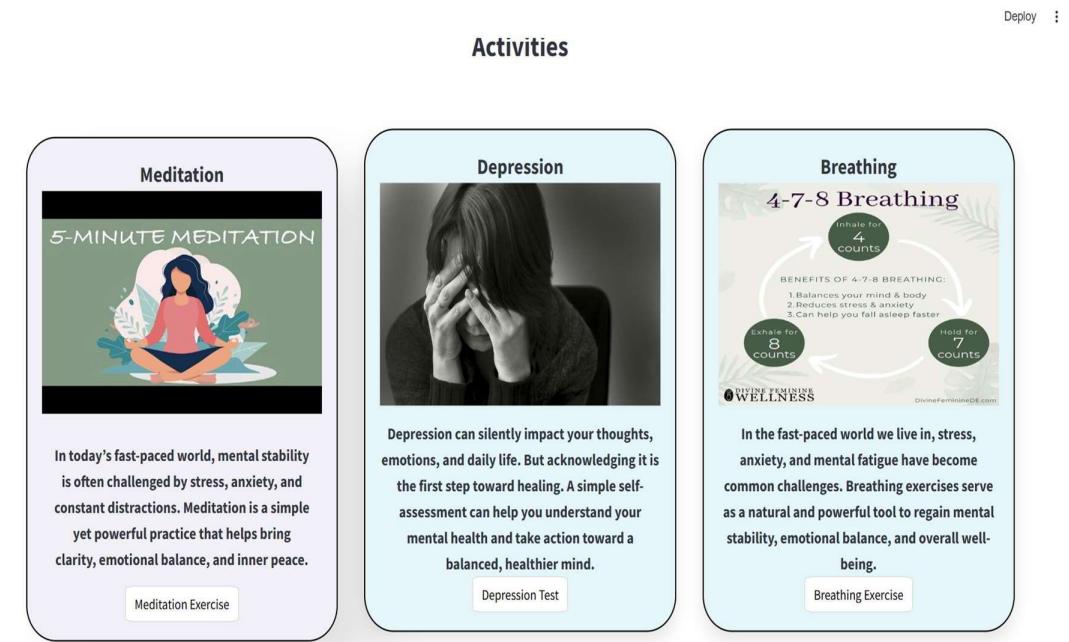


Fig. 7.1.7 Activities of CompanionMind

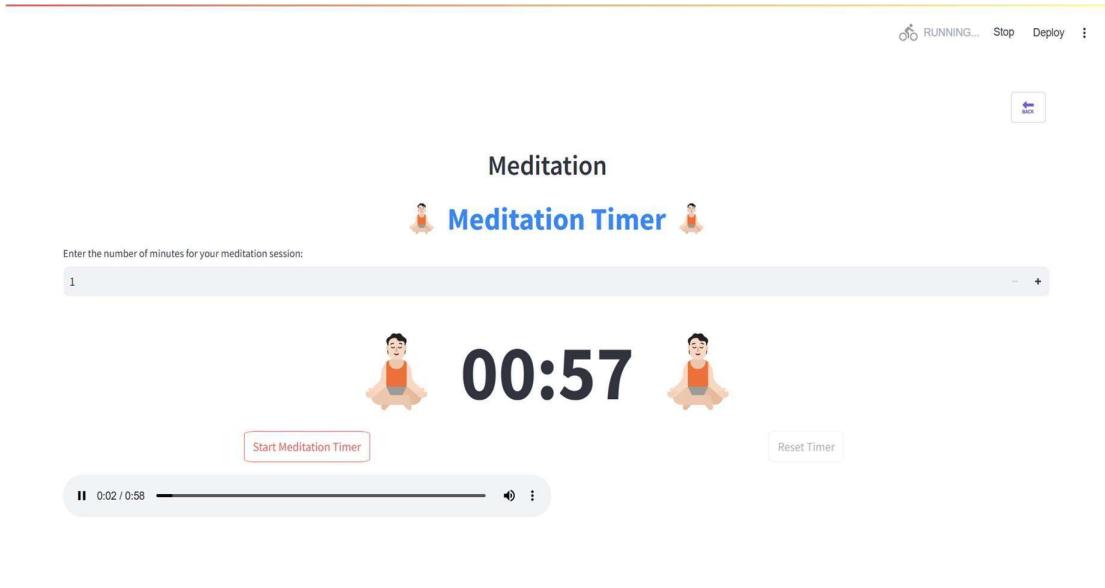


Fig. 7.1.8 Meditation timer started along with soothing Background Music

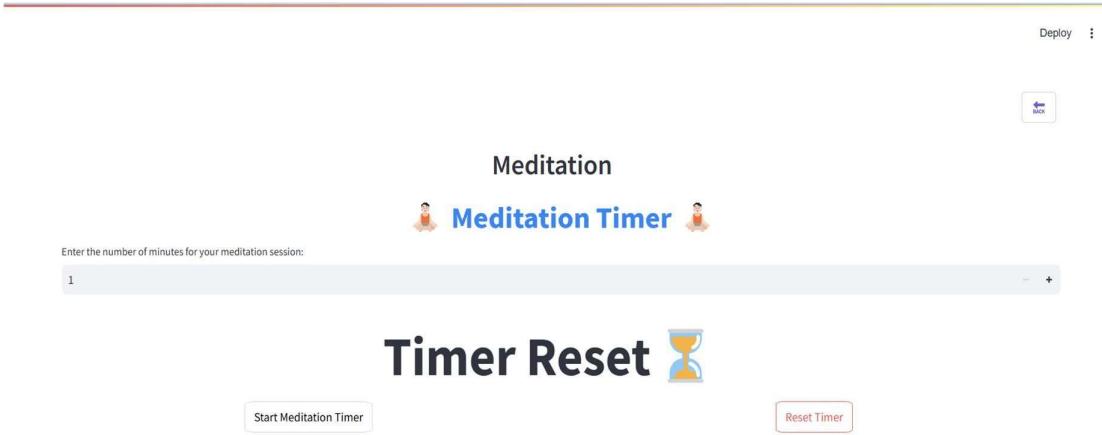


Fig. 7.1.9 Meditation timer reset

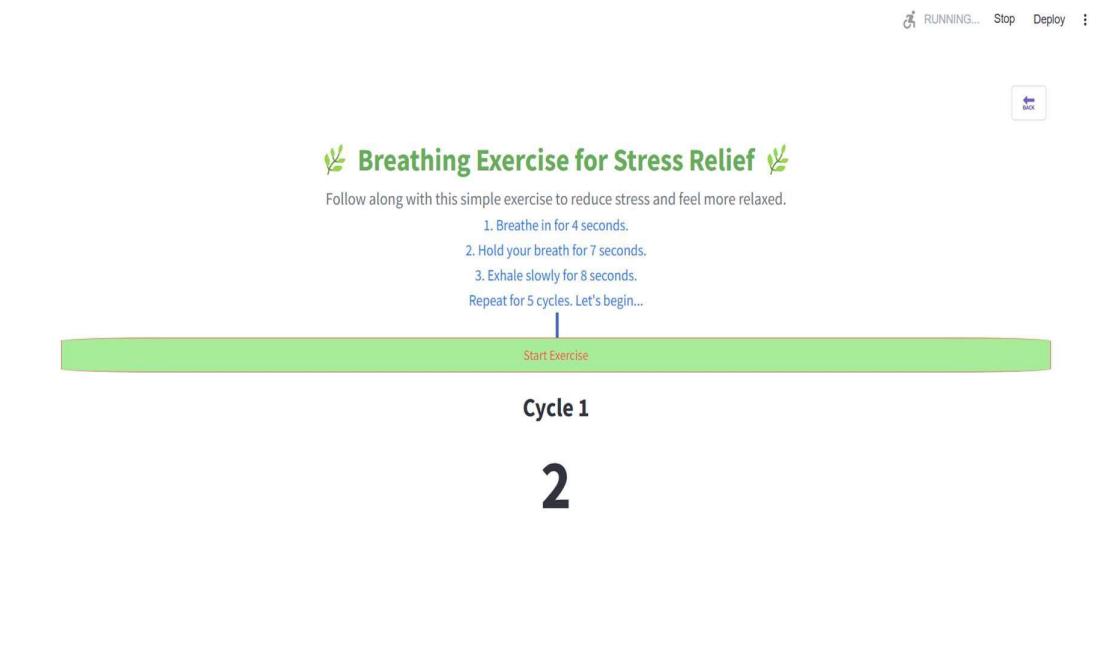


Fig. 7.1.10 Breathing Exercise for Stress Relief along with Instructions

The screenshot shows a web-based depression test interface. At the top right are 'Deploy' and three-dot buttons. Below them is a 'Run' button. The main title is 'Depression Test'. Underneath are three category buttons: 'Student', 'Working/Business', and 'Common_for_all'. The form fields include:

- Patient Name: A long input field.
- Mobile Number: An input field containing '10000000' with minus and plus buttons to its right.
- Landline Number: An empty input field.
- self_employed: A dropdown menu showing 'nan'.
- select your country: A dropdown menu showing 'United States'.

Fig. 7.1.11 Depression Test for Students, Working Professionals, Common People

The screenshot shows a report download page. At the top right are 'Deploy' and three-dot buttons. Below them is a 'Download Report' button. The main content area includes:

- 'Download Report' and 'Back to home' buttons.
- A 'Student' category button.
- A message: 'Report Generated and Downloaded Successfully.'

Fig. 7.1.12 Downloading the reports of Depression test

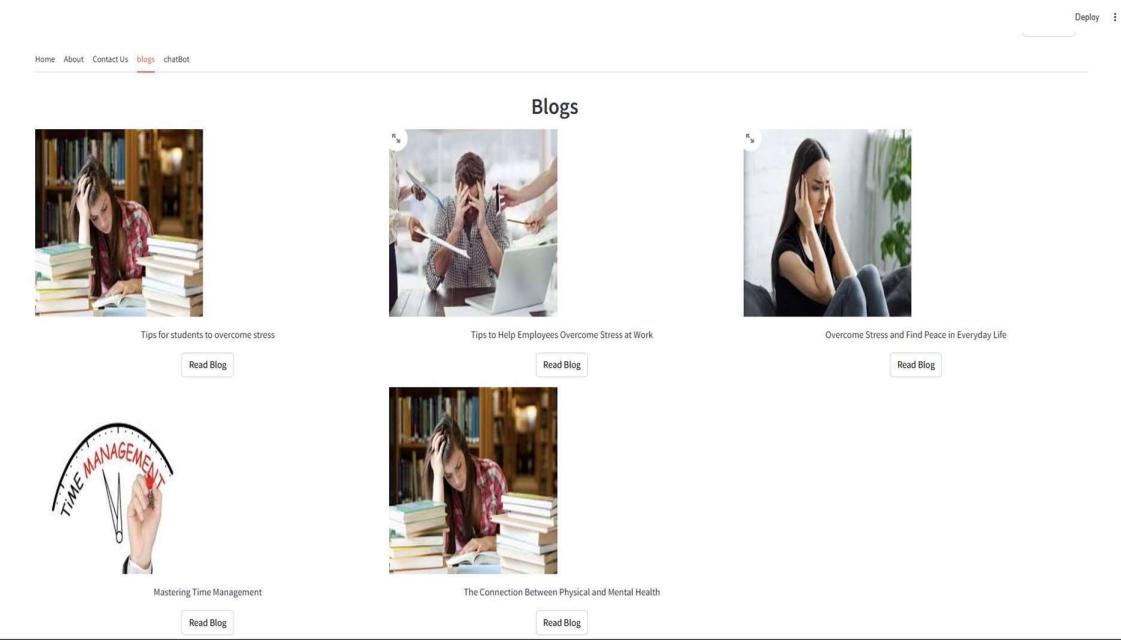


Fig. 7.1.13 Blogs to read regarding various concepts of Mental Health

The screenshot shows the CompanionMind AI-powered chatbot interface. At the top right are "Deploy" and three-dot menu icons. The main heading is "CompanionMind – Your AI Chat Companion" flanked by robot icons. Below the heading is a subtext: "Your personal AI-powered support for mindfulness, well-being, and guidance." A message input field contains the question "How can I manage Stress better?". A response card follows, starting with "Managing stress is essential for maintaining your overall well-being. Here are some effective ways to help you cope with stress:" followed by a numbered list of four tips. At the bottom is another message input field with the placeholder "How can I help you today?" and a send arrow icon.

Fig. 7.1.14 AI – Powered Chatbot for users

CHAPTER 8

CONCLUSION AND FUTUREWORK

8.1 CONCLUSION

In conclusion, this study introduces CompanionMind, an AI-driven approach designed to support mental well-being through personalized assessment and intervention. By integrating advanced Machine Learning models and Large Language Models, the system offers a comprehensive solution for understanding and addressing mental health challenges. Features such as personalized depression assessments, chatbot-based empathetic support, mindfulness meditation, and guided breathing exercises represent a significant advancement in mental health technology. The system's effectiveness is validated through extensive testing, ensuring accurate assessments and tailored recommendations, while its scalability and adaptability make it a valuable tool for individuals from diverse backgrounds. CompanionMind empowers users with self-care strategies, facilitates early identification of emotional distress, and fosters mental resilience, ultimately reducing barriers to mental health care. This innovative approach demonstrates the transformative potential of AI in mental health support, moving us closer to a future where mental health care is proactive, personalized, and inclusive, enhancing emotional well-being and quality of life worldwide.

8.2 CHALLENGES AND FUTURE WORK

Challenges:

Ensuring accurate assessments is challenging as mental health evaluations rely on user inputs, which may be influenced by emotions or reluctance to share personal details, requiring robust model interpretation. Personalization and adaptability are crucial since every individual's mental health journey is unique, demanding a chatbot and recommendation system that caters to diverse needs while maintaining sensitivity. Data privacy and security must be prioritized, implementing secure storage and encryption to protect sensitive user information. Additionally, bias and ethical concerns in AI need to be addressed, ensuring fairness, inclusivity, and ethical AI practices to prevent misinterpretations and incorrect recommendations.

Future Work:

The platform aims to integrate multiple mental health tests, covering stress, anxiety, and emotional well-being for a more comprehensive assessment system. Enhancements in AI-driven recommendations will focus on real-time user feedback and behavioral tracking for more personalized interventions. Gamification elements like streaks, achievements, and progress tracking will be introduced to keep users engaged and motivated in their self-care routines. Expanding CompanionMind into a cross-platform mobile application will improve accessibility and user experience. Additionally, multi-language support will be implemented to make mental health assistance available to a broader global audience.

CHAPTER - 9

REFRENCES

REFERENCES

- [1] Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation | npj Mental Health Research
- [2] The Opportunities and Risks of Large Language Models in Mental Health
- [3] Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data
- [4] Large Language Models for Mental Health Applications: Systematic Review
- [5] Large Language Models in Mental Health Care: a Scoping Review
- [6] Opportunities and risks of large language models in psychiatry
- [7] PsycoLLM: Enhancing LLM for Psychological Understanding and Evaluation
- [8] Ollama
- [9] Streamlit documentation
- [10] Llama 3.2 | Model Cards and Prompt formats
- [11] Running models with Ollama step-by-step | by Gabriel Rodewald | Medium
- [12] Saving a machine learning Model - GeeksforGeeks
- [13] SVC — scikit-learn 1.6.1 documentation
- [14] Introduction to the Gradient Boosting Algorithm | by Anjani Kumar | Analytics Vidhya | Medium

CHAPTER - 10

BASE PAPER

<https://doi.org/10.1038/s44184-024-00056-z>

Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation



Elizabeth C. Stade^{1,2,3} , Shannon Wiltsey Stirman^{1,2}, Lyle H. Ungar⁴, Cody L. Boland¹, H. Andrew Schwartz⁵, David B. Yaden⁶, João Sedoc⁷, Robert J. DeRubeis⁸, Robb Willer⁹ & Johannes C. Eichstaedt³

Large language models (LLMs) such as Open AI's GPT-4 (which power ChatGPT) and Google's Gemini, built on artificial intelligence, hold immense potential to support, augment, or even eventually automate psychotherapy. Enthusiasm about such applications is mounting in the field as well as industry. These developments promise to address insufficient mental healthcare system capacity and scale individual access to personalized treatments. However, clinical psychology is an uncommonly high stakes application domain for AI systems, as responsible and evidence-based therapy requires nuanced expertise. This paper provides a roadmap for the ambitious yet responsible application of clinical LLMs in psychotherapy. First, a technical overview of clinical LLMs is presented. Second, the stages of integration of LLMs into psychotherapy are discussed while highlighting parallels to the development of autonomous vehicle technology. Third, potential applications of LLMs in clinical care, training, and research are discussed, highlighting areas of risk given the complex nature of psychotherapy. Fourth, recommendations for the responsible development and evaluation of clinical LLMs are provided, which include centering clinical science, involving robust interdisciplinary collaboration, and attending to issues like assessment, risk detection, transparency, and bias. Lastly, a vision is outlined for how LLMs might enable a new generation of studies of evidence-based interventions at scale, and how these studies may challenge assumptions about psychotherapy.

Large language models (LLMs), built on artificial intelligence (AI) – such as Open AI's GPT-4 (which power ChatGPT) and Google's Gemini – are breakthrough technologies that can read, summarize, and generate text. LLMs have a wide range of abilities, including serving as conversational agents (chatbots), generating essays and stories, translating between languages, writing code, and diagnosing illness¹. With these capacities, LLMs are influencing many fields, including education, media, software

engineering, art, and medicine. They have started to be applied in the realm of behavioral healthcare, and consumers are already attempting to use LLMs for quasi-therapeutic purposes².

Applications incorporating older forms of AI, including natural language processing (NLP) technology, have existed for decades³. For example, machine learning and NLP have been used to detect suicide risk⁴, identify the assignment of homework in psychotherapy sessions⁵, and identify

¹Dissemination and Training Division, National Center for PTSD, VA Palo Alto Health Care System, Palo Alto, CA, USA. ²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. ³Institute for Human-Centered Artificial Intelligence & Department of Psychology, Stanford University, Stanford, CA, USA. ⁴Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA. ⁵Department of Computer Science, Stony Brook University, Stony Brook, NY, USA. ⁶Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

⁷Department of Technology, Operations, and Statistics, New York University, New York, NY, USA. ⁸Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA. ⁹Department of Sociology, Stanford University, Stanford, CA, USA. e-mail: betsystade@stanford.edu;johannes.stanford@gmail.com

patient emotions within psychotherapy⁶. Current applications of LLMs in the behavioral health field are far more nascent – they include tailoring an LLM to help peer counselors increase their expressions of empathy, which has been deployed with clients both in academic and commercial settings^{2,7}. As another example, LLM applications have been used to identify therapists' and clients' behaviors in a motivational interviewing framework^{8,9}.

Similarly, while algorithmic intelligence with NLP has been deployed in patient-facing behavioral health contexts, LLMs have not yet been heavily employed in these domains. For example, mental health chatbots Woebot and Tessa, which target depression and eating pathology respectively^{10,11}, are rule-based and do not use LLMs (i.e., the application's content is human-generated, and the chatbot's responds based on predefined rules or decision trees¹²). However, these and other existing chatbots frequently struggle to understand and respond to unanticipated user responses^{10,13}, which likely contributes to their low engagement and high dropout rates^{14,15}. LLMs may hold promise to fill some of these gaps, given their ability to flexibly generate human-like and context-dependent responses. A small number of patient-facing applications incorporating LLMs have been tested, including a research-based application to generate dialog for therapeutic counseling^{16,17}, and an industry-based mental-health chatbot, Youper, which uses a mix of rule-based and generative AI¹⁸.

These early applications demonstrate the potential of LLMs in psychotherapy – as their use becomes more widespread, they will change many aspects of psychotherapy care delivery. However, despite the promise they may hold for this purpose, caution is warranted given the complex nature of psychopathology and psychotherapy. Psychotherapy delivery is an unusually complex, high-stakes domain vis-à-vis other LLM use cases. For example, in the productivity realm, with a “LLM co-pilot” summarizing meeting notes, the stakes are failing to maximize efficiency or helpfulness; in behavioral healthcare, the stakes may include improperly handling the risk of suicide or homicide.

While there are other applications of artificial intelligence that may involve high-stakes or life-or-death decisions (e.g., self-driving cars), prediction and mitigation of risk in the case of psychotherapy is very nuanced, involving complex case conceptualization, the consideration of social and cultural contexts, and addressing unpredictable human behavior. Poor outcomes or ethical transgressions from clinical LLMs could run the risk of harming individuals, which may also be disproportionately publicized (as has occurred with other AI failures¹⁹), which may damage public trust in the field of behavioral healthcare.

Therefore, developers of clinical LLMs need to act with special caution to prevent such consequences. Developing responsible clinical LLMs will be a challenging coordination problem, primarily because the technological developers who are typically responsible for product design and development lack clinical sensitivity and experience. Thus, behavioral health experts will need to play a critical role in guiding development and speaking to the potential limitations, ethical considerations, and risks of these applications.

Presented below is a discussion on the future of LLMs in behavioral healthcare from the perspective of both behavioral health providers and technologists. A brief overview of the technology underlying clinical LLMs is provided for the purposes of both educating clinical providers and to set the stage for further discussion regarding recommendations for development. The discussion then outlines various applications of LLMs to psychotherapy and provides a proposal for the cautious, phased development and evaluation of LLM-based applications for psychotherapy.

Overview of clinical LLMs

Clinical LLMs could take a wide variety of forms, spanning everything from brief interventions or circumscribed tools to augment therapy, to chatbots designed to provide psychotherapy in an autonomous manner. These applications could be patient-facing (e.g., providing psychoeducation to the patient), therapist-facing (e.g., offering options for interventions from which the therapist could select), trainee-facing (e.g., offering feedback on qualities of the trainee's

performance), or supervisor/consultant facing (e.g., summarizing supervisees' therapy sessions in a high-level manner).

How language models work

Language models, or computational models of the probability of sequences of words, have existed for quite some time. The mathematical formulations date back to²⁰ and original use cases focused on compressing communication²¹ and speech recognition^{22–24}. Language modeling became a mainstay for choosing among candidate phrases in speech recognition and automatic translation systems but until recently, using such models for generating natural language found little success beyond abstract poetry²⁴.

Large language models

The advent of *large* language models, enabled by a combination of the deep learning technique transformers²⁵ and increases in computing power, has opened new possibilities²⁶. These models are first trained on massive amounts of data^{27,28} using “unsupervised” learning in which the model’s task is to predict a given word in a sequence of words. The models can then be tailored to a specific task using methods, including prompting with examples or fine-tuning, some of which use no or small amounts of task-specific data (see Fig. 1)^{28,29}. LLMs hold promise for clinical applications because they can parse human language and generate human-like responses, classify/score (i.e., annotate) text, and flexibly adopt conversational styles representative of different theoretical orientations.

LLMs and psychotherapy skills

For certain use cases, LLM show a promising ability to conduct tasks or skills needed for psychotherapy, such as conducting assessment, providing psychoeducation, or demonstrating interventions (see Fig. 2). Yet to date, clinical LLM products and prototypes have not demonstrated anywhere near the level of sophistication required to take the place of psychotherapy. For example, while an LLM can generate an alternative belief in the style of CBT, it remains to be seen whether it can engage in the type of turn-based, Socratic questioning that would be expected to produce cognitive change. This more generally highlights the gap that likely exists between simulating therapy skills and implementing them effectively to alleviate patient suffering. Given that psychotherapy transcripts are likely poorly represented in the training data for LLMs, and that privacy and ethical concerns make such representation challenging, prompt engineering may ultimately be the most appropriate fine-tuning approach for shaping LLM behavior in this manner.

Clinical LLMs: stages of integration

The integration of LLMs into psychotherapy could be articulated as occurring along a continuum of stages spanning from assistive AI to fully autonomous AI (see Fig. 3 and Table 1). This continuum can be illustrated by models of AI integration in other fields, such as those used in the autonomous vehicle industry. For example, at one end of this continuum is the assistive AI (“machine in the loop”) stage, wherein the vehicle system has no ability to complete the primary tasks – acceleration, braking, and steering – on its own, but provides momentary assistance (e.g., automatic emergency breaking, lane departure warning) to increase driving quality or decrease burden on the driver. In the collaborative AI (“human in the loop”) stage, the vehicle system aids in the primary tasks, but requires human oversight (e.g., adaptive cruise control, lane keeping assistance). Finally, in fully autonomous AI, vehicles are self-driving and do not require human oversight. The stages of LLM integration into psychotherapy and their related functionalities are described below.

Stage 1: assistive LLMs

At the first stage in LLM integration, AI will be used as a tool to assist clinical providers and researchers with tasks that can easily be “offloaded” to AI assistants (Table 1; first row). As this is a preliminary step in integration, relevant tasks will be low-level, concrete, and circumscribed, such that they present a low level of risk. Examples of tasks could include assisting with collecting information for patient intakes or assessment, providing basic

Fig. 1 | Methods for tailoring clinical large language models. Figure was designed using image components from Flaticon.com.

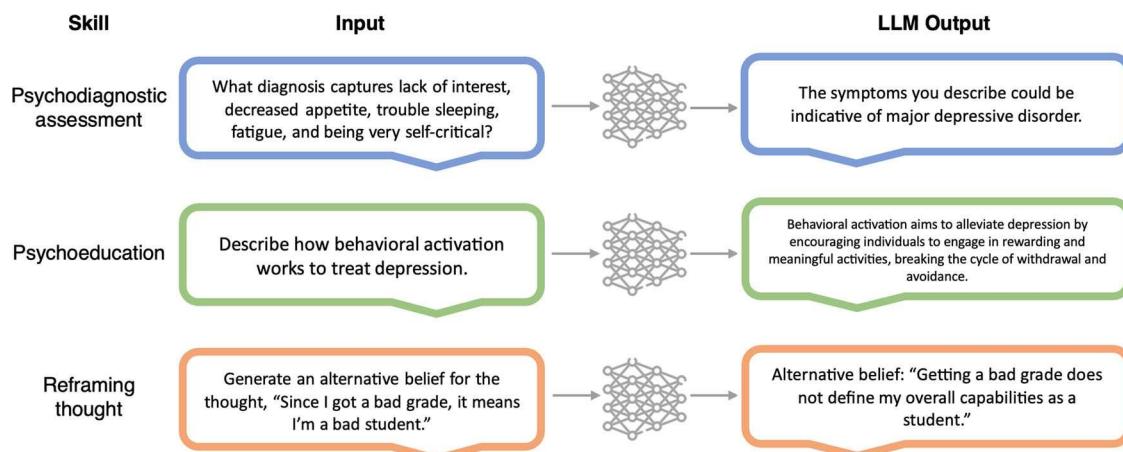
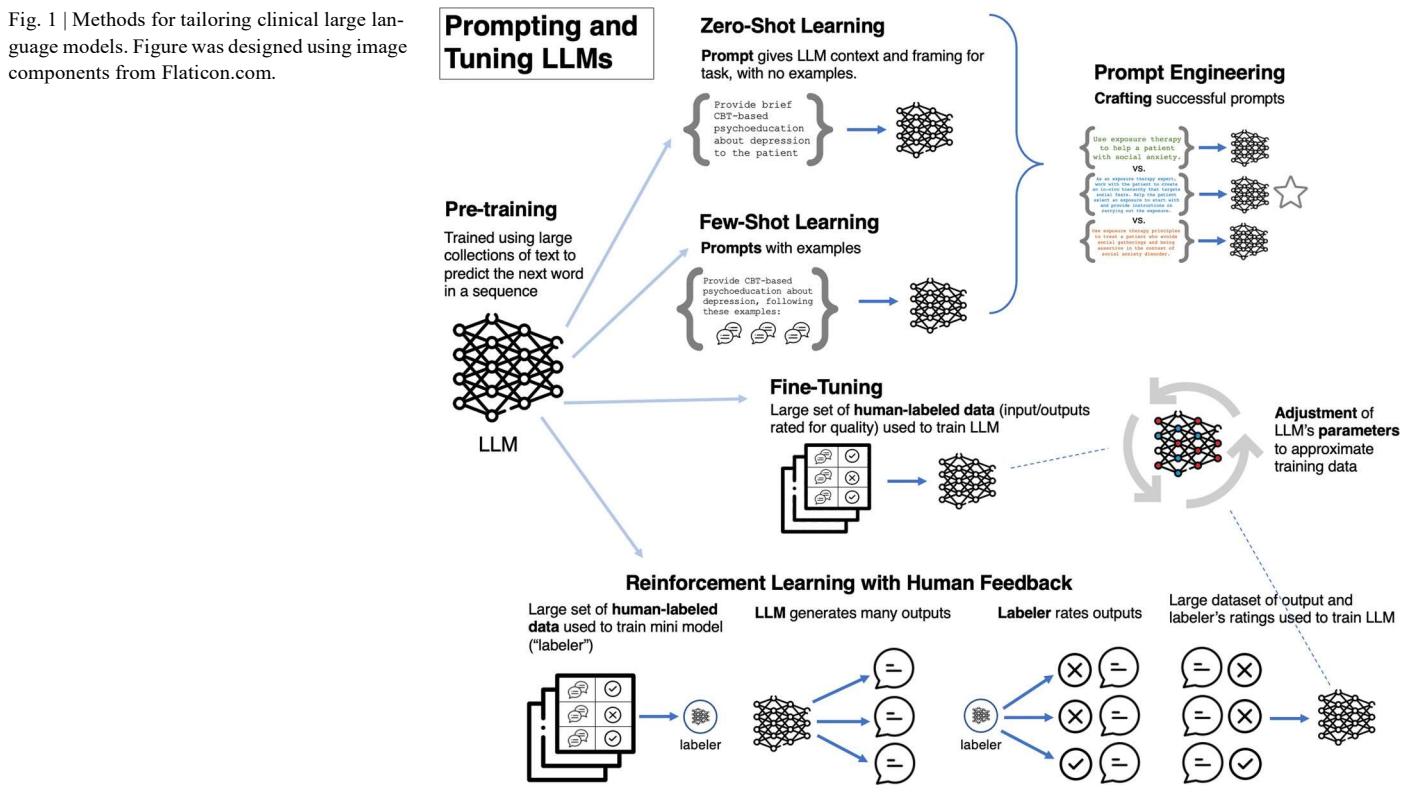


Fig. 2 | Example clinical skills of large language models. Note. Figure was designed using image component from Flaticon.com.

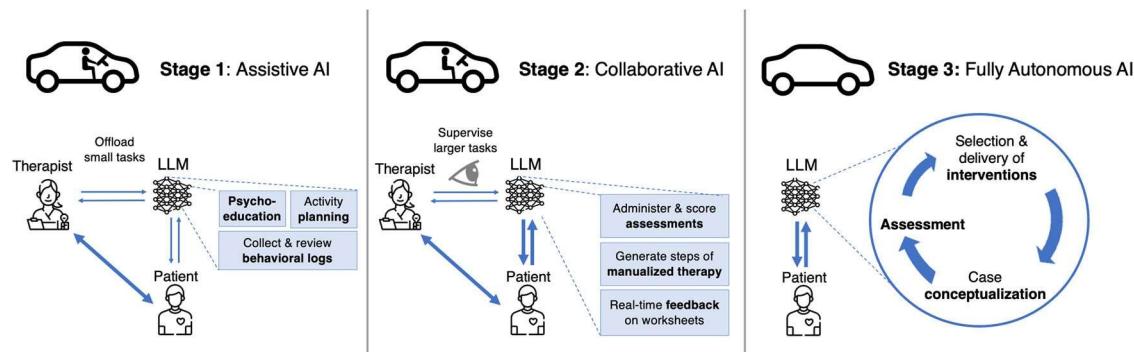


Fig. 3 | Stages of integrating large language models into psychotherapy. Figure was designed using image components from Flaticon.com.

Table 1 | Stages of Development of Clinical LLMs

Stage	Car Analogy	Characteristics of Assessment	Intervention Focus/Scope	Intervention Nature	Clinical Example	Potential Risks or Costs
Assistive AI ("machine in the loop")	AI-based features (e.g., automatic emergency breaking, lane departure warning) in the vehicle.	Standalone, modularized (e.g., assessments hand-picked by therapist and administered by survey).	Limited to concrete/ circumscribed (e.g., activity planning).	No full intervention packages; limited to components of interventions.	LLM trained to conduct skills from CBT-I might converse with the patient to collect their sleep diary data from the previous week to expedite a traditional therapy session.	Overhead and complexity for therapist for AI supervision.
Collaborative AI ("human in the loop")	Vehicle mostly completing the primary task; human in the driver seat actively monitors the vehicle's progress and overrides it as needed (e.g., adaptive cruise control, lane keeping assist).	Increasingly integrated (e.g., assessments recommended by LLM and summarized with context for therapist review).	Includes less concrete, more abstract interventions (e.g., planning and processing exposures).	Limited to structured/standar-dized (e.g., CBT for insomnia).	CBT-I LLM might generate a) an overview of the sleep diary data, b) a rationale for sleep restriction and stimulus control, and c) a sleep schedule prescription based on the diary data. This content would be reviewed and tailored by the psychotherapist before being discussed with the patient.	Drafts that require significant corrections may not save much time; Busy therapists may fail to check or tailor content, especially if given higher caseloads due to AI assistance.
Fully autonomous AI	Fully autonomous vehicles that operate without direct human oversight.	Fully integrated; informs intervention (e.g., unobtrusive, automated symptom assessment running in background).	Includes very abstract/ diffuse interventions (e.g., Socratic questioning).	Includes unstructured/unstan-dardized (e.g., acceptance and commitment therapy, idio-graphic or modular approaches).	LLM could implement a full course of CBT-I. The LLM would directly deliver multi-session therapy interventions and content to the patient, which would not be subject to tailoring or initial oversight by the psychotherapist.	Critical information could be mis-sed (e.g., suicide risk); Provision of inappropriate or harmful care.

AI: artificial intelligence, LLM: large language model, CBT-I: cognitive behavioral therapy for insomnia.

psychoeducation to patients, suggesting text edits for providers engaging in text-based care, and summarizing patient worksheets. Administratively, systems at this stage could also assist with clinical documentation by drafting session notes.

Stage 2: collaborative LLMs

Further along the continuum, AI systems will take the lead by providing or suggesting options for treatment planning and much of the therapy content, which humans will use their professional judgement to select from or tailor. For example, in the context of a text- or instant-message delivered structured psychotherapeutic intervention, the LLM might generate messages containing session content and assignments, which the therapist would review and adapt as needed before sending (Table 1; second row). A more advanced use of AI within the collaborative stage may entail a LLM providing a structured intervention in a semi-independent manner (e.g., as a chatbot), with a provider monitoring the discussion and stepping in to take control of the conversation as needed. The collaborative LLM stage has parallels to "guided self-help" approaches³⁰.

Stage 3: fully autonomous LLMs

In the fully autonomous stage, AIs will achieve the greatest degree of scope and autonomy wherein a clinical LLM would perform a full range of clinical skills and interventions in an integrated manner without direct provider oversight (Table 1; third row). For example, an application at this stage might theoretically conduct a comprehensive assessment, select an appropriate intervention, and deliver a full course of therapy with no human intervention. In addition to clinical content, applications in this stage could integrate with the electronic health record to complete clinical documentation and report writing, schedule appointments and process billing. Fully autonomous applications offer the most scalable treatment method³⁰.

Progression across the stages

Progression across the stages may not be linear; human oversight will be required to ensure that applications at greater stages of integration are safe for real world deployment. As different forms of psychopathology and their accompanying interventions vary in complexity, certain types of interventions will be simpler than others to develop as LLM applications. Interventions that are more concrete and standardized may be easier for models to deliver (and may be available sooner), such as circumscribed behavior change interventions (e.g., activity scheduling), as opposed to applications which include skills that are abstract in nature or emphasize cognitive change (e.g., Socratic questioning). Similarly, when it comes to full therapy protocols, LLM applications for interventions that are highly structured, behavioral, and protocolized (e.g., CBT for insomnia [CBT-I] or exposure therapy for specific phobia) may be available sooner than applications delivering highly flexible or personalized interventions (for example³¹).

In theory, the final stage in the integration of LLMs into psychotherapy is fully autonomous delivery of psychotherapy which does not require human intervention or monitoring. However, it remains to be seen whether fully autonomous AI systems will reach a point at which they have been evaluated to be safe for deployment by the behavioral health community. Specific concerns include how well these systems are able to carry out case conceptualization on individuals with complex, highly comorbid symptom presentations, including accounting for current and past suicidality, substance use, safety concerns, medical comorbidities, and life circumstances and events (such as court dates and upcoming medical procedures). Similarly, it is unclear whether these systems will prove sufficiently adept at engaging patients over time³² or accounting for and addressing contextual nuances in treatment (e.g., using exposure to treat a patient experiencing PTSD-related fear of leaving the house, who also lives in a neighborhood with high rates of crime). Furthermore, several skills which may be viewed as central to clinical work currently fall outside the purview of LLM systems, such as interpreting nonverbal behavior (e.g., fidgeting, eye-rolling), appropriately challenging a patient, addressing alliance ruptures, and making decisions about termination. Technological advances, including the

Table 2 | Imminent possibilities for clinical LLMs

Task	Target Audience	Example Input to LLM	Example LLM Output
Aid in administrative tasks	Clinician	Psychotherapy session recording	"... Met with patient for cognitive behavioral therapy for depression. Reviewed homework; patient completed three thought records over the past week. Introduced 'thinking biases' worksheet; assisted patient in identifying patterns of problematic thinking applicable to her automatic negative thoughts from the past week..."
Offer feedback on therapy homework worksheets	Patient	Digital CBT worksheet; Patient writes, "I've always felt this way," as evidence in support of the negative automatic thought: "I'm unlovable" on the worksheet	"Remember, 'evidence' means facts that support the belief. Sometimes it's helpful to think about facts so strong they would stand up in a court of law. What is the evidence that you are unlovable?"
Produce adherence and competence ratings for elements of therapy	Researcher	Psychotherapy session recording	"...Therapist helped patient identify negative automatic thoughts Adherence rating (0-1): 1 Competence rating (0-6): 5..."
Identify trainee psychotherapist's areas of success and areas for improvement	Psychotherapy trainee	Psychotherapy session recording	"...In the following exchange, the therapist successfully used Socratic questioning to ask open-ended, non-leading questions: [Patient: I should have known that it wasn't safe to get in that car. Therapist: Hm, help me understand... how could you have known that it wasn't safe?...]"
Suggest an improved therapeutic response, offer education about therapeutic exchanges	Peer counselor or lay mental health worker	Message-based exchange between patient and peer counselor; peer counselor has drafted a response: "You'll be fine"	"This could be improved by offering validation of the client's feelings. For instance, you might say, 'it sounds like you're going through a difficult time, and it's understandable to feel overwhelmed.' Would you like to rewrite before sending?"

approaching advent of multimodal language models that integrate text, images, video, and audio, may eventually begin to fill these gaps.

Beyond technical limitations, it remains to be decided whether complete automation is an appropriate end goal for behavioral healthcare, due to safety, legal, philosophical, and ethical concerns³³. While some evidence indicates that humans can develop a therapeutic alliance with chatbots³⁴, the long-term viability of such alliance building, and whether or not it produces undesirable downstream effects (e.g., altering an individual's existing relationships or social skills) remains to be seen. Others have documented potentially harmful behavior of LLM chatbots, such as narcissistic tendencies³⁵ and expressed concerns about the potential for their undue influence on humans in addition to articulating societal risks associated with LLMs more generally^{36,37}. The field will also need to grapple with questions of accountability and liability in the case of a fully autonomous clinical LLM application causing damage (e.g., identifying the responsible party in an incident of malpractice³⁸). For these and other reasons, some have argued against the implementation of fully autonomous systems in behavioral healthcare and healthcare more broadly^{39,40}. Taken together, these issues and concerns may suggest that in the short and medium term, assistive or collaborative AI applications will be more appropriate for the provision of behavioral healthcare.

Applications of clinical LLMs

Given the vast nature of behavioral healthcare, there are seemingly endless applications of LLMs. Outlined below are some of the currently existing, imminently feasible, and potential long-term applications of clinical LLMs. Here we focus our discussion on applications directly related to the provision of, training in, and research on psychotherapy. As such, several important aspects of behavioral healthcare, such as initial symptom detection, psychological assessment and brief interventions (e.g., crisis counseling) are not explicitly discussed herein.

Imminent applications

Automating clinical administration tasks. At the most basic level, LLMs have the potential to automate several time-consuming tasks associated with providing psychotherapy (Table 2, first row). In addition to using session transcripts to summarize the session for the provider, there is potential for such models to integrate within electronic health

records to aid with clinical documentation and conducting chart reviews. Clinical LLMs could also produce a handout for the patient that provides a personalized overview of the session, skills learned and assigned homework or between-session material.

Measuring treatment fidelity. A clinical LLM application could automate measurement of therapist fidelity to evidence-based practices (EBPs; Table 2, second row), which can include measuring *adherence* to the treatment as designed, *competence* in delivering a specific therapy skill, *treatment differentiation* (whether multiple treatments being compared actually differ from one another), and *treatment receipt* (patient comprehension of, engagement with, and adherence to the therapy content)^{41,42}. Measuring fidelity is crucial to the development, testing, dissemination, and implementation of EBPs, yet can be resource intensive and difficult to do reliably. In the future, clinical LLMs could computationally derive adherence and competence ratings, aiding research efforts and reducing therapist drift⁴³. Traditional machine-learning models are already being used to assess fidelity to specific modalities⁴⁴ and other important constructs like counseling skills⁴⁵ and alliance⁴⁶. Given their improved ability to consider context, LLMs will likely increase the accuracy with which these constructs are assessed.

Offering feedback on therapy worksheets and homework. LLM applications could also be developed to deliver real-time feedback and support on patients' between-session homework assignments (Table 2, third row). For example, an LLM tailored to assist a patient to complete a CBT worksheet might provide clarification or aid in problem solving if the patient experiences difficulty (e.g., the patient was completing a thought log and having trouble differentiating between the thought and the emotion). This could help to "bridge the gap" between sessions and expedite patient skill development. Early evidence outside the AI realm⁴⁷ points to increasing worksheet competence as a fruitful clinical target.

Automating aspects of supervision and training. LLMs could be used to provide feedback on psychotherapy or peer support sessions, especially for clinicians with less training and experience (i.e., peer counselors, lay health workers, psychotherapy trainees). For example, an LLM might be used to offer corrections and suggestions to the dialog of peer counselors (Table 2,

fourth row). This application has parallels to “task sharing,” a method used in the global mental health field by which nonprofessionals provide mental health care with the oversight by specialist workers to expand access to mental health services⁴⁸. Some of this work is already underway, for example, as described above, using LLMs to support peer counselors⁷.

LLMs could also support supervision for psychotherapists learning new treatments (Table 2, fifth row). Gold-standard methods of reviewing trainees’ work, like live observation or review of recorded sessions⁴⁹, are time-consuming. LLMs could analyze entire therapy sessions and identify areas of improvement, offering a scalable approach for supervisors or consultants to review.

Potential long-term applications

It is important to note that many of the potential applications listed below are theoretical and have yet to be developed, let alone thoroughly evaluated. Furthermore, we use the term “clinical LLM” in recognition of the fact that when and under what circumstances the work of an LLM could be called psychotherapy is evolving and depends on how psychotherapy is defined.

Fully autonomous clinical care. As previously described, the final stage of clinical LLM development could involve an LLM that can independently conduct comprehensive behavioral healthcare. This could involve all aspects related to traditional care including conducting assessment, presenting feedback, selecting an appropriate intervention and delivering a course of therapy to the patient. This course of treatment could be delivered in ways consistent with current models of psychotherapy wherein a patient engages with a “chatbot” weekly for a prescribed amount of time, or in more flexible or alternative formats. LLMs used in this manner would ideally be trained using standardized assessment approaches and manualized therapy protocols that have large bodies of evidence.

Decision aid for existing evidence-based practices. Even without full automation, clinical LLMs could be used as a tool to guide a provider on the best course of treatment for a given patient by optimizing the delivery of existing EBPs and therapeutic techniques. In practice, this may look like a LLM that can analyze transcripts from therapy sessions and offer a provider guidance on therapeutic skills, approaches or language, either in real time, or at the end of the therapy session. Furthermore, the LLM could integrate current evidence on the tailoring of specific EBPs to the condition being treated, and to demographic or cultural factors and comorbid conditions. Developing tailored clinical LLM “advisors” based on EBPs could both enhance fidelity to treatment and maximize the possibility of patients achieving clinical improvement in light of updated clinical evidence.

Development of new therapeutic techniques and EBPs. To this point, we have discussed how LLMs could be applied to current approaches to psychotherapy using extant evidence. However, LLMs and other computational methods could greatly enhance the detection and development of new therapeutic skills and EBPs. Historically, EBPs have traditionally been developed using human-derived insights and then evaluated through years of clinical trial research. While EBPs are effective, effect sizes for psychotherapy are typically small^{50,51} and significant proportions of patients do not respond⁵². There is a great need for more effective treatments, particularly for individuals with complex presentations or comorbid conditions. However, the traditional approach to developing and testing therapeutic interventions is slow, contributing to significant time lags in translational research⁵³, and fails to deliver insights at the level of the individual.

Data-driven approaches hold the promise of revealing patterns that are not yet realized by clinicians, thus generating new approaches to psychotherapy; machine learning is already being used, for example, to predict behavioral health treatment outcomes⁵⁴. With their ability to parse and summarize natural language, LLMs could add to existing data-driven approaches. For example, an LLM could be provided with a large historical dataset containing psychotherapy transcripts of different therapeutic

orientations, outcome measures and sociodemographic information, and tasked with detecting therapeutic behaviors and techniques associated with objective outcomes (e.g., reduction in depressive symptoms). Using such a process might make it possible for an LLM to yield fine-grained insights about what makes existing therapeutic techniques work best (e.g., Which components of existing EBPs are the most potent? Are there therapist or patient characteristics that moderate the efficacy of intervention X? How does the ordering of interventions effect outcomes?) or even to isolate previously unidentified therapeutic techniques associated with improved clinical outcomes. By identifying what happens in therapy in such a fine-grained manner, LLMs could also play a role in revealing mechanisms of change, which is important for improving existing treatments and facilitating real-world implementation⁵⁵.

However, to realize this possibility, and make sure that LLM-based advances can be integrated and vetted by the clinical community, it is necessary to steer away from the development of “black box,” LLM-identified interventions with low explainability (e.g., interpretability⁵⁶). To guard against interventions with low interpretability, work to finetune LLMs to improve patient outcomes could include inspectable representations of the techniques employed by the LLM. Clinicians could examine these representations and situate them in the broader psychotherapy literature, which would involve comparing them to existing psychotherapy techniques and theories. Such an approach could speed up the identification of novel mechanisms while guarding against the identification of “novel” interventions which overlap with existing techniques or constructs (thus avoiding the jangle fallacy, the erroneous assumption that two constructs with different names are necessarily distinct⁵⁷).

In the long run, by combining this information, it might even be possible for an LLM to “reverse-engineer” a new EBP, freed from the constraints of traditional therapeutic protocols and instead maximizing on the delivery of the constituent components shown to produce patient change (in a manner akin to modular approaches, wherein an individualized treatment plan is crafted for each patient by curating and sequencing treatment modules from an extensive menu of all available options based on the unique patient’s presentation³¹). Eventually, a self-learning clinical LLM might deliver a broad range of psychotherapeutic interventions while measuring patient outcomes and adapting its approach on the fly in response to changes in the patient (or lack thereof).

Toward a precision medicine approach to psychotherapy

Current approaches to psychotherapy often are unable to provide guidance on the best approach to treatment when an individual has a complex presentation, which is often the rule rather than being the exception. For example, providers are likely to have greatly differing treatment plans for a patient with concurrent PTSD, substance use, chronic pain, and significant interpersonal difficulties. Models that use a data-driven approach (rather than a provider’s educated guess) to address an individual’s presenting concern alongside their comorbidities, sociodemographic factors, history, and responses to the current treatment, may ultimately offer the best chance at maximizing patient benefit. While there have been some advances in precision medicine approaches in behavioral healthcare^{54,58}, these efforts are in their infancy and limited by sample sizes⁵⁹.

The potential applications of clinical LLMs we have outlined above may come together to facilitate a personalized approach to behavioral healthcare, analogous to that of precision medicine. Through optimizing existing EBPs, identifying new therapeutic approaches, and better understanding mechanisms of change, LLMs (and their future descendants) may provide behavioral healthcare with an enhanced ability to identify what works best for whom and under what circumstances.

Recommendations for responsible development and evaluation of clinical LLMs

Focus first on evidence-based practices

In the immediate future, clinical LLM applications will have the greatest chance of creating meaningful clinical impact if developed based on EBPs or

a “common elements” approach (i.e., evidence-based procedures shared across treatments)⁶⁰. Evidence-based treatments and techniques have been identified for specific psychopathologies (e.g., major depressive disorder, posttraumatic stress disorder), stressors (e.g., bereavement, job loss, divorce), and populations (e.g., LGBTQ individuals, older adults)^{55,61,62}. Without an initial focus on EBPs, clinical LLM applications may fail to reflect current knowledge and may even produce harm⁶³. Only once LLMs have been fully trained on EBPs can the field start to consider using LLMs in a data-driven manner, such as those outlined in the previous section on potential long-term applications.

Focus next on improvement (engagement is not enough)

Others have highlighted the importance of promoting engagement with digital mental health applications¹⁵, which is important for achieving an adequate “dose” of the therapeutic intervention. LLM applications hold the promise of improving engagement and retention through their ability to respond to free text, extract key concepts, and address patients’ unique context and concerns during interventions in a timely manner. However, engagement alone is not an appropriate outcome on which to train an LLM, because engagement is not expected to be sufficient for producing change. A focus on such metrics for clinical LLMs will risk losing sight of the primary goals, clinical improvement (e.g., reductions in symptoms or impairment, increases in well-being and functioning) and prevention of risks and adverse events. It will behoove the field to be wary of attempts to optimize clinical LLMs on outcomes that have an explicit relationship with a company’s profit (e.g., length of time using the application). An LLM that optimizes only for engagement (akin to YouTube recommendations) could have high rates of user retention without employing meaningful clinical interventions to reduce suffering and improve quality of life. Previous research has suggested that this may be happening with non-LLM digital mental health interventions. For instance, exposure is a technique with strong support for treating anxiety, yet it is rarely included in popular smartphone applications for anxiety⁶⁴, perhaps because developers fear that the technique will not appeal to users, or have concerns about how exposures going poorly or increasing anxiety in the short term, which may prompt concerns about legal exposure.

Commit to rigorous yet commonsense evaluation

An evaluation approach for clinical LLMs that hierarchically prioritizes risk and safety, followed by feasibility, acceptability, and effectiveness, would be in line with existing recommendations for the evaluation of digital mental health smartphone apps⁶⁵. The first level of evaluation could involve a demonstration that a clinical LLM produces no harm or very minimal harm that is outweighed by its benefits, similar to FDA phase I drug tests. Key risk and safety related constructs include measures of suicidality, non-suicidal self harm, and risk of harm to others.

Next, rigorous examinations of clinical LLM applications will be needed to provide empirical evidence of their utility, using head-to-head comparisons with standard treatments. Key constructs to be assessed in these empirical tests are feasibility and acceptability to the patient and the therapist as well as treatment outcomes (e.g., symptoms, impairment, clinical status, rates of relapse). Other relevant considerations include patients’ user experience with the application, measures of therapist efficiency and burnout, and cost.

Lastly, we note that given that possible benefits of clinical LLMs (including expanding access to care), it will be important for the field to adopt a commonsense approach to evaluation. While rigorous evaluation is important, the comparison conditions on which these evaluations are based should reflect real-world risk and efficacy rates, and perhaps employ a graded hierarchy with which to classify risk and error (i.e., missing a mention of suicidality is unacceptable, but getting a patient’s partner’s name wrong is nonideal but tolerable), rather than holding clinical LLM applications to a standard of perfection which humans do not achieve. Furthermore, developers will need to strike the appropriate balance of prioritizing constructs in a manner expected to be most clinically beneficial,

for example, if exposure therapy is indicated for the patient, but the patient does not find this approach acceptable, the clinical LLM could recommend the intervention prioritizing effectiveness before offering second-line interventions which may be more acceptable.

Involve interdisciplinary collaboration

Interdisciplinary collaboration between clinical scientists, engineers, and technologists will be crucial in the development of clinical LLMs. While it is plausible that engineers and technologists could use available therapeutic manuals to develop clinical LLMs without the expertise of a behavioral health expert, this is ill-advised. Manuals are only a first step towards learning a specific intervention, as they do not provide guidance on how the intervention can be applied to specific individuals or presentations, or how to handle specific issues or concerns that may arise through the course of treatment.

Clinicians and clinician-scientists have expertise that bears on these issues, as well as many other aspects of the clinical LLM development process. Their involvement could include a) testing new applications to identify limitations and risks and optimize their integration into clinical practice, b) improving the ability of applications to adequately address the complexity of psychological phenomena, c) ensuring that applications are developed and implemented in an ethical manner, and d) testing and ensuring that applications don’t have iatrogenic effects, such as reinforcing behaviors that perpetuate psychopathology or distress.

Behavioral health experts could also provide guidance on how best to finetune or tailor models, including addressing the question of whether and how real patient data should be used for these purposes. For example, most proximately, behavioral health experts might assist in *prompt engineering*, or the designing and testing of a series of prompts which provide the LLM framing and context for delivering a specific type of treatment or clinical skill (e.g., “Use cognitive restructuring to help the patient evaluate and reappraise negative thoughts in depression”), or a desired clinical task, such as evaluating therapy sessions for fidelity (e.g., “Analyze this psychotherapy transcript and select sections in which the therapist demonstrated the particularly skillful use of CBT skills, and sections in which the therapist’s delivery of CBT skills could be improved”). Similarly, in *few-shot learning*, behavioral health experts could be involved in crafting example exchanges which are added to prompts. For example, treatment modality experts might generate examples of clinical skills (e.g., high-quality examples of using cognitive restructuring to address depression) or of a clinical task (e.g., examples of both high- and low-quality delivery of CBT skills). For *finetuning*, in which a large, labeled dataset is used to train the LLM, and *reinforcement learning from human feedback* (RLHF), in which a human-labeled dataset is used to train a smaller model which is then used for LLM “self-training,” behavioral health experts could build and curate (and ensure informed patient consent for use of) appropriate datasets (e.g., a dataset containing psychotherapy transcripts rated for fidelity to an evidence-based psychotherapy). The expertise that behavioral health experts could draw on to generate instructive examples and curate high-quality datasets holds particular value in light of recent evidence that *quality* of data trumps *quantity* of data for training well-performing models⁶⁶.

In the service of facilitating interdisciplinary collaboration, it would benefit clinical scientists to seek out a working knowledge about LLMs, while it would benefit technologists to develop a working knowledge of therapy in general and EBPs in particular. Dedicated venues that bring together behavioral health experts and clinical psychologists for interdisciplinary collaboration and communication will aid in these efforts. Historically, venues of this type have included psychology-focused workshops at NLP conferences (e.g., the Workshop on Computational Linguistics and Clinical Psychology [CLPsych], held at the Annual Conference of the North American Chapter of the Association for Computational Linguistics [NAACL]) and technology-focused conferences or workgroups hosted by psychological organizations (e.g., APA’s Technology, Mind & Society conference; Association for Behavioral and Cognitive Therapies’ [ABCT] Technology and Behavior Change special interest group). This work has

also been done at nonprofits centered on technological tools for mental health (e.g., the Society for Digital Mental Health). Beyond these venues, it may be fruitful to develop a gathering that brings together technologists, clinical scientists, and industry partners with a dedicated focus on AI/LLMs, which could routinely publish on its efforts, akin to the efforts of the World Health Organization's Infodemic Management Conference, which has employed this approach to address misinformation⁶⁷. Finally, given the numerous applications of AI to behavioral health, it is conceivable that a new "computational behavioral health" subfield could emerge, offering specialized training that would bridge the gap between these two domains.

Focus on trust and usability for clinicians and patients
It is important to engage therapists, policymakers, end-users, and experts in human-computer interactions to understand and improve levels of trust that will be necessary for successful and effective implementation. With respect to applications of AI to augment supervision and support for psychotherapy, therapists have expressed concern about privacy, the ability to detect subtle non-verbal cues and cultural responsiveness, and the impact on therapist confidence, but they also see benefits for training and professional growth⁶⁸. Other research suggests that while therapists believe AI can increase access to care, allow individuals to disclose embarrassing information more comfortably, continuously refine therapeutic techniques⁶⁹, they have concerns about privacy and the formation of a strong therapeutic bond with machine-based therapeutic interventions⁷⁰. Involvement of individuals who will be referring their patients and using LLMs in their own practice will be essential to developing solutions they can trust and implement, and to make sure these solutions have the features that support trust and usability (simple interfaces, accurate summaries of AI-patient interactions, etc.).

Regarding how much patients will trust the AI systems, following the stages we outlined in Fig. 3, initial AI-patient interactions will continue to be supervised by clinicians, and the therapeutic bond between the clinician and the patient will continue to be the primary relationship. During this stage, it is important that clinicians talk to the patients about their experience with the LLMs, and that the field as a whole begins to accumulate an understanding and data on how acceptable interfacing with LLMs is for what kind of patient for what kind of clinical use case, in how clinicians can scaffold the patient-LLM relationship. This data will be critical for developing collaborative LLM applications that have more autonomy, and for ensuring that the transition from assistive to collaborative stage applications is not associated with large unforeseen risk. For example, in the case of CBT for insomnia, once an assistive AI system has been iterated on to reliably collect information about patients' sleep patterns, it is more conceivable that it could be evolved into a collaborative AI system that does a comprehensive insomnia assessment (i.e., it also collects and interprets data on patients' clinically significant distress, impairment of functioning, and ruling out of sleep-wake disorders, like narcolepsy)⁷¹.

Design criteria for effective clinical LLMs

Below, we propose an initial set of desirable design qualities for clinical LLMs.

Detect risk of harm.

- Accurate risk detection and mandated reporting are crucial aspects that clinical LLMs must prioritize, particularly in the identification of suicidal/homicidal ideation, child/elder abuse, and intimate partner violence. Algorithms for detecting risks are under development⁴. One threat to risk detection is that current LLMs have limited context windows, meaning they only "remember" a limited amount of user input. Functionally, this means a clinical LLM application could "forget" crucial details about a patient, which could impact safety (e.g., an application "forgetting" that the patient owns firearms would threaten its ability to properly assess and intervene around suicide risk). However, context windows have been rapidly expanding with each subsequent model release, so this issue may not be a problem for long.

In addition, it is already possible to augment the memory of LLMs with "vector databases," which would have the added benefit of retaining inspectable learnings and summaries across clinical encounters⁷².

In the future, and especially given much larger context windows, clinical LLMs could prompt clinicians with ethical guidelines, legal requirements (e.g., the Tarasoff rule, which requires clinicians to warn intended victims when a patient presents a serious threat of violence), or evidence-based methods for decreasing risk (e.g., safety planning⁷³), or even provide interventions targeting risk directly to patients. This type of risk monitoring and intervention could be particularly useful in supplementing existing healthcare systems during gaps in clinician coverage like nights and weekends⁴.

b) Be "Healthy." There is growing concern that AI chat systems can demonstrate undesirable behaviors, including expressions akin to depression or narcissism^{35,74}. Such poorly understood, undesirable behaviors risk harming already vulnerable patients or interfering with their ability to benefit from treatment. Clinical LLM applications will need training, monitoring, auditing, and guardrails to prevent the expression of undesirable behaviors and maintain healthy interactions with users. These efforts will need to be continually evaluated and updated to prevent or address the emergence of new undesirable or clinically contraindicated behavior.

Aid in psychodiagnostic assessment. Clinical LLMs ought to integrate psychodiagnostic assessment and diagnosis, facilitating intervention selection and outcome monitoring⁷⁵. Recent developments show promise for LLMs in the assessment realm⁷⁶. Down the line, LLMs could be used for diagnostic interviewing (e.g., Structured Clinical Interview for the *DSM-5*⁷⁷) using chatbots or voice interfaces. Prioritizing assessment enhances diagnostic accuracy and ensures appropriate intervention, reducing the risk of harmful interventions⁶³.

Be responsive and flexible. Given the frequency with which ambivalence and poor patient engagement arise in clinical encounters, clinical LLMs which use evidence-based and patient-centered methods for handling these issues (e.g., motivational enhancement techniques, shared decision making), and have options for second-line interventions for patients not interested in gold-standard treatments, will have the best chance of success.

Stop when not helping or confident. Psychologists are ethically obligated to cease treatment and offer appropriate referrals to the patient if the current course of treatment has not helped or likely will not help. Clinical LLMs can abide by this ethical standard by drawing on integrated assessment (discussed above) to assess the appropriateness of the given intervention and detect cases that need more specialized or intensive intervention.

Be fair, inclusive, and free from bias. As has been written about extensively, LLMs may perpetuate bias, including racism, sexism, and homophobia, given that they are trained on existing text³⁶. These biases can contribute to both error disparities – where models are less accurate for particular groups – or outcome disparities – where models tend to over-capture demographic information⁷⁸ – which would in turn contribute to the disparities in mental health status and care already experienced by minoritized groups⁷⁹. The integration of bias countermeasures into clinical LLM applications could serve to prevent this^{78,80}.

Be empathetic—to an extent. Clinical LLMs will likely need to demonstrate empathy and build the therapeutic alliance in order to engage patients. Other skills used by therapists include humor, irreverence, and gentle methods of challenging the patient. Incorporating these into clinical LLMs might be beneficial, as appropriate human likeness may facilitate engagement and interaction with AI⁸¹. However, this needs to be balanced against associated risks, mentioned above, of

incorporating human likeness in systems³⁶. Whether and how much human likeness is necessary for a psychological intervention remains a question for future empirical work.

Be transparent about being AIs. Mental illness and mental health care is already stigmatized, and the application of LLMs without transparent consent can erode patient/consumer trust, which reduces trust in the behavioral health profession more generally. Some mental health startups have already faced criticism for employing generative AI in applications without disclosing this information to the end user². As laid out in the White House Blueprint for an AI Bill of Rights, AI applications should be explicitly (and perhaps repeatedly/consistently) labeled as such to allow patients and consumers to “know that an automated system is being used and understand how and why it contributes to outcomes that impact them”⁸².

Discussion

Unintended consequences may change the clinical profession

The development of clinical LLM applications could lead to unintended consequences, such as changes to the structure of and compensation for mental health services. AI may permit increased staffing by non-professionals or paraprofessionals, causing professional clinicians to supervise large numbers of non-professionals or even semi-autonomous LLM systems. This could reduce clinicians' direct patient contact and perhaps increase their exposure to challenging or complicated cases not suitable for the LLM, which may lead to burnout and make clinical jobs less attractive. To address this, research could determine the appropriate number of cases for a clinician to oversee safely and guidelines could be published to disseminate these findings. The 24-hour availability of LLM-based intervention may also change consumer expectations of psychotherapy in a way that is at odds with many of the norms of psychotherapy practice (e.g., waiting for a session to discuss stressors, limited or emergency-only contact between sessions).

LLMs could pave the way for a next generation of clinical science
Beyond the imminent applications described in this paper, it is worth considering how the long-term applications of clinical LLMs might also facilitate significant advances in clinical care and clinical science.

Clinical practice. In terms of their effects on therapeutic interventions themselves, clinical LLMs might promote advances in the field by allowing for the pooling of data on what works with the most difficult cases, perhaps through the use of practice research networks⁸³. At the level of health systems, they could expedite the implementation and translation of research findings into clinical practice by suggesting therapeutic strategies to psychotherapists, for instance, promoting strategies that enhance inhibitory learning during exposure therapy⁸⁴. Lastly, clinical LLMs could increase access to care if LLM-based psychotherapy chatbots are offered as low intensity, low-cost options in stepped-care models, similar to the existing provision of computerized CBT and guided self-help⁸⁵.

As the utilization of clinical LLMs expands, there may be a shift towards psychologists and other behavioral health experts operating at the top of their degree. Presently, a significant amount of clinician time is consumed by administrative tasks, chart review, and documentation. The shifting of responsibilities afforded by the automation of certain aspects of psychotherapy by clinical LLMs could allow clinicians to pursue leadership roles, contribute to the development, evaluation, and implementation of LLM-based care, or lead policy efforts, or simply to devote more time to direct patient care.

Clinical science. By facilitating supervision, consultation, and fidelity measurement, LLMs could expedite psychotherapist training and increase the capacity of study supervisors, thus making psychotherapy research less expensive and more efficient.

In a world in which fully autonomous LLM applications screen and assess patients, deliver high-fidelity, protocolized psychotherapy, and collect

outcome measurements, psychotherapy clinical trials would be limited largely by the number of willing participants eligible for the study, rather than by the resources required to screen, assess, treat, and follow these participants. This could open the door to unprecedentedly large-N clinical trials. This would allow for well-powered, sophisticated dismantling studies to support the search for mechanisms of change in psychotherapy, which are currently only possible using individual participant level meta-analysis (for example, see ref. 86). Ultimately, such insights into causal mechanisms of change in psychotherapy could help to refine these treatments and potentially improve their efficacy.

Finally, the emergence of LLM treatment modalities will challenge (or confirm) fundamental assumptions about psychotherapy. Does therapeutic (human) alliance account for a majority of the variance in patient change? To what extent can an alliance be formed with a technological agent? Is lasting and meaningful therapeutic change only possible through working with a human therapist? LLMs hold the promise of empirical answers to these questions.

In summary, large language models hold promise for supporting, augmenting, or even in some cases replacing human-led psychotherapy, which may improve the quality, accessibility, consistency, and scalability of therapeutic interventions and clinical science research. However, LLMs are advancing quickly and will soon be deployed in the clinical domain, with little oversight or understanding of harms that they may produce. While cautious optimism about clinical LLM applications is warranted, it is also crucial for psychologists to approach the integration of LLMs into psychotherapy with caution and to educate the public about the potential risks and limitations of using these technologies for therapeutic purposes. Furthermore, clinical psychologists ought to actively engage with the technologists building these solutions. As the field of AI continues to evolve, it is essential that researchers and clinicians closely monitor the use of LLMs in psychotherapy and advocate for responsible and ethical use to protect the wellbeing of patients.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Received: 24 July 2023; Accepted: 30 January 2024;

Published online: 02 April 2024

References

1. Bubeck, S. et al. Sparks of artificial general intelligence: Early experiments with GPT-4. Preprint at <http://arxiv.org/abs/2303.12712> (2023).
2. Broderick, R. People are using AI for therapy, whether the tech is ready for it or not. *Fast Company* (2023).
3. Weizenbaum, J. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 36–45 (1966).
4. Bantilan, N., Malgaroli, M., Ray, B. & Hull, T. D. Just in time crisis response: Suicide alert system for telemedicine psychotherapy settings. *Psychother. Res.* 31, 289–299 (2021).
5. Peretz, G., Taylor, C. B., Ruzek, J. I., Jefroykin, S. & Sadeh-Shavit, S. Machine learning model to predict assignment of therapy homework in behavioral treatments: Algorithm development and validation. *JMIR Form. Res.* 7, e45156 (2023).
6. Tanana, M. J. et al. How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behav. Res. Methods* 53, 2069–2082 (2021).
7. Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C. & Althoff, T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat. Mach. Intell.* 5, 46–57 (2023).
8. Chen, Z., Flemotomas, N., Imel, Z. E., Atkins, D. C. & Narayanan, S. Leveraging open data and task augmentation to automated behavioral coding of psychotherapy conversations in low-resource scenarios. Preprint at <https://doi.org/10.48550/arXiv.2210.14254> (2022).

9. Shah, R. S. et al. Modeling motivational interviewing strategies on an online peer-to-peer counseling platform. *Proc. ACM Hum.-Comput. Interact.* 6, 1–24 (2022).
10. Chan, W. W. et al. The challenges in designing a prevention chatbot for eating disorders: Observational study. *JMIR Form. Res.* 6, e28003 (2022).
11. Darcy, A. Why generative AI Is not yet ready for mental healthcare. *Woebot Health* <https://woebothealth.com/why-generative-ai-is-not-yet-ready-for-mental-healthcare/> (2023).
12. Abd-Alrazaq, A. A. et al. An overview of the features of chatbots in mental health: A scoping review. *Int. J. Med. Inf.* 132, 103978 (2019).
13. Lim, S. M., Shiau, C. W. C., Cheng, L. J. & Lau, Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. *Behav. Ther.* 53, 334–347 (2022).
14. Baumel, A., Muench, F., Edan, S. & Kane, J. M. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *J. Med. Internet Res.* 21, e14567 (2019).
15. Torous, J., Nicholas, J., Larsen, M. E., Firth, J. & Christensen, H. Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evid. Based Ment. Health* 21, 116–119 (2018b).
16. Das, A. et al. Conversational bots for psychotherapy: A study of generative transformer models using domain-specific dialogues. in *Proceedings of the 21st Workshop on Biomedical Language Processing* 285–297 (Association for Computational Linguistics, 2022). <https://doi.org/10.18653/v1/2022.bionlp-1.27>.
17. Liu, H. Towards automated psychotherapy via language modeling. Preprint at <http://arxiv.org/abs/2104.10661> (2021).
18. Hamilton, J. Why generative AI (LLM) is ready for mental healthcare. *LinkedIn* <https://www.linkedin.com/pulse/why-generative-ai-chatgpt-ready-mental-healthcare-jose-hamilton-md/> (2023).
19. Shariff, A., Bonnefon, J.-F. & Rahwan, I. Psychological roadblocks to the adoption of self-driving vehicles. *Nat. Hum. Behav.* 1, 694–696 (2017).
20. Markov, A. A. Essai d'une recherche statistique sur le texte du roman "Eugene Onegin" illustrant la liaison des épreuve en chaîne ('Example of a statistical investigation of the text of "Eugene Onegin" illustrating the dependence between samples in chain'). *Izvistia Imperatorskoi Akad. Nauk Bull. L'Academie Imp. Sci. St-Petersbourg* 7, 153–162 (1913).
21. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423 (1948).
22. Baker, J. K. Stochastic modeling for automatic speech understanding. in *Speech recognition: invited papers presented at the 1974 IEEE symposium* (ed. Reddy, D. R.) (Academic Press, 1975).
23. Jelinek, F. Continuous speech recognition by statistical methods. *Proc. IEEE* 64, 532–556 (1976).
24. Jurafsky, D. & Martin, J. H. N-gram language models. in *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (Pearson Prentice Hall, 2009).
25. Vaswani, A. et al. Attention is all you need. *31st Conf. Neural Inf. Process. Syst.* (2017).
26. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at <http://arxiv.org/abs/2108.07258> (2022).
27. Gao, L. et al. The Pile: An 800GB dataset of diverse text for language modeling. Preprint at <http://arxiv.org/abs/2101.00027> (2020).
28. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. Preprint at <http://arxiv.org/abs/1810.04805> (2019).
29. Kojima, T., Gu, S. S., Reid, M., Matsuo, Y. & Iwasawa, Y. Large language models are zero-shot reasoners. Preprint at <http://arxiv.org/abs/2205.11916> (2023).
30. Fairburn, C. G. & Patel, V. The impact of digital technology on psychological treatments and their dissemination. *Behav. Res. Ther.* 88, 19–25 (2017).
31. Fisher, A. J. et al. Open trial of a personalized modular treatment for mood and anxiety. *Behav. Res. Ther.* 116, 69–79 (2019).
32. Fan, X. et al. Utilization of self-diagnosis health chatbots in real-world settings: Case study. *J. Med. Internet Res.* 23, e19928 (2021).
33. Coghlan, S. et al. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digit. Health* 9, 1–11 (2023).
34. Beatty, C., Malik, T., Meheli, S. & Sinha, C. Evaluating the therapeutic alliance with a free-text CBT conversational agent (Wysa): A mixed-methods study. *Front. Digit. Health* 4, 847991 (2022).
35. Lin, B., Bounoufouf, D., Cecchi, G. & Varshney, K. R. Towards healthy AI: Large language models need therapists too. Preprint at <http://arxiv.org/abs/2304.00416> (2023).
36. Weidinger, L. et al. Ethical and social risks of harm from language models. Preprint at <http://arxiv.org/abs/2112.04359> (2021).
37. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (ACM, 2021). <https://doi.org/10.1145/3442188.3445922>.
38. Chamberlain, J. The risk-based approach of the European Union's proposed artificial intelligence regulation: Some comments from a tort law perspective. *Eur. J. Risk Regul.* 14, 1–13 (2023).
39. Norden, J. G. & Shah, N. R. What AI in health care can learn from the long road to autonomous vehicles. *NEJM Catal. Innov. Care Deliv.* <https://doi.org/10.1056/CAT.21.0458> (2022).
40. Sedlakova, J. & Trachsel, M. Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? *Am. J. Bioeth.* 23, 4–13 (2023).
41. Gearing, R. E. et al. Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clin. Psychol. Rev.* 31, 79–88 (2011).
42. Wiltsey Stirman, S. Implementing evidence-based mental-health treatments: Attending to training, fidelity, adaptation, and context. *Curr. Dir. Psychol. Sci.* 31, 436–442 (2022).
43. Waller, G. Evidence-based treatment and therapist drift. *Behav. Res. Ther.* 47, 119–127 (2009).
44. Flemotomas, N. et al. "Am I a good therapist?" Automated evaluation of psychotherapy skills using speech and language technologies. *CoRR*, Abs, 2102 (10.3758) (2021).
45. Zhang, X. et al. You never know what you are going to get: Large-scale assessment of therapists' supportive counseling skill use. *Psychotherapy* <https://doi.org/10.1037/pst0000460> (2022).
46. Goldberg, S. B. et al. Machine learning and natural language processing in psychotherapy research: Alliance as example use case. *J. Couns. Psychol.* 67, 438–448 (2020).
47. Wiltsey Stirman, S. et al. A novel approach to the assessment of fidelity to a cognitive behavioral therapy for PTSD using clinical worksheets: A proof of concept with cognitive processing therapy. *Behav. Ther.* 52, 656–672 (2021).
48. Raviola, G., Naslund, J. A., Smith, S. L. & Patel, V. Innovative models in mental health delivery systems: Task sharing care with non-specialist providers to close the mental health treatment gap. *Curr. Psychiatry Rep.* 21, 44 (2019).
49. American Psychological Association. Guidelines for clinical supervision in health service psychology. *Am. Psychol.* 70, 33–46 (2015).
50. Cook, S. C., Schwartz, A. C. & Kaslow, N. J. Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics* 14, 537–545 (2017).
51. Leichsenring, F., Steinert, C., Rabung, S. & Ioannidis, J. P. A. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: An umbrella review and meta-analytic evaluation of recent meta-analyses. *World Psych.* 21, 133–145 (2022).

52. Cuijpers, P., van Straten, A., Andersson, G. & van Oppen, P. Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies. *J. Consult. Clin. Psychol.* 76, 909–922 (2008).
53. Morris, Z. S., Wooding, S. & Grant, J. The answer is 17 years, what is the question: Understanding time lags in translational research. *J. R. Soc. Med.* 104, 510–520 (2011).
54. Chekroud, A. M. et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psych.* 20, 154–170 (2021).
55. Kazdin, A. E. Mediators and mechanisms of change in psychotherapy research. *Annu. Rev. Clin. Psychol.* 3, 1–27 (2007).
56. Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I. & Atkinson, P. M. Explainable artificial intelligence: An analytical review. *WIREs Data Min. Knowl. Discov.* 11, (2021).
57. Kelley, T. L. *Interpretation of Educational Measurements*. (World Book, 1927).
58. van Bronswijk, S. C. et al. Precision medicine for long-term depression outcomes using the Personalized Advantage Index approach: Cognitive therapy or interpersonal psychotherapy? *Psychol. Med.* 51, 279–289 (2021).
59. Scala, J. J., Ganz, A. B. & Snyder, M. P. Precision medicine approaches to mental health care. *Physiology* 38, 82–98 (2023).
60. Chorpita, B. F., Daleiden, E. L. & Weisz, J. R. Identifying and selecting the common elements of evidence based interventions: A distillation and matching model. *Ment. Health Serv. Res.* 7, 5–20 (2005).
61. Chambliss, D. L. & Hollon, S. D. Defining empirically supported therapies. *J. Consult. Clin. Psychol.* 66, 7–18 (1998).
62. Tolin, D. F., McKay, D., Forman, E. M., Klonsky, E. D. & Thombs, B. D. Empirically supported treatment: Recommendations for a new model. *Clin. Psychol. Sci. Pract.* 22, 317–338 (2015).
63. Lilienfeld, S. O. Psychological treatments that cause harm. *Perspect. Psychol. Sci.* 2, 53–70 (2007).
64. Wasil, A. R., Venturo-Conery, K. E., Shingleton, R. M. & Weisz, J. R. A review of popular smartphone apps for depression and anxiety: Assessing the inclusion of evidence-based content. *Behav. Res. Ther.* 123, 103498 (2019).
65. Torous, J. B. et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr. Serv.* 69, 498–500 (2018).
66. Gunasekar, S. et al. Textbooks are all you need. Preprint at <http://arxiv.org/abs/2306.11644> (2023).
67. Wilhelm, E. et al. Measuring the burden of infodemics: Summary of the methods and results of the Fifth WHO Infodemic Management Conference. *JMIR Infodemiology* 3, e44207 (2023).
68. Creed, T. A. et al. Knowledge and attitudes toward an artificial intelligence-based fidelity measurement in community cognitive behavioral therapy supervision. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 49, 343–356 (2022).
69. Aktan, M. E., Turhan, Z. & Dolu, İ. Attitudes and perspectives towards the preferences for artificial intelligence in psychotherapy. *Comput. Hum. Behav.* 133, 107273 (2022).
70. Prescott, J. & Hanley, T. Therapists' attitudes towards the use of AI in therapeutic practice: considering the therapeutic alliance. *Ment. Health Soc. Incl.* 27, 177–185 (2023).
71. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (2013).
72. Yogatama, D., De Masson d'Autume, C. & Kong, L. Adaptive semiparametric language models. *Trans. Assoc. Comput. Linguist* 9, 362–373 (2021).
73. Stanley, B. & Brown, G. K. Safety planning intervention: A brief intervention to mitigate suicide risk. *Cogn. Behav. Pract.* 19, 256–264 (2012).
74. Behzadan, V., Munir, A. & Yampolskiy, R. V. A psychopathological approach to safety engineering in AI and AGI. Preprint at <http://arxiv.org/abs/1805.08915> (2018).
75. Lambert, M. J. & Harmon, K. L. The merits of implementing routine outcome monitoring in clinical practice. *Clin. Psychol. Sci. Pract.* 25, (2018).
76. Kjell, O. N. E., Kjell, K. & Schwartz, H. A. AI-based large language models are ready to transform psychological health assessment. Preprint at <https://doi.org/10.31234/osf.io/yfd8g> (2023).
77. First, M. B., Williams, J. B. W., Karg, R. S. & Spitzer, R. L. *SCID-5-CV: Structured Clinical Interview for DSM-5 Disorders: Clinician Version*. (American Psychiatric Association Publishing, 2016).
78. Shah, D. S., Schwartz, H. A. & Hovy, D. Predictive biases in natural language processing models: A conceptual framework and overview. in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* 5248–5264 (Association for Computational Linguistics, 2020). <https://doi.org/10.18653/v1/2020.acl-main.468>.
79. Adams, L. M. & Miller, A. B. Mechanisms of mental-health disparities among minoritized groups: How well are the top journals in clinical psychology representing this work? *Clin. Psychol. Sci.* 10, 387–416 (2022).
80. Viswanath, H. & Zhang, T. FairPy: A toolkit for evaluation of social biases and their mitigation in large language models. Preprint at <http://arxiv.org/abs/2302.05508> (2023).
81. von Zitzewitz, J., Boesch, P. M., Wolf, P. & Riener, R. Quantifying the human likeness of a humanoid robot. *Int. J. Soc. Robot.* 5, 263–276 (2013).
82. White House Office of Science and Technology Policy. Blueprint for an AI bill of rights. (2022).
83. Parry, G., Castonguay, L. G., Borkovec, T. D. & Wolf, A. W. Practice research networks and psychological services research in the UK and USA. in *Developing and Delivering Practice-Based Evidence* (eds. Barkham, M., Hardy, G. E. & Mellor-Clark, J.) 311–325 (Wiley-Blackwell, 2010). <https://doi.org/10.1002/9780470687994.ch12>.
84. Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T. & Vervliet, B. Maximizing exposure therapy: An inhibitory learning approach. *Behav. Res. Ther.* 58, 10–23 (2014).
85. Delgadillo, J. et al. Stratified care vs stepped care for depression: A cluster randomized clinical trial. *JAMA Psychiatry* 79, 101 (2022).
86. Furukawa, T. A. et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: A systematic review and component network meta-analysis using individual participant data. *Lancet Psychiatry* 8, 500–511 (2021).

Acknowledgements

This work was supported by the National Institute of Mental Health under award numbers R01-MH125702 (PI: H.A.S) and RF1-MH128785 (PI: S.W.S.), and by the Institute for Human-Centered A.I. at Stanford University to J.C.E. The authors are grateful to Adam S. Miner and Victor Gomes who provided critical feedback on an earlier version of this manuscript.

Author contributions

E.C.S., S.W.S., C.L.B., and J.C.E. wrote the main manuscript text. E.C.S., L.H.U., and J.C.E. prepared the figures. All authors reviewed the manuscript.

Competing interests

The authors declare the following competing interests: receiving consultation fees from Jimini Health (E.C.S., L.H.U., H.A.S., and J.C.E.).

Additional information

Correspondence and requests for materials should be addressed to Elizabeth C. Stade or Johannes C. Eichstaedt.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

